Alma Mater Studiorum - Università di Bologna
in cotutela con University of Luxembourg - Universitè du Luxembourg

DOTTORATO DI RICERCA IN

LAW, SCIENCE AND TECHNOLOGY

Ciclo 34

Settore Concorsuale: 12/H3 - FILOSOFIA DEL DIRITTO

Settore Scientifico Disciplinare: IUS/20 - FILOSOFIA DEL DIRITTO

THE EUROPEAN LEGAL APPROACH TO OPEN SCIENCE AND RESEARCH DATA

Presentata da: Ludovica Paseri

Coordinatore Dottorato

Monica Palmirani

Supervisore

Massimo Durante

Supervisore

Ugo Pagallo

Co-supervisore

Pascal Bouvry

Esame finale anno 2022

PhD-FSTM-2022-69
The Faculty of Science, Technology and Medicine

Alma Mater Studiorum University of Bologna – CIRSFID

DISSERTATION

Presented on 17/06/2022 in Bologna, Italy
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN INFORMATIQUE

AND

DOTTORE DI RICERCA IN

LAW, SCIENCE, AND TECHNOLOGY

By

Ludovica PASERI
Born on 31 October 1992 in Saluzzo (Italy)

The European Legal Approach to Open Science and Research Data

Dissertation Defence Committee

Prof. Dr. Giovanni Boniolo, *Università di Ferrara*, *Chairman*
Dr. Paolo Guarda, *Università di Trento*, *Vice-Chairman*
Dr. Ginevra Peruginelli, *IGSG CNR*, *Member*
Prof. Dr. Massimo Durante, *Università di Torino*, *Dissertation Supervisor*
Prof. Dr. Pascal Bouvry, *Université du Luxembourg*, *Dissertation Supervisor*

Abstract

This dissertation proposes an analysis of the governance of the European scientific research, focusing on the emergence of the Open Science paradigm. The paradigm of Open Science indicates a new way of doing science, oriented towards the openness of every phase of the scientific research process, and able to take full advantage of the digital Information and Communication Technologies (ICTs). The emergence of this paradigm is relatively recent, but in the last couple of years it has become increasingly relevant. The European institutions expressed a clear intention to embrace the Open Science paradigm, with several interventions and policies on this matter. Among many, consider, for example, the project of the European Open Science Cloud (EOSC), a federated and trusted environment for access and sharing of research data and services for the benefit of the European researchers; or the establishment of the new research funding programme, i.e., the Horizon Europe programme, laid down in the EU Regulation 2021/695, which links research funding to the adoption of the Open Science tenets. This dissertation examines the European approach to Open Science, providing a conceptual framework for the multiple interventions of the European institutions in the field of Open Science, as well as addressing the major legal challenges that the implementation of this new paradigm is generating. To this aim, the study first investigates the notion of Open Science, in order to understand what specifically falls under the umbrella of this broad term: it is proposed a definition that takes into account all its dimensions and an analysis of the human and fundamental rights framework in which Open Science is grounded. After that, the inquiry addresses the legal challenges related to the openness of research data, in light of the European legislative framework on Open Data. This also requires drawing attention to the European data protection framework, analysing the impact of the General Data Protection Regulation (GDPR) on the context of Open Science. The last part of the study is devoted to the infrastructural dimension of the Open Science paradigm, exploring the digital infrastructures that are increasingly an integral part of the scientific research process. In particular, the focus is on a specific type of computational infrastructure, namely the High Performance Computing (HPC) facility. The adoption of HPC for research is analysed both from the European perspective, investigating the EuroHPC project, and the local perspective, proposing the case study of the HPC facility of the University of Luxembourg, namely the ULHPC. This dissertation intends to underline the relevance of the legal coordination approach, between all actors and phases of the scientific research process, in order to develop and implement the Open Science paradigm, adhering to the underlying human and fundamental rights.

Table of Contents

List of Abbreviations

AI	Artificial Intelligence
AIA	Artificial Intelligence Act
APC	Article Processing Charge
BDVA	Big Data Value Association
CINI	National Interuniversity Consortium for Informatics
CLOUD	Clarifying Lawful Overseas Use of Data Act
CNIL	Commission Nationale Informatique & Libertés
CoEs	Centres of Excellence
CORD-19	COVID-19 Open Research Dataset
CORDIS	Community Research and Development Information Service
DGA	Data Governance Act
DMA	Digital Markets Act
DPIA	Data Protection Impact Assessment
DSA	Digital Services Act
ECHR	European Court on Human Rights
ECJ	European Court of Justice
EDI	European Data Infrastructure
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EMA	European Medicines Agency
ENISA	European Network and Information Security Agency
EOSC	European Open Science Cloud
EPI	European Processor Initiative Consortium

ERA	European Research Area
ESFRI	European Strategy Forum on Research Infrastructures
ETP4HPC	European Technology Platform for HPC
FAIR	Findable, Accessible, Interoperable, Reusable
GATS	General Agreement on Trade in Services
GDPR	General Data Protection Regulation
HPC	High Performance Computing
ICDI	Italian Computing and Data Infrastructure
ICO	Information Commissioner's Office
ICTs	Information and Communication Technologies
ISO	International Organization for Standardization
LERU	League of European Research Universities
LoA	Level of Abstractions
ML	Machine Learning
MLAT	Mutual Legal Assistance Treaty
MOOC	Massive Open Online Courses
NIST	National Institute of Standards and Technologies
ODD	Open Data Directive
OECD	Organization for Economic Co-operation and Development
PRACE	Partnership for Advanced Computing in Europe
PSI	Public Sector Information
RDA	Research Data Alliance
SRIA	Strategic Research and Innovation Agenda
UN	United Nations
TFEU	Treaty on the Functioning of the European Union

WHO	World Health Organization
WP29	Article 29 Data Protection Working Party
WTO	World Trade Organization

Chapter 1

Introduction

On December 21, 2020, the European Medicines Agency (EMA) authorised the dissemination, on the territory of the European Union[1], of the first vaccine against the SARS-CoV-2 virus, in light of the effectiveness emerged from the clinical trial data[2], in the experimental phase. Less than a year after the first official report of the disease to the World Health Organization (WHO)[3], we had a vaccine capable of coping with the emergency generated by the COVID-19 pandemic, that marked the beginning of the second decade of the twenty-first century. A shared effort of resources, knowledge, and skills allowed the fastest ever production of the only means able to limit an infectious disease that, in the first 12 months of spread, caused more than 2 million deaths[4].

On January 27, 2021, the 46th elected President of the United States, Joe Biden, just seven days after the inauguration day, issued a Memorandum for "the heads of executive departments and agencies on restoring trust in government through scientific integrity and evidence-based policymaking". In this Memorandum, President Biden stated that his Administration intends "[...] to make evidence-based decisions guided by the best available science and data. Scientific and technological information, data, and evidence are central to the development and iterative improvement of sound policies", stressing

[1] Similarly, the U.S. Food and Drug Administration (FDA) gave its authorisation, a few days earlier, on December 11, 2020. See: https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19.

[2] https://www.ema.europa.eu/en/news/ema-recommends-first-covid-19-vaccine-authorisation-eu.

[3] It occurred on January 31, 2019; see: https://www.who.int/news/item/13-04-2020-public-statement-for-collaboration-on-covid-19-vaccine-development.

[4] Based on data collected by WHO. See, the "WHO Coronavirus Disease (COVID-19) Dashboard": https://covid19.who.int.

that "[W]hen scientific or technological information is considered in policy decisions, it should be subjected to well-established scientific processes, including peer review where feasible and appropriate, with appropriate protections for privacy"[5].

The effort in developing the vaccine against the SARS-CoV-2 virus and the Memorandum of the US President are two emblematic examples of the relevance of science in society.

By contrast, a different drift is emerging, pushing in the opposite direction. On January 6, 2021, in Washington, for the first time since 1814, a large group of protesters broke into the Capitol of the United States, with the intention of contesting the outcome of the recent presidential elections. The relevant aspect is that all of them were supporters of sects alleging various conspiracies, all with the common ground of not being supported by any scientific evidence.

In the same direction, another event should be noted. The deluge of data and information related to the COVID-19 pandemic, which has led to talk of "*infodemic*"[6], connected with the restrictive measures of freedom of citizens issued by the States to deal with the emergency, have seen a considerable increase in conspiracy theories and denialist of various kinds. The extent of this phenomenon was so significant that

---

[5] JOE R. BIDEN, "Memorandum on Restoring Trust in Government Through Scientific Integrity and Evidence-Based Policymaking", The White House, January 27, 2020. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/memorandum-on-restoring-trust-in-government-through-scientific-integrity-and-evidence-based-policymaking/.

[6] Among others, see: LUCIANO FLORIDI, *Il verde e il blu. Idee ingenue per migliorare la politica*. (Milano: Cortina Editore, 2020), p. 269, where the origin of the phrase is mentioned, i.e., the 2003 SARS disease outbreak. It has even become the subject of a call for actions, organised by the WHO, aimed to mitigate harm from health misinformation among online and offline communities, see: https://www.who.int/news/item/11-12-2020-call-for-action-managing-the-infodemic. On the topic, an interesting research project has been conducted, showing that, in the infodemic of COVID-19, although false information is considerably retweeted, it is comparatively less so than science-based information, see: CRISTINA M. PULIDO, *et al*., "COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information." *International Sociology* 35.4 (2020): 377–392, doi: 10.1177/0268580920914755.

UNESCO and the European Commission deemed it necessary to launch a campaign to deal with the spread of these denialist theories[7].

In this scenario, characterised by such strongly opposing trends, the debate on the role of science in society and its relationship with politics and democracy is more vivid, alive, and necessary than ever. As a consequence, a study on the governance of scientific research and on the management of research data becomes of primary importance.

The intention is not to embark on the philosophical debate on the relationship between science, democracy, and society, as represented, for instance, by the contrasting perspectives of Robert Merton[8] and Michael Polanyi[9], in the twentieth century, or by the interpretation of Jürgen Habermas[10].

The purpose is, instead, to investigate the use of new technologies in the world of science, their impact on the way of doing research, and the

[7] https://en.unesco.org/themes/gced/thinkbeforesharing.

[8] In particular, on the normative structure of science and on the concept of *ethos* of science, see ROBERT K. MERTON, *The sociology of science. Theoretical and empirical investigations*. (Chicago: UCP, 1973), pp. 267-278. Here the Author identifies four sets of institutional imperatives that apply to science: universalism, communism, disinterestedness, and systematic skepticism. To them then, one must add "[...] the socially enforced value of humility", in order to counterbalance the emphasis on originality, as argued in ROBERT K. MERTON, "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22.6 (1957): p. 646. In addition, on the analysis of the social conditions in which science can thrive (and, conversely, those that result in hostility), see ROBERT K. MERTON, "Science and the Social Order." *Philosophy of Science* 5.3 (1938): 321–337.

[9] Traditionally opposed to Merton's vision, Michael Polanyi embodies a liberal perspective, that emerges from his work: MICHAEL POLANYI, *et al.*, "The Republic of Science: its political and economic theory", *Minerva* 1.1 (1962): 54-73. In Minerva, 38.1 (2000): 1-32. This work contains the essence of his most famous work "Personal Knowledge", in which Polanyi outlines a comparison between the concept of scientific community and society, on the basis that "[T]he organization of the community of scientists has an obvious bearing upon the problem of political organization", as described by De Jouvenel, about Polanyi's thought, in BETRAND DE JOUVENAL, "The Republic of Science", *The Logic of Personal Knowledge.* (London: Routledge and Kegan Paul, 1961): p. 140.

[10] See, in particular, "The idea of the theory of knowledge as a social theory", in JÜRGEN HABERMAS, *Knowledge and human interest*. (Boston: Beacon Press, 1972), 43-64; on the concept of deliberative democracy, it is, then, fundamental the reasoning exposed in: JÜRGEN HABERMAS, *Between facts and norms. Contributions to a discourse theory of law and democracy*. (Cambridge, Massachusetts: The MIT Press, 1996), 287-328, (in particular, chapter 7: "Deliberative politics: a procedural concept of democracy"). For the development of this analysis, with specific reference to the media society, at the basis of the collective deliberation: JÜRGEN HABERMAS, *Europe. The faltering project.* (Cambridge: Polity Press, 2008), pp. 138-183.

reaction of institutions, primarily European, to this changing science. In so doing, it is necessary to pay attention to the founding pillars of these dynamics: the opposing trends just described illustrate that nothing can be taken for granted.

Without delving into the debate on the *ethos* of science, it is worth drawing on that strand of knowledge to bring out an aspect that is extremely relevant to this investigation. This aspect, which is common to both Merton's and Polanyi's analyses, relates to the public nature of the dialogue between scientists[11], within the scientific community[12]. It is on this free and open exchange of ideas, perspectives, results, and data that science is founded. At the basis, there is, therefore, an uninterrupted dialogue that outlines a *status* of continuous research: there is never certainty, immutability, crystallisation; there is, rather, a continuous wondering, inquiring, and searching for answers.

From this perspective, scientific knowledge and democracy have much in common. In the same way as science, democracy is not given once and for all, it is constantly tested. As the philosopher of law

---

[11] This aspect has been clearly pointed out in: ROBERTO CASO, *La rivoluzione incompiuta: La scienza aperta tra diritto d'autore e proprietà intellettuale*. (Milano: Ledizioni, 2019), 94. In particular, Merton refers to the concept of social collaboration of science through the elaboration of the concept of "communism": "[T]he substantive findings of science are a product of social collaboration and are assigned to a community.", in ROBERT K. MERTON, *The sociology of science, op. cit*, p. 273. Polanyi, on the other hand, while being a firm believer in the total independence of scientists, focusing on communication and conviviality (see: MICHAEL POLANYI, *Personal knowledge: towards a post-critical philosophy*. (London: Routledge & Kegan Paul, 1958), p. 203), pays particular attention to the tacit communication of knowledge. On this particular aspect see: MICHAEL POLANYI, *The tacit dimension*. (Gloucester, Mass: Peter Smith ed., 1983). On the Polanyi's concept of "tacit knowledge" see, also: GIOVANNI BONIOLO, *Conoscere per vivere: Istruzioni per sopravvivere all'ignoranza*. (Milano: Mimesis, 2018), p. 54.

[12] In accordance then with their perspectives, this common aspect is differently declined. The consequence for Merton has to do with the concept of property, stating that: "[P]roperty rights in science are whittled down to a bare minimum by the rationale of the scientific ethic. [...] The communism of the scientific ethos is incompatible with the definition of technology as 'private property' in a capitalistic economy." in ROBERT K. MERTON, *The sociology of science, op. cit.*, pp. 273, 275. For Polanyi, instead, this dialogue – however public – is still limited to certain conditions: "[A] dialogue can be sustained only if both participants belong to a community accepting on the whole the same teaching and tradition for judging their own affirmations. A responsible encounter presupposes a common firmament of superior knowledge", in MICHAEL POLANYI, *Personal knowledge, op. cit.*, p. 378.

Norberto Bobbio said: “[D]emocracy is dynamic”[13]. It is fed by a continuous tension that is the source of its strength and weakness[14]. This tension that characterises democracy is made even more evident in societies whose well-being depends on the Information and Communication Technologies (ICTs), as argued by the philosopher of law Ugo Pagallo[15].

It is against this backdrop that this dissertation should be placed: it aims to examine the policies related to science and the technological revolution, with particular attention to the European dimension. These policies, in fact, enable an assessment of the regulatory framework for the governance of scientific research.

To achieve this purpose, two premises must be considered: (i) a methodological premise and (ii) an institutional premise[16].

First consider the methodological premise, which concerns the change that is taking place in the methodology of science[17]. The context

[13] On the dynamic nature of the democracy, see: “For a democratic system, the process of ‘becoming’, of transformation, is its natural state. Democracy is dynamic, despotism is static and always essential the same.” in NORBERTO BOBBIO, *The future of democracy: a defence of the rules of the game*. (Minneapolis: University of Minnesota Press, 1987), p. 17.

[14] Still following the comparison with democracy, Bobbio provides an example of dynamism as a strength. The Author, in fact, arguing that such dynamism, permitted and guaranteed by respecting the procedural rules of democracy, states: “[…] there is the ideal of gradual renewal of society via the free debate of ideas and the modification of attitudes and ways of life: only democracy allows silent revolutions to take shape and spread, as has happened in the case of the relationship between the sexes in the last few decades, which is probably the greatest revolution of our age”, in NORBERTO BOBBIO, *The future of democracy, op. cit.*, p. 42. In relation to weakness, to the contrary, an example of how such dynamism can pose a risk is provided, recently, by the first speech of Kamala Harris, first woman Vice President of the United States, who remarks: “Congressman John Lewis, before his passing, wrote: ‘Democracy is not a state. It is an act’. And what he meant was that America’s democracy is not guaranteed. It is only as strong as our willingness to fight for it, to guard it and never take it for granted. And protecting our democracy takes struggle. It takes sacrifice.”, in “Kamala Harris’ full victory speech”, CNN, Youtube, November 8, 2020, https://www.youtube.com/watch?v=ExPm_hJQYpQ.

[15] See, in particular: UGO PAGALLO, “The broken promises of democracy in the information era.” in CORIEN PRINS, *et al.* (eds), *Digital democracy in a globalized world*. (Cheltenham, UK: Edward Elgar Publishing, 2017), 77-99, doi: 10.4337/9781785363962.

[16] These two premises will be further analysed in Section 2.1.

[17] It should be specified that the methodological premise is not related to the methodology of this dissertation, but they are related to science itself. For a description of the methodology of this dissertation, see: Section 1.2.

of science traditionally understood is undergoing a significant revolution, given the impact of the ICTs and the Internet. This revolution can be represented as a real paradigm shift[18], with the establishment of the so-called "Open Science paradigm." The expression Open Science hints an "umbrella concept" that, in general, intends to describe the profound transformation to which science is subjected in recent decades, due to the impact of the digital revolution. The impact of the ICTs has, indeed, affected the deepest and traditionally established dynamics, changing the field of science.

Today science embraces a more open stance. The Open Science paradigm intends to affirm the openness of all the phases of scientific research, from data collection to publication of results, for the sake of a global research, based on collaboration and closely connected with the society. This openness, however, is not blind and absolute, but it pursues the formula "as open as possible, as closed as necessary"[19]. This formula begets a balance between the intention to openness and sharing of the research results and the need for closure of data and safeguards for the protection of specific interests (e.g., the right to privacy, national security, or public order).

[18] The concept of "paradigm shift" is specifically intended to hint at the nature of scientific revolutions, the role of theories and the progress resulting from revolutions, as exposed by THOMAS KUHN, *The structure of scientific revolutions*. (Chicago: University of Chicago Press, 1970), pp. 43-51, and also pp. 160-173. According to Kuhn, science goes through cycles: from normality; through crisis; into revolution, a situation that allows the shift into a new paradigm. In this dissertation, the aim is not to argue the emergence of a new paradigm of science from a philosophical point of view. On this matter, see: ANTONY JG. HEY, STEWART TANSLEY, KRISTIN MICHELE TOLLE, *The fourth paradigm: data-intensive scientific discovery*. (Redmond, WA: Microsoft research, 2009). The Authors have developed a temporal scanning that distinguishes four paradigms of science: the first paradigm, represented by the experimental science, thousand years ago; then, the theoretical science represents the second paradigm; the third paradigm, the so-called "computational", with the introduction of computers, able to simulate complex phenomena; and the fourth paradigm, the current one, represented by data-intensive science. Our purpose is to understand how, starting from this new paradigm of science, the actors of the system have operated so far or should operate in order to set the governance of this new context.

[19] https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science_en.

The second premise is institutional: it has to do with the fact that the European institutions have expressed a commitment to support the Open Science paradigm. This stand is evident in the design of policies on scientific research and innovation released in recent years. As will be analysed in Chapter 2, institutions intend to assume the burden of creating the conditions for the development of scientific research, embracing the Open Science paradigm. The Open Science paradigm is also an expression of the process of technological convergence[20]. It concerns, in fact, (i) increased quantity of research data created or collected; (ii) techniques for analysing such data and the infrastructures for implementing them; (iii) new computational power provided by High Performance Computing (HPC) facilities.

In other words, the first premise relates to the relationship between science and technology, and the second features the relationship between science and law: science is experiencing a paradigm shift, represented by Open Science, triggered by the introduction of new technologies; and the institutions, at various levels[21], support this paradigm shift by proposing attempts to implement it in practice.

Assumed these premises, this dissertation intends to be part of the legal research strand that investigates the proper governance of scientific research. Hence, the intention is to focus on the regulatory disciplines at European level, involved in scientific research and, in particular, in the Open Science paradigm, analysing the major legal

---

[20] On the concept of technological convergence, see, among others: UGO PAGALLO, "Algo-rhythms and the beat of the legal drum." *Philosophy & Technology* 31.4 (2018): 507-524, doi: 10.1007/s13347-017-0277-z; and UGO PAGALLO, MASSIMO DURANTE, SHARA MONTELEONE, "What Is New with the Internet of Things in Privacy and Data Protection? Four Legal Challenges on Sharing and Control in IoT." in: RONALD LEENES, *et. al.*, (eds) *Data Protection and Privacy: (In)visibilities and Infrastructures. Law, Governance and Technology Series*, 36 (2017): 59-78, doi: 10.1007/978-3-319-50796-5_3.

[21] On the several levels of the institutions involved, see Section 2.4.3.1. However, consider, first of all, the European level; but, also, the international level, in relation to the UNESCO initiatives; and also the national level, both within the European Union, and outside. See the example of Canada, with "Canada's 2018-2020 National Action Plan on Open Government," which devotes extensive attention to Open Science, with a commitment that is primarily directed at the internal "intramural" science of the federal government. See: https://open.canada.ca/en/content/canadas-2018-2020-national-action-plan-open-government#toc8.

issues. In the perspective of the technological convergence, attention will be also paid to the interaction between Open Science projects and the adoption of cloud computing and HPC capabilities for scientific research purposes.

After this general overview, Section 1.1 depicts the research problem of this study, exposing the related research questions.

Then, Section 1.2 illustrates the methodology adopted to conduct this dissertation.

Finally, Section 1.3 describes the structure of the study, briefly outlining the content of each chapter.

1.1 Research Problem and Questions

In light of the premises just described, the purpose of this study is to investigate the European approach to Open Science, understood as the profound transformation to which science is subjected in the recent decades, due to the impact of the digital revolution.

Recently, the expression Open Science appears as a buzzword, and it evokes many concepts such as "Open Access", "Open Data", "Open Source", "Open Innovation", etc. However, these concepts are extremely different and there is a need to understand the relationship between them and their respective roles in the EU policies. For this reason, the main research question is:

RQ: *Which approach has the European Union adopted on Open Science and what are the related legal challenges which arise?*

Many issues need to be untangled; therefore, a number of sub-research questions are identified.

Considering the Open Science scenario at European level, the first problem that emerges is the complexity resulting from a proliferation of policies, documents, and initiatives: there is a need for a systematisation of the EU interventions on Open Science.

Hence the first sub-set of research questions is:

RQ2: *What has been done so far at European level in the field of Open Science? What are the legislative disciplines involved? What is the approach to Open Science governance that should be followed?*

Having clarified the evolution of the European Open Science policies, the legislative disciplines involved and the governance approach to be adopted, it is necessary to focus on the notion of "Open Science" in itself. It is important to understand what – today – falls under the notion of Open Science, wondering:

RQ3: *What should be understood by the Open Science paradigm? What are its foundations? Which are its different dimensions and how do they interact with each other?*

Answering RQ3 will clarify the relationship between the general Open Science paradigm and its different dimensions. One of its dimensions is represented by the open research data. At this point, the first legal issue arises: the coordination between the EU Open Science policies and the European Open Data framework. The related set of sub-research questions is the following:

RQ4: *What is the relationship between EU Open Science policies and the European Open Data legal framework? How do the two interact?*

Afterwards, a further legal issue will be addressed, represented by the relationship between the European approach to Open Science and the data protection framework, answering the following RQ5:

RQ5: *Which issues arise from the protection of personal data in the context of scientific research? Does the emergence of the Open Science paradigm complexifies the compliance with the EU data protection discipline?*

Finally, the last set of research questions deals with the infrastructural dimension of the Open Science paradigm. For this reason, the following question will be posed:

RQ6: *What role do e-infrastructures for research play in the Open Science paradigm?*

In particular, the investigation prompted by this last question will be supported by a case study, represented by the HPC platform of the University of Luxemborug (ULHPC), used for scientific research purposes[22].

Table 1.1 summarises the research questions guiding this dissertation.

Table 1.1: A Summary of the Research Questions

Main Research Question	**Which approach has the European Union adopted on Open Science and what are the related legal challenges which arise?**
RQ2	What has been done so far at European level in the field of Open Science? What are the legislative disciplines involved? What is the approach to Open Science governance that should be followed?
RQ3	What should be understood by the Open Science paradigm? What are its foundations? Which are its different dimensions and how do they interact with each other?
RQ4	What is the relationship between Open Science policies and the European Open Data legal framework? How do the two interact?
RQ5	Which issues arise from the protection of personal data in the context of scientific research? Does the emergence of the Open Science scenario complexifies the compliance with the EU data protection discipline?
RQ6	What role do e-infrastructures for research play in the Open Science paradigm?

[22] On the ULHPC, see: https://hpc.uni.lu.

1.2 Notes on Methodology

As a preliminary remark, it is necessary to give some insights on the methodology adopted to conduct this dissertation.

Three main considerations, developed below, should be borne in mind: (i) the study is framed within the field of law and technology; (ii) it has also an interdisciplinary orientation; and, (iii) it includes a case study which aims to deepen the link between the applicable law and the practice of a specific scientific community, in the use of technology.

1.2.1 Law and Technology

This study is grounded in the field of law and technology, a strand of research that investigates the complex relationship between law and the development of technology that characterises our time. The need to develop this emerging strand of research is clarified by Stefano Rodotà: "[T]he interplay between technological innovation, social change and legal solutions poses problems every day which often make the old criteria and known approaches totally inadequate"[23].

The field of law and technology is strictly related to the domain of philosophy of law. The domain of philosophy of law is engaged in exploring the conceptual foundations of the law of its age[24], in the same way that "[A] meaningful philosophy is a philosophy of its time", as argued by the philosopher Mariarosaria Taddeo[25]. Thus, in this era marked by the digital revolution, attention must be directed to the rule of law of the information society[26]. Accordingly, this research project

---

[23] STEFANO RODOTÀ, *Tecnologie e diritti*. (Bologna: Il Mulino, 1995), p. 9. [Translation from the Italian original text].

[24] This assumption is sound to the extent that the philosophical level is considered as: "[…] the context in which reason operates argumentatively around the justification of problem solving without the support of any revelation." in GIOVANNI BONIOLO, *Il limite e il ribelle. Etica, naturalismo, darwinismo*. (Milano: Cortina Editore, 2003), p. 77.

[25] MARIAROSARIA TADDEO, "Philosophy and computing in Information Societies." *Minds & Machines* 26 (2016): p. 204, doi: 10.1007/s11023-016-9400-7.

[26] On the concept of information society, see: LUCIANO FLORIDI, "The information society and its philosophy: Introduction to the special issue on 'the Philosophy of Information, its Nature, and future developments'." *The Information Society* 25.3 (2009): p. 153, in which the Author, in investigating the

intends to fit into the debate of the contemporary philosophy of law represented by the interplay between law and technology[27], with specific attention to the field of science. In this context, in order to draw some considerations about "[...] what the law is supposed to do (requirements), and what it is called to do (functions)"[28], the attention will be focused on the major legal issues identified.

In the context of today's science, characterised by the emergence of the Open Science paradigm, there is a growing complexity from the point of view of the governance[29]. This complexity stems from the multiplicity of disciplines involved, the many actors in the field, as well as the nature of science itself, even amplified by the use of ICTs: "[T]he more an issue impacts on the whole infrastructure of the system, the more complex such an issue is; but, the more complex an issue is, the less traditional notions of legal and political thought can tackle such complexity in terms of physical sanctions, national jurisdiction, or self-referential rule of law"[30].

Hence, in such a complexity, the legal challenges that most affect the emerging paradigm of science will be examined, with specific attention on the Open Data issues and the data protection concerns. This analysis is meant to develop some considerations about the European governance of scientific research, in a perspective of what the philosopher of law Massimo Durante calls "a democratic project of good governance"[31].

Accordingly, part of the aim of this study is to investigate the role of institutions *de lege ferenda*, namely in their ability to design the future

---

origins of the concept of information society states that: "[O]nly very recently have human progress and welfare begun to depend mostly on the successful and efficient management of the information life cycle."

[27] Ugo Pagallo, Massimo Durante, "The philosophy of law in an information society." in Luciano Floridi, *The Routledge Handbook of Philosophy of Information*, (New York: Routledge, 2016): p. 401.

[28] Ugo Pagallo, Massimo Durante, "*The philosophy of law*", *op. cit.*, p. 398.

[29] This complexity is illustrated in Section 2.4.3.

[30] Ugo Pagallo, Massimo Durante, "*The philosophy of law*", *op. cit.*, p. 402.

[31] Massimo Durante, "The democratic governance of information societies. A critique to the theory of stakeholders." *Philosophy & Technology* 28.1 (2015): p. 29, doi: 10.1007/s13347-014-0162-y.

of research in Europe[32], taking into account the framework of rights and values whose respect is imperative.

1.2.2 Interdisciplinarity

This study has an interdisciplinary orientation: in order to analyse the legal issues of a field radically transformed by technology, it is crucial to focus on the technologies involved, on how they work and on their specificities.

Among the various taxonomies proposed in the literature regarding the interdisciplinarity of the legal research, this study should be considered closer to the typology of the so-called "basic interdisciplinary research" according to Siems[33]. This category may be defined by van Klink and Taekema as "[...] legal research that uses insights from other disciplines heuristically as a source of inspiration"[34].

Although the core of this dissertation remains legal, other disciplines will not only be inspirational sources, since they are ontologically connected to the topic of the scientific research governance.

In the taxonomy proposed by Klink, the types of research range from a strict monodisciplinarity to an integrated interdisciplinarity. The former is represented by Kelsen's vision, according to which "[...] combining perspectives from different disciplines would lead to a 'methodological syncretism' and is, therefore, 'inadmissible'"[35]. The

---

[32] In the relationship between mimetic and poietic approaches, the aim is to refer more to a poietic approach, considering that: "[P]oietic, as opposed to mimetic science, relies on a conceptual logic of construction that does not start from the system to analyse it in terms of a model, but actually starts from the model (the blueprint) to realise the system. In this case understanding is constructing". In particular: "The more science shifts from a mimetic to a poietic approach to the world the more we shall need a logic of design.", in LUCIANO FLORIDI, "The logic of design as a conceptual logic of information." *Minds and Machines* 27.3 (2017): 495-519, doi: 10.1007/s11023-017-9438-1.

[33] MATHIAS M. SIEMS, "The taxonomy of interdisciplinary legal research: Finding the way out of the desert." *Journal of Commonwealth Law and Legal Education* 7.1 (2009): pp. 5-17.

[34] BART VAN KLINK, SANNE TAEKEMA, "Limits and Possibilities of Interdisciplinary Research into Law.", in STEPHAN KIRSTE, *et al. Interdisciplinary research in jurisprudence and constitutionalism.* (Stuttgart: Steiner, 2012): p. 10.

[35] BART VAN KLINK, SANNE TAEKEMA, *"Limits and Possibilities", op. cit.*, p. 6.

latter evokes John Dewey's pragmatism, "[...] in a theory that addresses questions of scientific truth and inquiry in the context of problem-solving in everyday experience"[36]. In this spectrum of possibilities, this study, adopting a basic and heuristic[37] interdisciplinary approach, intends to draw on other domains of knowledge in addition to law: the research questions are mainly legal, although the investigation to answer these questions involves a crossover with other areas of knowledge[38].

1.2.3 The Case Study: ULHPC

This dissertation, focusing on the relationship between law and technology in scientific research, devotes particular attention to a case study: the ULHPC, namely an HPC platform, hosted by the University of Luxembourg, and used for scientific research purposes.

The premise is the connection that the European institutions make between the field of Open Science and the promotion of cloud computing services and HPC capabilities, evoking the process of technological convergence.

The analysis of the experience of the University of Luxembourg allows to bring out more clearly some aspects that are primarily investigated in the broader perspective of European projects.

The aim of the European Open Science projects is not to build a European research environment from scratch. Rather, the aim is to exploit what already exists across Europe (especially in terms of infrastructures) and align it with certain standards (in terms of values and rights, but also in technological terms), to foster sharing in scientific research.

---

[36] BART VAN KLINK, SANNE TAEKEMA, *"Limits and Possibilities"*, *op. cit.*, p. 2.

[37] BART VAN KLINK, SANNE TAEKEMA, *Law and method.* (Tübingen: Mohr Siebeck, 2011), pp. 10-11.

[38] In Klink's proposed taxonomy, this view of a moderate interdisciplinarity is linked to the perspective of the sociologist Nicklas Luhmann: "Luhmann's system theory seems to offer an attractive middle position between two extremes: Dewey's integrative vision in which disciplines are mixed and mixed up in one unified whole on the one hand and Kelsen's rigorous separation of disciplines on the other.", BART VAN KLINK, SANNE TAEKEMA, "*Limits and Possibilities*", *op. cit.*, p. 7-8.

Accordingly, the ULHPC case study can also be considered as a preliminary investigation of a research infrastructure, among many in Europe, that will be part of the European federated research environment, which is starting to take shape[39].

Hence, some of the considerations developed for this case study may be generalised to many other similar realities, scattered across the European territory[40].

1.3 Outline

The dissertation is structured in seven chapters. After this introduction, each chapter aims to answer a sub-research question, presented in Table 1.1.

Chapter 2 investigates the European governance of scientific research, answering to RQ2. In doing so, it starts by describing the scenario in which this study is set, further deepening the two premises – methodological and institutional – introduced at the beginning of this chapter. Then, starting from the complexity resulting from the proliferation of initiatives, policies and documents issued by European institutions in the context of Open Science, a mapping of the evolution of such policies is outlined, with the aim to provide a consistent conceptual framework.

Then, considering that these Open Science policies did not emerge in a normative vacuum, attention is drawn to the current European legislative framework.

In light of the systematisation of the EU Open Science policies and the analysis of the legislative disciplines involved, the most suitable

[39] The European federated environment for science is described in Section 2.2 and further analysed in Chapter 6.

[40] During the investigation of this dissertation, a comparison has been made with a few other similar initiatives, located across Europe, with the same issues. In particular, the HPC platform of the University of Turin, called "HPC4AI", see: https://hpc4ai.unito.it; and the infrastructure of the Research Center "Area Science Park", in Trieste, which holds an integrated environment of cloud computing and HPC capabilities, called "Orfeo Ecosystem", see: https://www.areasciencepark.it.

approach to the Open Science governance is identified. Between two polarised and extreme positions, i.e., supporters and opponents of Open Science, this dissertation opts for a third way, of legal coordination, which guides the whole study.

Chapter 3 will analyse the evolution of the Open Science paradigm, addressing RQ3.

From the conceptual framework of the evolution of the EU policies on Open Science conducted in Chapter 1, several dimensions of the phenomenon emerge: some of these dimensions have changed over time, others are unclear. Chapter 3 therefore aims to investigate the notion of Open Science and to propose a definition of the Open Science paradigm that fits today's needs, bringing out an all-encompassing outlook, which is still lacking at European level.

After proposing the interpretation of the Open Science paradigm, its foundations are explored, proposing an analysis which places the roots of the Open Science in the framework of fundamental and human rights.

Once the European state of the art of Open Science has been clarified (what it is) and an all-encompassing interpretation of the paradigm has been identified (what it should be), Chapter 3 concludes by shedding light on the possible pitfalls of the Open Science paradigm in its implementation.

Chapter 4 focuses on open research data, tackling RQ4.

The interpretation of the Open Science provided in Chapter 3 identifies research data as one of the key elements of the paradigm. For this reason, Chapter 4 first attempts to define research data and untangle the issues related to it, starting by distinguishing legal issues from ethical ones.

Legal issues are then explored in depth, focusing mainly on the recent European directive on Open Data, highlighting the hurdles that still remain to the effective openness of research data.

Afterwards, the discussion moves to ethical issues, mainly related to the quality of research data.

Finally, looking forward, some considerations are made about the impact of the new proposed regulation by the European Commission on data governance, namely the Data Governance Act (DGA), on the European Open Science framework.

Chapter 5 examines the issues related to the protection of personal data in the Open Science scenario, as represented by RQ5.

At first, it will be clarified which types of data processing activities are generally carried out in the context of scientific research.

Then, after providing an overview of the data protection framework for scientific research, the main challenges are identified and tackled.

Next, a position is taken on the alleged incompatibility between Open Science and data protection: a possible convergence will be argued. Nevertheless, the analysis recognises the persistence of certain barriers to an effective protection of personal data in the research field, trying to identify what actions can and should be taken in the direction of convergence.

Chapter 6 is dedicated to the e-infrastructures of the Open Science, in response to the last RQ6.

The purpose is to understand the role played by digital infrastructures in the Open Science paradigm. In doing so, an initial investigation on the evolution of the research infrastructures is proposed.

After that, in order to narrow down the scope of the research, a specific computational infrastructure is addressed, namely the High Performance Computing (HPC). Since much is happening in HPC domain at the European level, the chapter explores technology *per se*, its evolution in the EU policies and its link with Open Science.

This analysis on research e-infrastructures leads to develop the issue of the interplay between public and private sectors in the field of scientific research. In particular, the consequences of this interaction from a legal point of view are explored, in light of the enactment of the US CLOUD Act.

Finally, the chapter will end by presenting the case study of this dissertation, represented by the HPC facility of the University of Luxembourg, i.e., ULHPC.

Chapter 7 concludes the dissertation by pointing out the main findings, providing answers to the main research question.

Chapter 2

The European Scientific Research Governance

Chapter 2 investigates the scenario within which this study is conducted, analysing various aspects of the governance of scientific research in the European Union.

In Section 2.1, the intention is to deepen the relevance of the topic under investigation, focusing on the premises of this dissertation, i.e., the methodological premise (Section 2.1.1) and the institutional one (Section 2.1.2).

Then, in Section 2.2, the attention is on the European approach to scientific research, mapping the evolution of the multiple projects related to Open Science.

Starting from the assumption that the recent European projects on scientific research do not arise in a normative vacuum, Section 2.3 is devoted to an overview of the European legislative frameworks related to this domain.

Clarified the EU projects on scientific research and the different legislative frameworks involved, possible approaches to the Open Science governance are investigated in Section 2.4. The different stances are summarised in three approaches, represented by: (i) the position of the supporters of an *ad hoc* law for Open Science (Section 2.4.1); (ii) the opponents who argue the incompatibility between the EU legal framework and the Open Science (Section 1.4.2); and (iii) a suggested alternative way (Section 2.4.3).

Finally, Section 2.6 concludes the analysis on the European scientific research governance with some final remarks.

2.1 The Scenario: Between Science, Technology and, Law

This dissertation should be placed in a scenario marked by the interaction of science, technology, and law. This scenario is described starting from two premises underlying this study[1]. The first premise, the methodological one, specifically concerns the impact of technology on the field of science, in its various facets; while the second, defined as institutional, concerns the reaction of institutions to this changing science.

2.1.1 Methodological Premise

The philosopher James H. Moor, in 1975, examining what was meant by computer ethics, presented an interesting perspective on how technology impacts on human activities[2]. According to Moor, initially, when we are in the early stages of introducing technology into a specific human activity, we wonder how effectively the application of technology improves the activity itself: "How well does a computer do such and such activities?"[3].

The situation changes in the next stage, which Moor calls "the technological permeation stage", where technology has become pervasive, perceived as normal. In this phase, technology profoundly transforms the activity of which it has become an integral part. Here, we tend to wonder about the very nature of that activity because technology has now ontologically mutated it: "What is the nature and value of such and such an activity?"[4].

---

[1] These premises (i.e., methodological and institutional) have already been introduced. See: Chapter 1.
[2] JAMES H. MOOR, "What is computer ethics?." *Metaphilosophy* 16.4 (1985): 266-275. Consider that "[A]lthough some trace the start much earlier, many scholars cite James Moor's 1985 paper as the beginning of computer ethics as a sub-discipline of applied ethics." in MARIAROSARIA TADDEO, KEITH MILLER, "Ethics and Information Technologies: History and Themes of a Research Field.", in MARIAROSARIA TADDEO, KEITH MILLER (eds.), *The Ethics of Information Technologies* (Abingdon: Routledge, 2020).
[3] JAMES H. MOOR, "*What is computer ethics?.", op. cit.*, p. 271.
[4] *Ibid.*

Accordingly, something similar is happening to the link between science and technology of our time, i.e., the Information and Communication Technologies (ICTs). The impact is twofold: on the one hand, it changes the way of doing science; on the other hand, it changes the way of wondering about science.

2.1.1.1 The Way of Doing Science

As regards the way of conducting science, several activities have been radically transformed by the use and adoption of new technologies. The most affected aspects can be summarised as follow: (i) the daily tasks; (ii) the research methodology; (iii) the research cycle; and finally (iv) the research performance.

(i) Daily tasks. Chiefly, even the most basic and daily operations of the scientific research process have been totally transformed[5]. Today, it is often no longer necessary to physically go to a library to carry out a literature review, taking advantage of the digitisation of publications through databases[6]; or, for instance, it is no longer necessary to create datasets on which testing models or theories, being able to draw on an

---

[5] On the philosophical analysis of the impact of new technologies on our daily practices, see: MARIAROSARIA TADDEO, "Philosophy and computing in Information Societies." *Minds & Machines* 26 (2016): 203-204, doi: 10.1007/s11023-016-9400-7. The Author identifies all the issues related to this aspect as belonging to the conceptual axis of the investigation between philosophy and computing (where the other Cartesian axis is the methodological axis): "Philosophical enquiries located along the contextual axis face new, pressing problems, concerning the way we produce scientific knowledge (philosophy of science), the nature of this knowledge (epistemology), and of the tools we use to produce it (philosophy of computing, philosophy of mathematics, and philosophy of statistics), as well as their ethical implications (ethics). It is with the contextual axis in the picture that the area of philosophy and computing is fully delineated and its importance becomes evident.", p. 203.

[6] "The scientific researcher working on a new project hardly does any of the things that used to be the standard for this kind of activity: going to the library, consulting books, consulting journals, opening dozens of filing folders full of handwritten forms." in LUCIANO GALLINO, *Tecnologia e democrazia. Conoscenze tecniche e scientifiche come beni pubblici*. (Torino: Einaudi, 2014), p. 423, EPub. [Translation from the Italian original text]. In addition, also communication and collaboration between scientists and researchers is also facilitated by ICTs. On this issue, and specifically on the existence of persistent barriers to communication in the information society, see: MASSIMO DURANTE, "How to cross boundaries in the information society: vulnerability, responsiveness, and accountability.", *Acm Sigcas Computers and Society* 43.1 (2013): 9-21.

immense amount of real data, acquired and elaborated with data analytics techniques[7].

(ii) Research Methodology. Then, ICTs also modified the choices made by researchers when determining the methodology for conducting a project. In other words, digital technologies should not be considered only tools because they truly contribute to shape the research itself. The use of one technology over another has a much greater impact than is generally thought. As claimed by the philosopher Sabina Leonelli: "Choices that may seem purely technical – what form of probabilistic calculation to use, what kind of classifications to rely on – have implications for how the resulting knowledge can transform society. And although the scientists involved in Big Data analysis are primarily responsible for the decisions they make, their choices cannot be made outside the social context in which knowledge is produced and used"[8]. Leonelli, then, emphasises the extent of the consequences: "Within a democratic society, this implies a commitment to the pursuit of dialogue and discussion between researchers and other social groups whose experience of certain situations puts them in a position to contribute decisively to the evaluation of the assumptions and choices made in the production, dissemination, and interpretation of data"[9].

---

[7] Consider that, for a long time, the most pioneering research conducted in the field of machine learning required the actual creation of methods for generating data, in order to test the machine learning systems. See, for instance: GIULIA PAGALLO, DAVID HAUSSLER, "Boolean feature discovery in empirical learning." *Machine learning* 5.1 (1990): 71-99.

[8] SABINA LEONELLI, *La ricerca scientifica nell'era dei big data. Cinque modi in cui i Big Data danneggiano la scienza, e come salvarla.* (Milano: Meltemi Editore, 2018), p. 114. [Translation from the Italian original text]. Similarly, on this issue, see: BEN VAN CALSTER, *et al.*, "Predictive analytics in health care: how can we know it works?" *Journal of the American Medical Informatics Association*, 26.12 (2019):1651–1654, doi:10.1093/jamia/ocz130. Here, the Authors go a bit further than what was argued in our study regarding an actual impact of the method or technology used on the outcome of the research project. Indeed, they insist on the role of external validation to assess the actual validity of an analysis based on undisclosed predictive algorithms. This aspect is related to the topic of reproducibility of the scientific research, which will be discussed in Section 4.1.1, in defining the notion of "research data".

[9] SABINA LEONELLI, *La ricerca scientifica nell'era dei big data., op. cit.,* p. 114.

(iii) Research Cycle. The ICTs' impact on science cannot be interpreted only as the use of new or different tools: it is the representation of something deeper. Considering the impact of ICTs on society, the philosopher Luciano Floridi has referred to a revolution, specifically the so-called "Forth Revolution": a disruption of such a scale that it has led to examine how our very identity has changed[10]. Accordingly, the impact of ICTs on the specific aspect of our society represented by science has produced and is producing effects marked and intense, which require a more aware investigation: this transformation must be interpreted as a real paradigm shift, represented by the Open Science. This phenomenon describes a metamorphosis affecting the entire scientific research cycle – starting with pure mathematics[11] down to the dissemination – identifying an actual cultural transition[12].

(iv) Research Performance. The use of ICTs in the field of scientific research has led to remarkable progress. In the last thirty years there have been amazing developments in many and different fields of knowledge, e.g., genetic engineering, astronomy, medicine, etc[13].

---

[10] According to the Philosopher, in fact, human beings should be represented as informational entities that interacts with other informational organisms, whether human or artificial: human and artificial agents are both informational organisms, the so-called '*inforg*'. They operate, in a space, the so-called '*infosphere'*, which does not see a boundary between online and offline, making the experience '*onlife*', in LUCIANO FLORIDI, *The fourth revolution: How the infosphere is reshaping human reality.* (Oxford: OUP, 2014), pp. 25-58 EPub; LUCIANO FLORIDI, *The onlife manifesto: Being human in a hyperconnected era.* (Cham: Springer Nature, 2015), pp. 7-16. On the application of the notion of '*inforg*' to scientific research, see: Chapter 6 and also Section 7.1.8, describing the so-called "*inforg of science*".

[11] While ICTs have generally changed the way science is done, digital technologies have been truly disruptive in the field of pure mathematics. Consider the strand of research ranging from the Hungarian philosopher Imre Lakatos to the mathematician Gregory Chaitin, which investigates the epistemological aspect. Among others, see: IMRE LAKATOS, *Proofs and refutations: The logic of mathematical discovery.* (Cambridge: Cambridge university press, 2015); GREGORY J. CHAITIN, "Algorithmic information theory." *IBM journal of research and development* 21.4 (1977): 350-359.

[12] SÖNKE BARTLING, SASCHA FRIESIKE, *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing.* (Cham: Springer Nature, 2014), 3. Next chapter, specifically Section 3.2, discusses the relevance of the concept of "research cycle" or "research process" in the identification of the definition of the Open Science paradigm today.

[13] The notion of "scientific progress" is much debated. Here, the aim is merely to refer to a progressive advancement of knowledge due to technical developments. For a more detailed discussion of the notion

Indeed, the production of the vaccine against the SARS-CoV-2 virus, mentioned in the introduction to this analysis, is an outstanding example.

Now, based on the perspective provided by the philosopher James Moor on how technology impacts on human activities, the focus shifts from the way science is conducted to the way science is perceived in itself.

2.1.1.2 The Way of Wondering About Science

As regards the way of wondering about science, considerable changes have to be reported. Two aspects are crucial: the first relates to the role of the scientific method; the second to the interaction between science and society.

First, the massive use of ICTs has even led to the belief that the scientific method is useless, made obsolete by the great amount of data, the so-called Big Data[14]. The historian Yuval Harari describes this trend in terms of "Dataism", stating that: "[…] Dataism inverts the traditional pyramid of learning. Hitherto, data was seen as only the first step in a long chain of intellectual activity. Humans were supposed to distil data into information, information into knowledge, and knowledge into wisdom. However, Dataists believe that humans can no longer cope with the immense flows of data, hence they cannot distil data into information, let alone into knowledge or wisdom. The work of processing data should therefore be entrusted to electronic algorithms, whose capacity far exceeds that of the human brain. In practice, this

---

of "scientific progress", see: ILKKA NIINILUOTO, "Scientific Progress", in EDWARD N. ZALTA (ed.) *The Stanford Encyclopedia of Philosophy* (2019), https://plato.stanford.edu/archives/win2019/entries/scientific-progress/.

[14] CHIRS ANDERSON, "The end of theory: The data deluge makes the scientific method obsolete", *Wired magazine* (2008), https://www.wired.com/2008/06/pb-theory/. Here, the intention is to refer, without going into detail, to the phenomenon of data science, to the debate on the framing of data science as a real discipline, and to the foundational issues in the philosophy of data science currently ongoing. In this regard, see: JULES DESAI, DAVID WATSON, VINCENT WANG, MARIAROSARIA TADDEO, LUCIANO FLORIDI, "The epistemological foundations of data science: a critical analysis." SSRN (2022), doi: 10.2139/ssrn.4008316.

means that Dataists are skeptical about human knowledge and wisdom, and prefer to put their trust in Big Data and computer algorithms"[15].

As argued by Luciano Floridi, the situation slightly differs, considering that data do not speak for themselves: it is essential what he calls small-scale patterns, able to extract information and knowledge from huge amount of blind data[16].

This statement implies a distinction between data, information, and knowledge, conceived by Floridi, in the construction of the ethics of information[17]. In this study, the intention is not to delve into the philosophical thought of Floridi, rather it is relevant to bear in mind the distinction between the notion of data, information, and knowledge. In doing so, it is worth to mention the interpretation provided by the philosopher of law Massimo Durante, about these three concepts of Floridi's ethics[18]:

- data can be considered as a mere point of discontinuity, a lack of uniformity;
- information, instead, is built on the concept of data and, here, is understood as data that hold meaning[19];

---

[15] YUVAL NOAH HARARI, *Homo Deus: A brief history of tomorrow.* (New York: Harper Collins, 2016), p. 748, EPub.

[16] "The real, epistemological problem with big data is small patterns. […] Small patterns matter because they represent the new frontier of competition, from science to business, from governance to social policies. In a Baconian open market of ideas, if someone else can exploit them earlier and more successfully than you do, you might be out of business soon […] or miss a fundamental discovery.", in LUCIANO FLORIDI, "Big data and their epistemological challenge." *Philosophy & Technology* 25.4 (2012): 436, doi: 10.1007/s13347-012-0093-4.

[17] LUCIANO FLORIDI, *The philosophy of information.* (Oxford: OUP 2011), pp. 278-279; MASSIMO DURANTE, "Dealing with legal conflicts in the information society. An informational understanding of balancing competing interests." *Philosophy & Technology* 26.4 (2013): 437-457, doi: 10.1007/s13347-013-0105-z.

[18] MASSIMO DURANTE, *Computational Power: The Impact of ICT on Law, Society and Knowledge.* (New York: Routledge, 2021), pp. 50-75. See, also: MASSIMO DURANTE, *Ethics, Law and the Politics of Information: A Guide to the Philosophy of Luciano Floridi.* (Dordrecht: Springer, 2017), pp. 3-19, and also 103-116, doi: 10.1007/978-94-024-1150-8.

[19] In Luciano Floridi's framework of information ethics, several different profiles of information coexist, in addition to the one considered here, namely that of semantic information. On this aspect, see: LUCIANO FLORIDI, *Information: A very short introduction.* (Oxford: OUP 2011), pp. 31-34.

- and finally, starting from the ability to wonder the right questions, knowledge is understood as that mesh of questions and answers that stems from this ability to inquire (rather than a mere stratification of information[20]).

It is precisely from this interpretation of knowledge as the capacity to wonder and inquire that the role of researchers, as a community, has made explicit: this mesh of questions and answers is enriched precisely by their continuous dialogue and the transparency that characterises this dialogue.

Wondering about the nature and value of research activity as profoundly transformed by ICTs, attention should also be drawn to the second crucial aspect introduced at the beginning, i.e., the interaction between science and society. The impact of ICTs on scientific research is not an issue whose interest can be limited to the narrow circle of the scientific community, but necessarily engages the society as a whole. Assuming that the aim of science is the development of knowledge as a precondition for further improvements of society, then a wide and aware debate about scientific research and its governance must necessarily involve the entire society.

The ontologically collective nature of knowledge has been emphasised in the public debate precisely because of the COVID-19 pandemic: after a long time, the public at large seemed to have a renewed interest in the mechanisms proper to the functioning of science.

The relevance of societal engagement in science is also confirmed by the European institutions, which seem to be oriented towards a greater consideration of civil society in the management of research data. Think, for instance, about the new concept of data altruism, as envisaged in the proposed regulation of the Data Governance Act of the

[20] JANNIS KALLINIKOS, "The making of ephemeria: on the shortening life spans of information." *International Journal of Interdisciplinary Social Sciences* 4.3 (2009): 227-236.

European Commission[21]; or the increased relevance of the citizens participation in science as highlighted in the Horizon Europe, the new European funding programme for innovation and research[22].

The first premise of this dissertation, the methodological one, concerns the impact of technology on science. This impact has been investigated by adopting the perspective suggested by the philosopher Moor, taking into account two dimensions: (i) how technology has changed the human activity (the way of doing science); (ii) how the technology has changed the way of wondering about that human activity (the way of wondering about science).

Next section will be devoted to the second premise of this dissertation, the institutional one.

2.1.2 Institutional Premise

The institutional premise consists in the commitment that the European institutions have assumed in relation to the Open Science paradigm. The *corpus* of policies issued in recent years, which are mapped out in Section 2.2, clearly expresses the intention of the institutions to support this new paradigm of science, in order to foster scientific research in Europe.

Considering the geopolitical point of view, this dissertation will adopt the European perspective, for two main reasons.

---

[21] Recital 35 of the Data Governance Act states: "This Regulation aims at contributing to the emergence of pools of data made available on the basis of data altruism that have a sufficient size in order to enable data analytics and machine learning, including across borders in the Union". The participation of citizens in the management (or voluntary transfer) of data seems to be a consequence of the model of participatory democracy which, especially in certain areas (e.g. environmental policies) has been quite successful in recent years. This debate is particularly well developed in Germany where, recently, citizen participation has also been experimented in the field of foreign policy. For instance, see: HANNA PFEIFER, CHRISTIAN OPITZ, ANNA GEIS, "Deliberating ForeignPolicy: Perceptions and Effects of Citizen Participation in Germany." *German Politics* (2020): 1-18, doi:10.1080/09644008.2020.1786058. A further examination on the concept of data altruism will be provided in Section. 2.3.4, and also in Section 4.4.1.

[22] On this point, see: Section 2.2.6.

The first reason is that only at European level is possible to make concrete any effort that, instead, at a national level would lose effectiveness. Promoting scientific research guided by the use of new technologies can be an asset of competitiveness for Europe, in a global scenario that involves powers such as the United States, Russia and China, making enormous investments in this direction[23].

Traditionally, in the field of scientific research, in the relationship between the European Union and Member States, the former has always played a role of support, coordination, and complementarity with respect to the action of each Member State[24]. This dynamic emerges from the interpretation of the Article 6 of the Treaty on the Functioning of the European Union (TFEU)[25] and Title XIX of the TFEU, entitled "Research and Technological Development and Space".

Although Member States have historically been reluctant to grant sovereignty and powers in this field[26], today the trend seems to be changing. The traditional attitude has been diminishing over the decades. The transformation of the relationship between EU and Member States is believed to be the result of what the philosopher

[23] Consider that this awareness is even clearly expressed by the European institutions. Concerning the High Performance Computing (HPC), for instance, see: "No single Member State alone has the financial resources to develop the necessary HPC ecosystem, in competitive time frames with respect to the US, Japan or China.", in European Commission, *European Cloud Initiative - Building a competitive data and knowledge economy in Europe*, COM/2016/178 final, p. 5, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016DC0178;

[24] On the competences of the European Union, *ex multiis*, see: MARIA LUISA MANIS, "The processing of personal data in the context of scientific research. The new regime under the EU-GDPR." *BioLaw Journal* 11.3 (2017): 325-354, doi:10.15168/2284-4503-259. On the distribution of competences between the European Union and the Member States, see: Section 5.2.2.

[25] Consolidated version of the Treaty on the Functioning of the European Union (TFEU), OJ C 326, 26.10.2012, p. 47–390, http://data.europa.eu/eli/treaty/tfeu_2012/oj.

[26] For an overview of the European integration process in higher education and research, see: ERIC BEERKENS, "The emergence and institutionalisation of the European higher education and research area." *European journal of education* 43.4 (2008): 407-425, doi:10.1111/j.1465-3435.2008.00371.x. According to the Author, initially: "Due to the national sensitivity of education and some types of scientific research, countries are very hesitant to give up any sovereignty in these fields, and thus they will try to tackle issues with a European dimension in an inter-governmental manner, leaving European educational and research policy in the hands of the lowest common denominator in this bargaining process".

George Steiner called "a certain idea of Europe"[27], according to which the EU institutions are committed to empower "[…] researchers to cooperate freely across borders and at enabling undertakings to exploit the internal market potential to the full, in particular through the opening-up of national public contracts, the definition of common standards and the removal of legal and fiscal obstacles to that cooperation"[28].

The second reason for adopting a European perspective is that the European Union, to date, appears a pioneer in identifying policies to support the Open Science. The European Union has the opportunity to set a benchmark[29] for the rest of the world in establishing a model of scientific research that is open, collaborative but strongly anchored to European values, safeguarding the fundamental and human rights of individuals. It is precisely the strong link with the framework of fundamental and human rights and European values that may represent an extra value that – in an ideal, or perhaps idealistic vision – should characterise the architecture of the European Union.

A relevant example of how the European Union is a reference point for the rest of the world in the area under investigation is the recent "shared appeal on Open Science by UNESCO, WHO, CERN and the Office of the United Nations High Commissioner for Human Rights". This appeal concerns the definition of the first recommendation on Open Science by UNESCO, which has been released in November 2021, in the General Annual Conference in Paris. The first draft, released on October 27, 2020, had the merit of bringing to the forefront, at the

[27] GEORGE STEINER, *The idea of Europe. An essay*. (New York: Abrams, 2015).

[28] The Article 179 TFEU. An emblematic example is represented by the degree of flexibility related to the mobility of researchers: they can move from one European Country to another, for research purposes, without the bureaucratic burden that has to be faced by non-European colleagues.

[29] The opportunity (or ability) of the European institutions to be a reference in this area evokes the concept of the so-called "Brussels effect", see: ANU BRADFORD, "The Brussels effect." *Northwestern University Law Review* 107 (2012): p. 1. In contrast to Bradfort's assertions, this dissertation does not refer to the power of European legislation to impose itself unilaterally, but rather to an ability to direct, or, at least, have an influence on Open Science policies at a global level.

international level, the debate on Open Science: it conveyed many of the concepts which have been the core of European interventions in recent years[30].

The "Strategic Research and Innovation Agenda" (SRIA, hereinafter), released by the EOSC Executive Board[31], in its version 1.0, in February 2021, emphasised the leading role of EU on Open Science, arguing that: "[W]hile other regions in the world have launched their own efforts, none of them have done it at the scale on which Europe has invested"[32].

The European institutions, in order to ensure the fulfilment of their commitment, are discussing the identification of a model of governance – understood as a set of tools and methods of regulation – capable of promoting scientific research as much as possible[33].

Illustrating the first premise, it was argued that new technologies make possible previously unknown operations in the scientific field, in particular with reference to sharing and reuse of huge amounts of data. Consider that the sharing and reuse of data is an asset for the development of research and its economic consequences – as has been repeatedly stated by the European institutions[34]. Hence, it is crucial to have a legal framework that is sufficiently flexible, realistic, but also ambitious. Such a framework should be flexible in order to be able to

[30] See the following Section 2.2.

[31] The EOSC Executive Board is one of the bodies managing the European Open Science Cloud (EOSC) project, the central initiative of the European institutions on Open Science, which is analysed in Section 2.2.4.

[32] EOSC Executive Board, "Strategic Research and Innovation Agenda (SRIA)", Version 1.0, February 2021, https://www.eoscsecretariat.eu/; and, also: https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf.

[33] The European scientific research governance can also be a way to promote several fundamental rights related to scientific research, such as the free movement of knowledge, according to the Article 27 of the Universal Declaration of Human Rights, the academic freedom *ex* the Article 13 and the right to education *ex* the Article 14 of the Charter of Fundamental Rights of the European Union and, in general, innovation and development. On the foundations of Open Science in the framework of fundamental and human rights, see: Section 3.3.

[34] European Commission, *A European strategy for data*, COM/2020/66 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066.

allow operations that are technically feasible, without excessive bureaucratic burdens or uncertainties. It is also required to be realistic, i.e., practically implementable, taking into account the state of the art of technology and bearing in mind the process of technological convergence. Finally, this legal framework should have the ambition of revitalising scientific research in Europe by promoting collaboration as much as possible.

In light of the first methodological premise (i.e., the transformation occurred in science for the introduction of ICTs) and the second institutional premise (i.e., the commitment of European institutions in supporting the Open Science paradigm), next section analyses the current European framework for scientific research, mapping the various projects that are currently underway.

2.2 European Projects: From Cloud Computing to the European Open Science Cloud

The analysis proposed in these sections aims to illustrate the complex tangle of strategies and initiatives on Open Science at European level, from 2015 to date. This systematisation is based on various grounds.

First, it aims to provide a map of European policy developments, by dividing it into different phases.

Second, for each phase, this systematisation sheds light, on the one hand, on the elements recognised as barriers or obstacles to the evolution of the Open Science paradigm and, on the other hand, on the elements of the concrete development of the European policy making.

Further, the complexity of the scenario also hinges on the fact that these policies aim to exploit the phenomenon of technological convergence fruitful. These projects jointly consider the data management, the infrastructures, and also the aspects related to the computational power. This mapping is, therefore, useful also to understand the relationship between these various facets, avoiding confusing concepts[35].

An important condition for approaching the analysis of the *corpus* of policies under consideration is that the lawmaker, promoting Open Science, should not prioritise either any specific scientific domain nor any technical or technological way of ensuring this openness. Peters, already in 2010, in his analysis on the philosophy of Open Science, exploring the relationship between "the technologies of openness" (referring in particular to the development of web 2.0) and science, argues that although science is already traditionally open "[...] the internet/web potentially increases the openness"[36] in general, not embedding it in the development of a particular type of technology. In other words, it is relevant to analyse the developments of the European Open Science policies bearing in mind that their actual implementation should not depend on the choice of specific technologies. Rather, this implementation should be characterised, from a legal and policy-making point of view, by a technologically neutral approach[37].

Thus, having clarified the relevance of this systematisation and the condition of technological neutrality that lawmakers should adopt when approaching this matter, it is now time to proceed with the analysis of the origin of the concept of Open Science within the European architecture. This study identifies a distinction between a substantial origin of Open Science in EU (Section 2.2.1), and a formal one (Section 2.2.2).

---

[35] This refers, for instance, to the confusion created by the expression "European Open Science Cloud", to indicate a federated environment that is not a cloud computing system. This aspect is clarified in Section 2.2.2 and Section 2.2.3.

[36] MICHAEL A. PETERS, "On the philosophy of open science." *Review of Contemporary Philosophy* 9 (2010): p. 126.

[37] The concept of technological neutrality of law has been widely discussed by many scholars. Considering the topic under investigation see: UGO PAGALLO, *Il diritto nell'età dell'informazione: il riposizionamento tecnologico degli ordinamenti giuridici tra complessità sociale, lotta per il potere e tutela dei diritti*. (Torino: Giappichelli Editore, 2014), 142-145; CHRIS REED, *Making laws for cyberspace*. (Oxford: Oxford University Press, 2012), 160-201.

2.2.1 The Fifth European Freedom: The Free Movement of Knowledge

The origin of the concept of Open Science within the European architecture is connected to the Communication n. 192 of 2015, in which the European Commission sets out the "Digital Single Market Strategy for Europe"[38]. This Communication is represented as the formal starting point to map out the European initiatives and projects in the field of scientific research, directed towards digital innovation.

However, it is deemed that the substantial origin of the concept of Open Science in the European architecture dates back a little further. In fact, the intention to remove technical-legislative barriers to promote the maximum exploitation of knowledge in Europe should be connected to the so-called "fifth European freedom", the freedom of circulation of scientific results and knowledge, established in 2008, in the Council of Ljubljana[39]. The EU free movement of knowledge has been made concrete with the reinforcement of the European Research Area (ERA): a space of free movement for scientific research in Europe[40]. In fact, the concept of knowledge sharing was boosted by the European Commission, in 2012, in the Communication entitled "A reinforced European Research Area Partnership for Excellence and Growth"[41]. In this Communication, without any explicit mention of the concept of Open Science, was set the goal of strengthening the ERA. The free movement guaranteed by the ERA was twofold: first, it aimed to ensure an increasing movement of research personnel among

[38] European Commission, *A Digital Single Market Strategy for Europe*, COM/2015/192 final, ELI:https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015DC0192.

[39] The phrase was coined by Janez Potočnik, European Commissioner for Research from 2004 to 2010, and introduced at the Council of Europe in Ljubljana in 2008.

[40] The European Research Area (ERA) was introduced for the first time in a Communication of the European Commission, in 2000, see: European Commission, *Towards a European research area*, COM/2000/0006 final, ELI: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:52000DC0006. However, the intention here is to bring out in this discussion the relation between the ERA and the fifth European freedom of circulation of knowledge, of which it is considered a true embodiment.

[41] European Commission, *A Reinforced European Research Area Partnership for Excellence and Growth*, COM/2012/392 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012DC0392.

Member States; and, second, it promoted greater circulation of knowledge through the medium of the digital tools. In particular, the European Commission declared: "[R]esearch and innovation benefit from scientists, research institutions, business and citizens accessing, sharing and using existing scientific knowledge and the possibility to express timely expectations or concerns on such activities. [...] As most knowledge creation and transfer uses digital means, all barriers preventing seamless online access to digital research services for collaboration, computing and accessing scientific information (e-Science) and to e-infrastructures must also be removed by promoting a digital ERA"[42]. Thus, although the starting point for the Open Science policies is formally traced back to 2015, with the establishment of the Digital Single Market, it is worth emphasising that the concept of free circulation of knowledge, which characterises the Open Science, was already part of previous European policies. In other words, the establishment and the reinforcement of the ERA is a real milestone in the European process of scientific knowledge sharing[43].

Now, having clarified the substantial origin of the concept of Open Science within the European architecture – in the fifth European freedom of circulation and in its practical realisation with the ERA – it is time to proceed with the analysis of the formal starting point of the European *corpus* of policies of Open Science.

2.2.2 Digital Single Market Strategy

The Communication n. 192 of 2015, setting out the "Digital Single Market Strategy for Europe"[44], is considered to be the formal start[45] of

---

[42] COM/2012/392, p. 13.

[43] The fact that the establishment of the ERA has been a milestone in the development of scientific research in Europe is confirmed by the release of a new Communication of the European Commission, concerning the ERA: European Commission, *A new ERA for Research and Innovation*, COM/2020/628 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:628:FIN. This document is analysed in Section 2.2.5, as part of the set of European initiatives that characterise the year 2020 as a turning point for Open Science.

[44] COM/2015/192, p. 1.

[45] It should be noted that the foundations of the Digital Single Market Strategy can be traced back to

the European Commission's Open Science policies: here, the concept of Open Science is explicitly mentioned, and it is part of a concrete strategy.

The Commission aimed to outline a clear European policy for a digital development, with a cross-cutting approach, which would go beyond the boundaries of individual Member States. Each Member State could only provide too limited solutions, hardly matched with the global nature of digital development. This strategy hinges on how ICTs are reshaping the world and every aspect of our lives.

The main pillars of this Strategy were: "(1) better access for consumers and business to online goods and services across Europe; (2) creating the right conditions for digital networks and services to flourish; (3) maximising the growth potential of our European Digital Economy".

In order to facilitate implementation of these broad and general goals, sixteen specific initiatives were outlined[46]. By setting out the intention to maximise the potential growth of the digital economy, the European Commission drew an explicit connection between Cloud services, High Performance Computing (HPC) and Open Science, which will mark the projects carried out in the following years (and, partially, still being implemented). The Commission declared that: "Big

---

the year 2010, in: European Commission, *A digital agenda for Europe*, COM/2010/245 final, ELI: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52010DC0245. In this Communication the European Commission introduced the concept of "a vibrant digital single market", claiming that "[I]t is time for a new single market to deliver the benefits of the digital era".

[46] The initiatives are about: (1) the establishment of rules to make cross-border e-commerce easier; (2) the review of the Regulation on Consumer Protection Cooperation; (3) a creation of more efficient parcel delivery; (4) the end of unjustified geo-blocking; (5) to identification of potential competition concerns affecting European e-commerce markets; (6) the review of the European Copyright Law; (7) the review of the Satellite and Cable Directive; (8) the reduction of the administrative burden businesses face from different VAT regimes; (9) the overhaul of the EU telecoms rules; (10) the review of the audiovisual media framework; (11) the analysis of the online platforms role's in the market; (12) the review of EU data protection rules and the e-Privacy Directive; (13) the establishment of a partnership with the industry on cybersecurity for online network security; (14) the proposal of a "European free flow of data initiative"; (15) the definition of priorities for standards and interoperability; (16) the establishment of an inclusive digital society and a new e-government action plan.

data, cloud services and the Internet of things are central to the EU's competitiveness. Data is often considered as a catalyst for economic growth, innovation and digitalisation [...]. Big data and High Performance Computing are also changing the way research is performed and knowledge is shared, as part of a transition towards a more efficient and responsive Open Science"[47]. Among the sixteen initiatives outlined with the intention of putting into practice the three pillars expressed in the Digital Single Market Strategy, the fourteenth regards a "European free flow of data initiative". This initiative was meant to promote the free movement of data within the European Union. The press release issued by the European Commission on May 6, 2015, by then President Junker, announced: "The Commission will: [...] [P]ropose a 'European free flow of data initiative' to promote the free movement of data in the European Union. Sometimes new services are hampered by restrictions on where data is located or on data access – restrictions which often do not have anything to do with protecting personal data. This new initiative will tackle those restrictions and so encourage innovation. The Commission will also launch a European Cloud Initiative covering certification of cloud services, the switching of cloud service providers and a 'research cloud'"[48].

With the expression "research cloud", the European Commission was referring, for the first time, to the project that would have been launched two years later, known as the "European Open Science Cloud" (EOSC, hereinafter). The EOSC is an essential part of the European projects on Open Science, and it will be analysed in detail in next section.

2.2.3 European Cloud Initiative

In 2016, the European Commission presented the European Cloud Initiative, a roadmap describing the European strategy aimed to

[47] COM/2015/192, p. 14.

[48] https://ec.europa.eu/commission/presscorner/detail/en/IP_15_4919.

maximise the potential of cloud computing services. The goal was to achieve the optimal results from the management and analysis of data, with particular attention to data resulting from scientific research, to strengthen innovation, since the European Union was defined as "[...] the largest producer of scientific knowledge in the world"[49].

The European Cloud Initiative was divided into two parts: the first part outlined five reasons why EU was not yet fully exploiting the potential of data; the second part, in light of the identified barriers, was dedicated to the solutions proposed by the Commission to achieve this goal.

The first part, dedicated to the obstacles to the maximisation of benefits, identified five reasons: (i) data coming from publicly funded research not always open; (ii) lack of interoperability; (iii) fragmentation; (iv) lack of a world-class High Performance Computing infrastructure to process data; (v) lack of advanced analytics techniques for the reuse of research data[50].

These five barriers are further specified below.

(i) The closure of data resulting from publicly funded scientific research was recognised as a barrier to innovation. This closure was identified as a failure of investment. The European Commission indicated three elements as main causes of the lack of openness: the lack of awareness of the value of the data by researchers and academia; the lack of "clear structure of incentives and rewards"[51] for researchers in committing efforts to sharing data; and the lack of "clear legal basis"[52] for sharing. As regards the first cause, i.e., the lack of awareness of researchers, it is hard to agree, considering that there is no evidence about it. While the other two causes are certainly problematic, partially still persisting today[53].

---

[49] COM/178/2016, p. 2.
[50] COM/2016/178, pp. 3-5.
[51] COM/2016/178, p. 4.
[52] *Ibid.*
[53] "Partially" because the introduction of the new European Open Data discipline has helped to change

(ii) The lack of interoperability is certainly a problematic issue from a technical point of view. Besides, it has been recently confirmed by the SRIA document[54]. In addition to the technical perspective, also another type of interoperability has been identified, related to sharing of research data: the so-called "legal interoperability". This notion is rather vague and will be investigated later[55].

(iii) The fragmentation referred to by the European Commission in the European Cloud Initiative is infrastructural: "Data infrastructures are split by scientific and economic domains, by countries and by governance models"[56]. Indeed, fragmentation is an obstacle that arises also in other contexts, first of all the legal one[57]. However, fragmentation in research infrastructures is not necessarily a barrier to research development, as will be investigated below[58].

(iv) The lack of HPC infrastructures and (v) the lack of advanced data analysis techniques are two aspects related to the technological convergence process already mentioned above[59]. In fact, the Commission here reported a profound weakness of the European Union in technological development, which still persists despite the excellent research centres and universities, and the considerable scientific knowledge[60].

---

the scenario. This aspect is then further developed in Chapter 3.

[54] The SRIA document – already mentioned in Section 2.1.2 – is the Strategic Research and Innovation Agenda (SRIA, hereinafter), released by the EOSC Executive Board, in its version 1.0, in February 2021. It states that: "[A]nother barrier is the lack of interoperability across datasets, resulting in a fragmented data landscape". See: SRIA, p. 148.

[55] See: Section 4.2.1.

[56] COM/2016/178, p. 4.

[57] For instance, think about the fragmentation of the discipline of the protection of personal data in the field of scientific research. This aspect is further explored in Section 5.2.2 and Section 5.5.1.

[58] See: Section 6.1 on the role of e-infrastructures in science, but also Section 6.4.3, describing the findings of the legal analysis of the case study ULHPC.

[59] See: Section 2.2.

[60] The European approach to technological development is investigated in Section 2.2.5 and described in terms of the "The von der Leyen Doctrine"; it will then be addressed in Section 6.2.3, with specific regard to the HPC.

The second part of the European Cloud Initiative concerned the solutions proposed by the Commission. They are represented by the creation of (i) the European Open Science Cloud (EOSC) and (ii) the European Data Infrastructure (EDI).

(i) The EOSC, i.e. the European Open Science Cloud, is a federated and trusted environment, which aims to facilitate as much as possible the sharing and reuse of scientific research, in terms of research data, research results (i.e. publications), and services, for the benefit – mainly – of European researchers, universities and research centers. This openness should not be understood as an indiscriminate sharing, but it follows the formula "as open as possible, as closed as necessary", in order to ensure the fairest balance of opposing interests at stake. To have an effective data sharing, data must be rigorous from a technical and structural point of view, therefore it must follow the so-called "FAIR Data Principles": FAIR is an acronym that stands for findable, accessible, interoperable, and reusable[61]. According to Budroni *et. al.*[62] the EOSC can be represented as a sort of Internet of FAIR data and services, to make scientific research accessible, in Europe, "[...] under the same terms of use and distribution"[63].

(ii) The European Data Infrastructure (EDI) is the infrastructural architecture, to be placed at the basis of the EOSC, starting from the assumption that "Europe needs integrated world-class HPC capability, high-speed connectivity and leading-edge data and software services for its scientists and for other lead users from industry (including SMEs) and the public sector"[64]. In the vision of the European institutions, the construction of the EDI should be

[61] Mark D. Wilkinson, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1 (2016): 1-9, doi:10.1038/sdata.2016.18. For an in-depth analysis on the concept of data FAIRness, see: 4.1.2.

[62] Paolo Budroni, Jean-Claude Burgelman, Michel Schouppe, "Architectures of knowledge: The European Open Science Cloud." *ABI Technik* 39.2 (2019): 136, doi:10.1515/abitech-2019-2006.

[63] Paolo Budroni, *et al.*, "*Architecture of knowledge*", *op. cit.*, p. 130.

[64] COM/2016/178, p. 8.

considered closely related to the implementation of the HPC strategy[65], namely the European strategy for the promotion of High-Performance Computing, directed to provide the European Union with the enormous computing power of exascale supercomputers[66]. The intention to converge the efforts of the development of Open Science with the development of HPC emerges also from the Recital 9 of the first Council Regulation on Euro-HPC[67]. Recital 9 states that: "[…] a mechanism should be set up at Union level to combine and concentrate the provision of support to the establishment of a world-class European High Performance Computing infrastructure and for research and innovation in High Performance Computing by Member States, the Union and the private sector. This infrastructure should provide access to the public sector users, users from industry, including small and medium-sized enterprises (SMEs), and users from academia, including the communities of the emerging European Open Science Cloud"[68]. Here it is clear that the European institutions aim to make the phenomenon of technological convergence fruitful. In the same direction, consider a recent statement by one of the central bodies of the EOSC[69], the EOSC Executive Board: "EOSC will be instrumental in stimulating many areas of the European private sector, for example, the cloud and artificial intelligence (AI) industries, that are willing to align to

---

[65] European Commission, *High-Performance Computing: Europe's place in a Global Race*, COM/2012/45 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012DC0045.

[66] "The European Data Infrastructures will gather the necessary resources and capabilities, to close the chain from research and development to the delivery and operation of the exascale HPC systems co-designed between users and suppliers.", in COM/2016/178, p. 8. See: Section 6.2.

[67] On the analysis of the EuroHPC project, see: Section 6.2.3.

[68] Council Regulation (EU) 2018/1488 establishing the European High Performance Computing Joint Undertaking (2018) OJ L 252, 2, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R1488.

[69] The bodies governing the EOSC are specified and analysed in Section 2.2.4.

these principles while, at the same time, it will ensure that European researchers remain in control of their data […]"[70].

In this context, a clarification regarding terminology should be made. Although the EOSC is defined as a "cloud", it is not a traditional cloud computing technology from a technical point of view. Without delving into the technicalities of what can be defined as a cloud infrastructure[71], the federated nature of the EOSC ecosystem differs from the properly technical cloud computing systems, whether they are based on a centralised or federated data management model.

Besides, even what is commonly referred to as "the European cloud for business", namely the Gaia X project, is not a real cloud computing system. The Gaia X project is related to cloud services, but it does not intend to build a European cloud, as it is very often wrongly claimed.

This dissertation is focused on scientific research, excluding the European cloud projects for the private and industrial sector from the research scope. However, below there are a few comments to trace the major differences between the EOSC, i.e., a space that is called "cloud" without being so, and Gaia X, i.e., a project that is commonly defined as the establishment of a European cloud for enterprises without being so.

2.2.3.1 EOSC and Gaia X.

As already seen, the EOSC is a trusted and federated environment, for the benefit of researchers, research centers and universities in Europe, which aims to foster sharing and reuse of research data, providing services for data management, analysis, and archiving[72]. The EOSC is the outcome of a project of the European Commission, in close

[70] SRIA, p. 147.

[71] The commonly accepted definition is provided by the National Institute of Standards and Technology (NIST), see: PETER MELL, TIM GRANCE, "The NIST definition of cloud computing." *Special Publication* 800-145 (2011), p. 2, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

[72] "[…] 1.7 million European researchers and 70 million professionals in science and technology a virtual environment with free at the point of use, open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines.", in COM/2016/178, p. 6.

collaboration with stakeholders and Member States. On the other hand, Gaia X is an initiative of Member States, in particular led by France and Germany[73], with the participation of other European partners. The goal is to create a transparent digital ecosystem that provides data processing, analysis and storage services to enable European companies to become globally competitive[74]. The first step seems to be the establishment of a certification system[75] to ensure that companies that meet specific requirements are able to take part in this federated European ecosystem[76].

In many discussions regarding the development of the project Gaia X, however, it is quite usual to come across the expression "a European cloud for companies". On the contrary, the Gaia X project – as well as the EOSC – does not envisage the establishment of a European cloud. The intention is rather to create an environment including different cloud services and several data infrastructures, based on a federated model[77]. If the concept of a European cloud computing may appear evocative, it is not correct, because the cloud services – however important – must be considered only one of the many capabilities available. In other words, the project appears much more complex and refers to data infrastructures that also includes other aspects (e.g., HPC, edge computing, etc.).

---

[73] The European Commission in "A European strategy for data" (already mentioned, which will be discussed in Section 2.2.5) refers to the Gaia X project as one of the Member States' initiative projects, which must be further supported at the European level in close connection with the development of the EU data strategy, in order to foster synergy of action. See: COM/2020/66, p. 18.

[74] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html.

[75] The caveat in stating that the first step of Gaia X is the establishment of a certification system stems from the fact that the project is currently *in fieri*, undergoing major and continuous changes.

[76] Recently, Gaia X has been defined as follow: "Gaia-X is an initiative to develop an open software layer of control, governance, and the implementation of a common set of policies and rules to be applied to any existing cloud/ edge technology stack to obtain transparency, sovereignty and interoperability across data and services. This can be deployed with any cloud player that implements this open SW layer in conjunction with the associated policies and rules.", FRANCESCO BONIFIGLIO, *Gaia X Vision & Strategy*, 2021, https://gaia-x.eu/news/gaia-x-association-releases-its-vision-and-strategy-document.

[77] On this aspect, there is currently a debate regarding the rules of participation in the GAIA X ecosystem, given the participation of the US-based company Plantir. See: https://medium.com/palantir/palantir-and-gaia-x-85ab98845144d.

Analysing the current debate on the implementation of EOSC and Gaia X, some considerations can be made. First, these two projects start from the same premise. Yet this premise is developed differently.

The basic premise is that Europe is a significant producer of valuable data, both research and non-research data[78], but does not have the digital resources needed to exploit them. From the technological point of view, it becomes necessary to rely on the large commercial digital players, non-European, with all the legal, ethical, and economic concerns related[79]. As a result, the need to set limits on the overwhelming power[80] of these big players emerges[81].

In order to avoid this situation, however, EOSC and Gaia X develop the premise differently.

The scientific sector, with the EOSC, aims to embody the paradigm of the Open Science, where science is data-driven and deeply connected with the use of ICTs, but also inspired by the tenets of openness, collaboration and globality, with the overcoming of geographical boundaries[82].

On the opposite, the private sector, with Gaia X and its slogan "initiated by Europe, for Europe", intends to encompass the principle

---

[78] On the definition of research data, see: Section 4.1.1.

[79] But also, with some undeniable – and perhaps too often denied – advantages, such as the efficiency of the services provided.

[80] On this aspect see, in particular: ORESTE POLLICINO, MARCO BASSINI, "Bridge is down, data truck can't get through... a critical view of the Schrems Judgment in the context of European constitutionalism." *A Critical View of the Schrems Judgment in The Context of European Constitutionalism* (2017): 1-28; and GIOVANNI DE GREGORIO, "The Rise of Digital Constitutionalism in the European Union." *International Journal of Constitutional Law* (2020): 1-28.

[81] In any case, it should be pointed out that we do not currently have the means to say whether restricting non-European players is always the best decision. The case (or failure) of the national COVID-19 tracking apps has reminded us that the efficiency of the services provided must be strongly considered from an economic, ethical, and deontological, and – not least – legal, point of view.

[82] Besides, this direction is the one indicated by the very nature of research, which is intrinsically free from geographical boundaries. This statement is not intended to express a too idealistic, if not idealised, view of the matter under investigation. It is intended only to describe the conceptual starting point that drives – or should drive – the implementation of the various projects. There is the awareness that in the practical implementation different interests may prevail or diverge significantly from this starting point. This initial vision does not invalidate an objective and critical evaluation of project implementation. On this aspect, see: Section 3.4.2 and Section 7.2.

of digital sovereignty. According to this principle – very debated in the recent discussions on digitisation[83] – the control over digital technologies and their impacts should be exercised on a national or territorial level[84].

Given these different developments of the initial premise, however, EOSC and Gaia X are still connected, especially from the point of view of data management[85]. It would be worthwhile to develop some forms of coordination between the EOSC and Gaia X[86]. Both projects also include non-European players that are frequently at the centre of the debate. This aspect may raise some concerns from a legal point of view, which will be investigated later[87].

However, currently, the Gaia X project is experiencing what can be called a real downsizing[88]: the risk is that it will remain a half-

---

[83] On the concept of digital sovereignty, see: LUCIANO FLORIDI, "The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU." *Philosophy & Technology* 33.3 (2020): 371, doi: 10.1007/s13347-020-00423-6. See, also, ELEONORA BASSI, "Dati, Sovranità, Nuovi Modelli di Governance." in UGO PAGALLO, MASSIMO DURANTE (eds.), *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società*. (Milano-Udine: Mimesis, 2022), pp. 203-220.

[84] In particular, on the concept of digital sovereignty in the context of the cities, see: FRANCESCA BRIA, MALCOM BAIN, "Manifesto in favour of technological sovereignty and digital rights for cities", version 2.0 (2018), https://www.barcelona.cat/digitalstandards/manifesto/0.2/.

[85] This also emerged from the "Widening to the public and private sectors" meeting on October 21, 2020, at the EOSC Symposium 2020, specifically in the talk by Andrea Weiss, Head of Digital Business Models, on the structure of Gaia X. In the meeting, in identifying the points of contact between EOSC and Gaia X, a relevant aspect was the one related to FAIR principles in data production: the so-called "fairification" of data, i.e., the verification of data under FAIR principles, a typical feature of EOSC, can represent a point of valid connection between the two projects. See: "Widening to the public and private sectors - Oct 21 #EOSCSymposium2020", EOSC Portal, YouTube, October 21, 2020, https://www.youtube.com/watch?v=dX_lPdiC24M. Consider that, in this regard, the European strategy for data (COM/2020/66) mentions, in fact, the FAIR principles, in general, in relation to the issue of interoperability. On the FAIR Data Principles, see: Section 4.1.2.

[86] The SRIA document states, in fact, the need of some sort of coordination between these two projects. "The Rules will also need to reflect changes in the wider environment, such as the development of the GAIA-X initiative", see: SRIA, p. 116.

[87] On the interplay between public and private sector in the scientific research domain and the legal consequences, see: Section 6.4. There, an analysis of the impact of the US CLOUD Act will be realised.

[88] Although in December 2021 two reports (on the vision and strategy and on the architecture of Gaia X) have been released, the implementation of this project is much debated, mainly because the project's governance authorities have delayed the release of policy rules for participation in Gaia X, originally expected in September 2021. See: Gaia-X European Association for Data and Cloud AISBL, *Gaia-X Architecture Document*, 2021, https://www.gaia-x.eu/publications.

developed project. Precisely for this reason, designing forms of structural coordination between EOSC and Gaia X could become mutually beneficial.

In any case, both projects are in full progress, and it will take time to see next developments, in order to understand in which direction they intend to steer.

2.2.4 European Open Science Cloud First Steps

Back to the developments of the Open Science in EU, the attention should be drawn to the mid-term review on the implementation of the Digital Single Market Strategy, in 2017[89]. The European Commission, in the section entitled "Developing a European Open Science Cloud, High Performance Computing and European Data Infrastructure" stressed again the need to invest in and implement the HPC infrastructures in support of scientific research. The Commission announced that "Europe cannot take the risk that data produced by EU research and industry will be processed elsewhere because of the lack of High Performance Computing capabilities. This would increase our dependency on facilities in third countries and would encourage innovation to leave Europe"[90]. Once again it is clear that the European Commission supports a connection between two initiatives, namely the establishment of the EOSC and the HPC implementation projects, revealing that the ambition is to take part in the process of technological convergence already mentioned.

Consider, however, that the EOSC is an ecosystem designed primarily – at least in its initial phase – for scientific research; while HPC represents capabilities that provide great computational power and considerable space for data storage and analysis, which certainly

---

[89] European Commission, *The Mid-Term Review on the implementation of the Digital Single Market Strategy. A Connected Digital Single Market for all*, COM/2017/228 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017DC0228.

[90] COM/2017/228, p. 20.

can be decisive in the advancement of many scientific research projects but, equally, can be used by companies for commercial purposes.

A further important step, then, is represented by the Recommendation of the European Commission, regarding access to and preservation of scientific information, released on April 25, 2018[91]. By arguing that "[T]echnological progress has over time caused a major shift in the world of science towards increasingly collaborative methods and has steadily contributed to an increasing volume of scientific material"[92], this Recommendation intends to be an enabler of coordination between the various initiatives developed at European level and the subsequent – and still essential – implementation at national level. In particular, it is established that within eighteen months from the publication of these Recommendation in the Official Journal of the EU, Member States shall communicate to the European Commission the initiatives undertaken in the field of access to and preservation of scientific information[93]. This Recommendation is an important step because it can be interpreted as a real mandate from the European Commission to the Member States to implement Open Access policies, namely a free and online access to the scientific results[94].

[91] Commission Recommendation (EU) 2018/790, *On access to and preservation of scientific information*, of 25 April 2018, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018H0790. Consider that this Recommendation was intended to replace: Commission Recommendation (EU) 2012/417, *On access to and preservation of scientific information*, of 17 July 2012, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012H0417.

[92] Recital 9, C/2018/2375.

[93] This Recommendation, issued by the then European Commissioner for Research, Science and Innovation Carlos Moedas, is certainly intended to represent a significant boost in the Open Science policies of the Union, which although in line with the developments described so far, appears somewhat innovative compared to the traditional European policy design in the field of scientific research. On the perspective of Carlos Moedas, see: ELENA GIGLIA, "Open Access e Open Science: per una scienza più efficace." *Journal of Biomedical Practitioners* 1.1 (2017): 22-23.

[94] In general, the expression Open Access refers to access to scientific research literature (e.g., publications) that is supposed to be free, without additional fees, digital and online. However, currently the expression Open Access also refers to the access to research data. For further analysis of the concept of Open Access, as part of the Open Science paradigm, see Section 3.1.2 and Section 3.2.

Against this backdrop, after a few months, on November 23, 2018, the EOSC was launched, in Vienna, at the end of a long period of consultations with the various stakeholders involved in the field. On this occasion, the "Vienna Declaration on the European Open Science Cloud"[95] was released, emphasising the key points of the initiative. It recommends considering the EOSC implementation as a process and not as a project, to highlight its dynamic nature; it also points out the need for a joint effort, between EU and Member States, in identifying a common governance framework; and, finally, it stresses the potential of the EOSC "[...] to enable first-class data-driven science and to stimulate new business models benefiting our society and the economy"[96].

The EOSC Declaration also shows the intention to govern, together, data, infrastructures, and computational power, evoking the concept of the technological convergence. This Declaration asserts that: "[...] Europe is well placed to take a global leadership position in the development and application of cloud services for Science"[97]. This optimistic statement is due to the fact that "[I]n an initial phase of development from 2017 to 2020, the EC made a financial investment of approximately €320 million to begin building the foundations of EOSC through project calls in Work Programmes in Horizon 2020"[98].

Even at this early stage of development, the EOSC is beginning to take shape. In light of the substantial investments made since its launch, it can be interpreted as the project through which the European Union participates in addressing the major global challenges of our time[99].

---

95 University of Vienna, "The Vienna Declaration on the European Open Science Cloud", November 23, 2018, https://eosc-launch.eu/declaration/.

96 "The Vienna Declaration", Section 11.

97 "The Vienna Declaration", Section 4.

98 SRIA, p. 47.

99 Consider that the EOSC Executive Board is in line with this vision, arguing: "[T]he climate crisis, the extinction of species, global poverty and social inequality are only a few of the challenges that

After the description of the first steps of the EOSC implementation (i.e., the preparatory European Cloud Initiative and the consequent launch of the EOSC), it is now time to focus on the next phase of the Open Science policy developments, represented by the year 2020, as a real watershed.

2.2.5 The Year 2020: A Tipping Point for Open Science

The events of the year 2020 and their implications for scientific research can hardly be ignored. This year represented a real tipping point for the Open Science paradigm essentially for four reasons: (i) the concretisation of what can be called the "von der Leyen Doctrine"; (ii) the COVID-19 pandemic; (iii) the developments on the governance of EOSC project; and, finally, (iv) the reform of the European Research Area (ERA). Below follows a further analysis of these four reasons.

(i) von der Leyen Doctrine. In January 2020, the current President of the European Commission, Ursula von der Leyen, expressed a relevant statement, at the World Economic Forum. In exposing the policy direction of the Commission in the field of data, von der Leyen dwelt specifically on the EOSC. The President stressed that "[D]ata is a renewable resource such as wind or sun. Every 18 months we double the amount of data we produce [...] 85 per cent of which is never used"[100]. More importantly, it was also declared the intention to co-create a framework to allow the use of these data through the EOSC, a means by which we can take advantage and ensure sustainability of such data, defined as a hidden treasure. The President of the European Commission, in other words, made explicit a very precise vision, which can be defined as the "von der Leyen Doctrine". The President, in fact, renewed the debate on science but to highlight a very specific aspect of

humankind is facing in the 21st century [...]. Research plays a crucial role in addressing these challenges and, against this background, EOSC will be a major European vehicle for joining forces to help transform individual research efforts into collective efforts.", SRIA, p. 142.

[100] "Ursula von der Leyen speaks at the World Economic Forum", Euronews, Youtube, live January 22, 2020, https://www.youtube.com/watch?v=_A7Q514z_dw&feature=youtu.be&t=649.

it: the economic value that can be obtained from sharing and reusing data, including scientific data.

In the following month, the European Commission released the Communication n. 66 of 2020, entitled “A European Strategy for Data”, outlining the direction to be followed for the strengthening of a digital Europe. The Communication aims to achieve the considerable benefits arising from data-driven innovation, but “in accordance with European values, fundamental rights and rules”[101]. This twofold objective is based on the idea that “[C]itizens will trust and embrace data-driven innovations only if they are confident that any personal data sharing in EU will be subject to full compliance with the EU’s strict data protection rules”[102].

At first glance, this approach seems to embody the characteristic European model of building trust on the basis of a lack of trust: the idea that citizens will trust data-driven innovation only if they can rely on a system of safeguards ensured either by technological security or by legal certainty (understood as a system of sanctions and responsibilities that provide certainty)[103].

Yet, this Strategy is not only this. It can be considered as an official concretisation of the “von der Leyen Doctrine” which aims, whilst respecting European values, to emphasise the economic benefit deriving from the exploitation of data, their sharing and reuse.

Furthermore, the EU data strategy entailed the establishment of nine “Common European data spaces in strategic sectors and domains of public interest”[104], taking the EOSC initiative as an example for their implementation. The EOSC, in fact, considering the progress

[101] COM/2020/66, p. 1.
[102] *Ibid.*
[103] MASSIMO DURANTE, “Sicurezza e fiducia nell'età della tecnologia.” *Filosofia politica* 29.3 (2015): pp. 439-458.
[104] The sectors considered strategic are: Common European industrial (manufacturing) data space; Common European Green Deal data space; Common European mobility data space; Common European health data space; Common European financial data space; Common European agricultural data space; Common European data spaces for public administrations; Common European skills data space.

made in recent years, and aiming not only to the management but also to the sharing and reuse of research data was represented as a model on which to base further developments of these European Common Data Spaces.

(ii) COVID-19 Pandemic. The second reason for making 2020 a breakthrough year for scientific research is the COVID-19 pandemic. One of the many effects of this global crisis is surely to have brought the debate on science, how it works and the traditional dynamics of scientific research under renewed scrutiny by the public at large. The need to tackle a crisis that affected everyone, without exceptions, gave new strength to a vision of scientific research characterised by collaboration and openness of methods, data, and results, typical of the Open Science paradigm.

The research projects seeking to develop a vaccine against the SARS-CoV-2 virus, in fact, made more evident than ever the need for a science that is "as open as possible, as closed as necessary". In this regard, projects such as ERAvsCORONA have been launched, with the aim of developing a collective effort in the fight against the virus[105].

However, a renewed interest in a more open and collaborative science was not only the expression of top-down projects, but was, perhaps above all, the expression of a bottom-up demand, directly from the scientific community and some publishers. As an illustration, consider that in February 2020, in an editorial published in the journal "Nature", it was explicitly stated the need for a great data sharing to cope with the spread of the virus, presenting an agreement between some publishers, to ensure the fastest and most effective dissemination

[105] https://ec.europa.eu/info/sites/info/files/research_and_innovation/research_by_area/documents/ec_rtd_era-vs-corona.pdf. Another initiative worth mentioning is the one that led to the establishment of the "European COVID.19 Data Platform". This platform – part of the ERAvsCORONA Action Plan – "[…] aims at providing an open, trusted, and scalable European and global environment where researchers can store and share datasets, such as DNA sequences, protein structures, data from pre-clinical research and clinical trials, as well as epidemiological data.", see: https://joinup.ec.europa.eu/collection/digital-response-covid-19/news/european-covid-19-data-platform.

of knowledge[106]. Besides, this event is also showing the full feasibility of such an operation[107].

(iii) EOSC Governance Developments. In the meantime, the EOSC began to operate from its portal, "The EOSC Portal", albeit in an experimental phase[108].

The governance of the initiative was organised in its multiple layers, to tackle the problems inherent to the implementation. The structure of the governance is quite complex, and it has four main bodies:

- the Governance Board, composed by representatives of the different Member States;
- the Executive EOSC Board, which is made up of 11 members, selected through a call for applications;
- the Stakeholders Forum, which includes users, representatives of the various Open Science projects implemented in recent years, both at European and national level, but also representatives of the public sector and industry;
- five Working Groups, composed of selected representatives of stakeholders, focusing on the five strategic priorities of the

[106] "Researchers must ensure that their work on this outbreak is shared rapidly and openly. […] Nature and its publisher Springer Nature have now signed a joint statement with other publishers, funders and scientific societies to ensure the rapid sharing of research data and findings relevant to the coronavirus. […] For researchers, the message is simple: work hard to understand and combat this infectious disease; make that work of the highest standard; and make results quickly available to the world.": "Calling all coronavirus researchers: keep sharing, stay open" *Nature*, 576 (2020): 7, doi: 10.1038/d41586-020-00307-x.

[107] On the role of scientific journals, from a philosophical perspective, see: GIOVANNI BONIOLO, LISA ONAGA, "Seeing clearly through COVID-19: current and future questions for the history and philosophy of the life sciences." *History and Philosophy of the Life Sciences*, 43 (2021): 81-83, doi: 10.1007/s40656-021-00434-2. On the feasibility of the openness of scientific results, made concrete during the COVID-19 pandemic, see: LUDOVICA PASERI, "COVID-19 Pandemic and GDPR: When Scientific Research becomes a Component of Public Deliberation", in DARA HALLINAN, RONALD LEENES, PAUL DE HERT (eds.), *Data Protection and Privacy: Enforcing Rights in a Changing World.* (London: Hart Publishing, 2022): pp. 171-172. This paper develops an examination of the role of scientific research as part of public deliberation, starting from the analysis presented in: GIOVANNI BONIOLO, *Il pulpito e la piazza. Democrazia, deliberazione e scienze della vita.* (Milano: Cortina Editore, 2011).

[108] https://eosc-portal.eu.

implementation (landscape; FAIR Data; architecture; rules of participations; sustainability)[109].

(iv) New ERA. In September 2020, the European Commission issued the Communication n. 628 of 2020, aimed to establish "A new ERA for Research and Innovation"[110]. The European Research Area (ERA), born in 2000 and formalised in 2008, is a cornerstone of the Open Science policies in Europe. In Section 2.2.1, the ERA has been interpreted as the substantial origin of the concept of Open Science within the architecture of the European Union. In the establishment of the ERA, the European institutions never explicitly mentioned the Open Science: this concept, rather, laid the ground for a vision of science characterised by openness, circulation of knowledge and researchers. As an illustration of its founding role in the Open Science paradigm, its reform has been proposed twenty years after its establishment, in connection the recognition of the Open Science paradigm by the European institutions[111]: "The ERA has enhanced access to open, free of charge, re-usable scientific information initiative through the Open Science and the recently launched European Open Science Cloud (EOSC) creating a cloud area for research data in Europe allowing for better science through open and collaborative knowledge sharing"[112].

This Communication, therefore, lays the foundations for a reform of the ERA designed to adapt it to contemporary needs, as a reaction to the profound transformation that scientific research is undergoing: "[…] a new ERA will need to boost Europe's recovery and to support its green and digital transitions by supporting innovation based competitiveness and fostering technological sovereignty in key strategic areas (e.g. Artificial Intelligence and data, microelectronics,

[109] See: SARAH JONES, JEAN-FRANÇOIS ABRAMATIC, "European Open Science Cloud (EOSC) Strategic Implementation Plan." *Publications Office of the European Union* (2019), doi: 10.2777/202370.
[110] European Commission, *A new ERA for Research and Innovation*, COM/2020/628 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:628:FIN.
[111] See: Section 3.2.
[112] COM/2020/628, p. 1.

quantum computing, 5G, batteries, renewable energy, hydrogen, zero-emission and smart mobility, etc.) in line with the model of open strategic autonomy"[113].

Bearing in mind these four circumstances that made the year 2020 a turning point for Open Science, it is now time to focus on the latest developments related to the Open Science initiatives, in order to clarify the general framework of policies implemented so far at European level.

2.2.6 Horizon Europe

In April 2021, the European Union launched a new grant programme for research and innovation, i.e., the Horizon Europe programme. This new funding plan, following the previous Horizon 2020 programme, has been established by the Regulation (EU) 2021/695[114].

Horizon Europe allocated EUR 95.5 billion to scientific research, for the four-year period 2022-2027. This investment represents an upward trend of 30% compared to the first framework programme, launched in 1994[115].

The Horizon Europe programme is a relevant step in the definition of the EU Open Science approach because the Open Science becomes a backbone of the programme[116]: Recital 8 states that "Open science,

[113] COM/2020/628, p. 4. In addition, in this Communication there is a clear reference to the so-called the "von der Leyen Doctrine". One of the key points of the reform is "Translating R&I results into the economy": the idea of promoting science as a driver for the economy, as expressed by the President of the European Commission at the World Economic Forum, clearly emerges.

[114] The Horizon Europe framework programme has been proposed in 2018: European Commission, *Proposal for a Regulation of the European Parliament and of the Council, establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination*, COM/2018/435 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018PC0435. Then, it has been approved in 2021, see: Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021, *establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013*, OJ L 170, 12.5.2021, p. 1–68, ELI: http://data.europa.eu/eli/reg/2021/695/oj.

[115] QUIRIN SCHIERMEIER, "How Europe's€ 100-billion science fund will shape 7 years of research." *Nature* 591.7848 (2021): pp. 20-21, doi: 10.1038/d41586-021-00496-z.

[116] On the link between Horizon 2020 and Open Science and the Horizon Europe programme and Open Innovation, see: DOUGLAS K. R. ROBINSON, ANGELA SIMONEC, MARZIA MAZZONETTO, "RRI legacies: co-

including open access to scientific publications and research data, as well as optimal dissemination and exploitation of knowledge have the potential to increase the quality, impact and benefits of science. They also have the potential to accelerate the advancement of knowledge by making it more reliable, efficient and accurate, more easily understood by society and responsive to societal challenges"[117].

The Article 14 is precisely entitled "Open Science" and establishes that "[T]he Programme shall encourage open science as an approach to the scientific process based on cooperative work and diffusing knowledge"[118]. In addition, the Article 39 embraces Open Science, stating that "[T]he work programme may provide for additional incentives or obligations for the purpose of adhering to open science practices"[119].

The paradigm of Open Science is thus officially recognised, finalising the evolution of the EU approach to scientific research that in this dissertation has been traced back to the early 2000s, with the establishment of the fifth European freedom, i.e., the free movement of knowledge. In other words, the idea of knowledge sharing, which has been part of the European design since decades[120], is recognised in hard law, through the establishment of Open Science.

---

creation for responsible, equitable and fair innovation in Horizon Europe." *Journal of Responsible Innovation* 8.2 (2021): 209-216, doi: 10.1080/23299460.2020.1842633. I agree with the emphasis on "Open Innovation" emerging in the new programme, which I believe is in line with the "von der Leyen Doctrine" outlined above. However, I do not support the position that relegates Open Science to the previous programme, in order to give space in Horizon Europe only to Open Innovation. Open Science and Open Innovation are two distinct phenomena, although in continuity (on the definition of Open Science and its different dimensions, see: Chapter 3). Consider, in fact, that in the Regulation establishing the Horizon 2020 framework programme, Open Science was not even mentioned.

[117] Recital 8, Regulation (EU) 2021/695.

[118] The Article 14, Regulation (EU) 2021/695.

[119] The Article 39, Regulation (EU) 2021/695.

[120] The intention is to emphasise that the circulation of knowledge has become a pillar of the European design at exactly the same time as the European Union has lost its purely economic meaning and has pursued the ambition of being a union of European citizens and cultures.

In order to track the progress set in the Horizon Europe programme, the Annex V presents the "Key impact pathway indicators"[121]: the category "fostering diffusion of knowledge and Open Science" represents one of the goal to be tracked and monitored. However, the short-, medium- and long-term indicators identified for this category appears to be very broad and general[122].

Furthermore, another significant aspect is that the Horizon Europe programme gives considerable attention to the social engagement in the research domain. This new funding programme encourages greater civil society participation in scientific research. This involvement, traditionally represented by the expression "citizen science", is often considered one of the possible strands of the Open Science paradigm[123]. While waiting to see how this aspect of the programme will be realised, it should be considered that the participation of citizens in science, although it may bring certain benefits, also entails a considerable number of concerns[124]. Think about the resources to be deployed by researchers in terms of time, training efforts, rewards, coordination, establishment of procedural criteria, etc. For this reason, such initiatives can only be successful if supported by concrete and effective strategies at national and local level[125].

---

[121] Annex V, Regulation (EU) 2021/695.

[122] The sort-term indicator is "Shared knowledge – Share of research outputs (open data/ publication/software etc.) resulting from the Programme shared through open knowledge infrastructures"; the medium-term is "Knowledge diffusion – Share of open access research outputs resulting from the Programme actively used/cited"; and the longer-term indicator is "New collaborations – Share of Programme beneficiaries which have developed new transdisciplinary/transsectoral collaborations with users of their open access research outputs resulting from the Programme", see: Annex V, Regulation (EU) 2021/695.

[123] On the meaning of Open Science, on its different facets see: Chapter 3, in particular Section 3.1, where a more in-depth analysis is provided.

[124] The intention of the European institutions to increase social involvement in science can be interpreted as part of a wider project to broaden citizen participation in the activities of the European Union, exemplified by projects such as the "European Citizens' Initiative". On this aspect, see: ERIK LONGO, "The European Citizens' initiative: too much democracy for EU polity?" *German Law Journal* 20 (2019): 181-200, doi: 10.1017/glj.2019.12.

[125] On a topic pertaining to the citizen science, i.e. crowdfunding of science, see: LUDOVICA PASERI, "Crowdfunding of Science and Open Data: Opportunities, Challenges, and Policies." in ANDREA KŐ, *et*

However, it is essential to wait for the actual development of these more recent stages, in order to be able to elaborate a more accurate analysis. Next section will conclude the analysis on the evolution of the concept of Open Science in Europe by investigating the latest progress of the EOSC.

2.2.7 Looking Forward: EOSC Association and Partnership

At this stage of the mapping of the European Open Science developments, the complexity of the dynamics is evident, given the great number of actors and the breadth of topics involved.

In order to reduce the complexity arising from the multiple interests involved, and the respective multiple stakeholders, a legal entity has been created: the EOSC Association, based in Belgium, an entity authorised to be part of binding contractual agreements[126]. This Association aims to be a unique touchstone for EOSC governance decisions. The project, which was born from a considerable bottom-up drive, with the leading role of the scientific community, has *de facto* established a centralised management and decision-making body. As an illustration, consider that the SRIA document states that "[L]essons learned in the first implementation phase have shown that while the project-based approach is very successful in involving the many stakeholders and communities in developing the EOSC ecosystem, the individuality and freedom of projects has led to a fragmented landscape of systems and stakeholders"[127]. The identification of an Association represents an evolution from the first steps of the project's implementation. The purpose specified in the Statute of the Association

---

al., (eds.) *Electronic Government and the Information Systems Perspective: 8th International Conference – EGOVIS 2019*, 11709 (Cham: Springer 2019): 3-15, doi: 10.1007/978-3-030-27523-5_1. Here, starting from the experience of the "Ricerca e Talenti" Foundation of the University of Turin, it was analysed how the lack of specific support in policies makes the development of such projects very difficult in practice. Other successful examples of citizen science are described in: JENNIFER ECKHARDT, *et al.*, "Ecosystems of Co-Creation" *Frontiers in Sociology* 30 (2021), p. 6, doi: 10.3389/fsoc.2021.642289.

[126] The EOSC Association was established on 29 July 2020, see: https://eosc.eu/documents.

[127] SRIA, p. 48.

is to provide "[…] a single voice for advocacy and representation for the broader EOSC stakeholder community"[128].

According to the Statute, members of the Association are research funding and performing organisations; service providing organisations; or other organisations with an interest in the context of the European scientific research.

The most important aspect of the establishment of this Association is that it is functional to the identification of a co-programmed EOSC Partnership, under the new European funding programme for innovation and research, i.e., Horizon Europe. In other words, the EOSC Association, as a legal entity, is legally capable of becoming the effective signatory, together with the European Commission, of proper contractual agreements, with all the public and private partners intending to participate in the EOSC.

This partnership pursues two main goals. First, it becomes the means by which to guarantee participation: "[T]his EOSC Partnership brings together all relevant stakeholders to co-design and deploy a European Research Data Commons where data are findable, accessible, interoperable and reusable (FAIR)"[129]. In addition, it enables the take-up of financial grants from the Horizon Europe research programme. Consider that the managing of relations with the various public and private partners is one of the most delicate issues: it is closely related to the identification of the rules of participation in the EOSC federated environment[130].

Regarding the second goal, the role of the EOSC Partnership is crucial in relation to the issue of research funding in Europe. The establishment of a legal entity such as the EOSC Association, as well as the identification of the EOSC Partnership, allow the effective

[128] See, the EOSC Association Statute: https://eosc.eu/sites/default/files/EOSC_Statutes.pdf.
[129] SRIA, p. 53.
[130] An entire working group "Rules of Participation" is dedicated to this issue (already mentioned in Section 2.2.4), see: https://www.eoscsecretariat.eu/working-groups/rules-participation-working-group.

management of grants resulting from the Horizon Europe programme, among the various partners and stakeholders involved.

The systematisation of the European policies on Open Science has been completed: it started from the establishment of the ERA at the beginning of 2000s; went through the foundation of the EOSC project; and ended today with the establishment of the EOSC Association and the launch of the new Horizon Europe programme.

This analysis represents the starting point for understanding the complex and multifaceted scenario of scientific research in Europe.

Next section casts light on the legislative disciplines involved in the scenario just described.

2.3 Regulatory Frameworks

The topic under investigation is the European governance of scientific research and its commitment to the paradigm of Open Science: collaborative, global, closely related to the use of cloud services and computational power provided by HPC platforms. The previous Section 2.2 explored the different stages of the European policies in the field of Open Science. This analysis illustrated a complex set of strategies developed over years. These EU initiatives, however, do not arise in a regulatory vacuum: the Open Science polices must be placed within the existing European legal framework.

Four main strands of investigation are primarily involved: (i) the discipline of Open Data; (ii) the discipline of data protection; (iii) the set of rules on copyright and the ownership of data; and, finally, (iv) looking forward, some considerations should be made about the package of Regulation proposals recently presented by the European Commission.

2.3.1 Open Data

First, attention is drawn to the domain of Open Data. The expression Open Data is generally understood as the set of initiatives aimed to open up the data produced, collected and held by the public sector, in order to generate new value, both in economic terms and for society, through their reuse. The European Open Data framework should necessarily be taken into account with regard to Open Science, since both phenomena involve the sharing and reuse of data.

The backbone of this domain is the so-called Open Data Directive (ODD), namely the Directive (EU) 2019/1024[131], which in June 2019 substantially amended the previous PSI 2003/98/EC[132].

The key issue is to understand the relationship between Open Science policies and provisions of the European Open Data framework. Very often the concept of Open Data is erroneously overlapped with the concept of Open Science. Besides, sometimes this overlapping and consequent confusion is triggered by the European institutions themselves, handling the two phenomena, (i.e., the Open Science and the Open Data) as a whole[133].

---

[131] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019, *on open data and the re-use of public sector information*, OJ L 172, 26.6.2019, 56–83, ELI: http://data.europa.eu/eli/dir/2019/1024/oj.

[132] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003, *on the re-use of public sector information*, OJ L 345, 31.12.2003, 90-96, ELI: http://data.europa.eu/eli/dir/2003/98/oj, as modified by the Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC, *on the re-use of public sector information*, OJ L 175, 27.6.2013, p. 1–8, ELI: http://data.europa.eu/eli/dir/2013/37/oj.

[133] According to the Communication of the European Commission, *Towards a common European Data Space,* COM 2018/232, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52018DC0232, three key steps are identified on the basis of which to constitute the "Common Data Spaces": (1) a review of the PSI directive, for data produced or collected by the public sector; (2) the promotion of private sector data sharing, with reference to B2B; (3) the update of the Recommendation on access to and preservation of scientific information (the update of this recommendation was analysed in the previous Section 2.2.4, with reference to Recommendation EU 2018/790). The fact that the Open Data sharing, in the public sector, and the scientific research data sharing are both potentially components of the same European Data Space should not be confusing. First, because the aforementioned Recommendation 2018/790, which is the update of the Recommendation on Access to and Preservation of Scientific Information – mentioned in Communication COM 2018/232 – is primarily directed to scientific research results, rather than data. It in fact, in Recital 6, expressly refers to the concept of Open Access, as "[...] a means of dissemination for researchers who may decide to publish their work, in particular in the context of publicly-funded research". Open Access, which is one of the facets of Open Science, can refers both to data and to publications; in this case the focus is on the final publication, i.e., the result of the research project, not only the data processed conducting the research project. See: Section 3.1.2 and Section 4.1.

Although these two concepts are clearly connected, they should not be considered synonymous[134].

The boundary between these two categories tends to blur to the extent that the Article 1(1)c of the ODD includes in its scope the category of research data. According to the Article 2(9) of the ODD, the research data are defined as "documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results".

In addition, both Recitals 27 and 28 of the ODD underlines the importance of exploiting the huge amount of data generated by scientific research by defining policies, at national level, so that "[...] certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidized by public funding or co-funded by public and private-sector entities".

The Article 10 of the ODD, specifically addressed to research data, makes an explicit reference to the FAIR Data Principles, i.e., the set of rules that indicate how data should be, from a technical point of view (findable, accessible, interoperable, and reusable)[135].

It is established that Member States shall support the availability of research data, with the adoption of national policies that are compliant with the FAIR Data Principles: this is the first time that this concept is organically provided for in a European Union legislative act.

Despite the official adoption of this concept into legislation, the sharing of FAIR research data is still far from being established as a practice by the scientific community: there are still open issues, both in

[134] On the relationship between the general concept of Open Science and the Open Data, see: Chapter 4.

[135] The FAIRness of research data has already been mentioned, in relation to the establishment of the EOSC, in Section 2.2.3. However, for an in-depth analysis of the concept see: Section 4.1.2.

legal[136] and in ethical terms[137]. Since "[…] the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year"[138].

The first legislative stance on FAIR Data – represented by the ODD and its national transpositions – and the divergent practices still persisting among part of the scientific community are two opposing trends that need further investigation[139].

2.3.2 Data Protection

The data protection discipline, whose pillar in Europe is the General Data Protection Regulation (GDPR)[140], is necessarily involved in science: many scientific research projects entail the processing of personal data. This discipline is crucial, *a fortiori*, in the context of Open Science, which is based on sharing and reuse of research data.

Think about research projects conducted in the medical sector, which processes not only personal data, but health data, one of the special categories of personal data, protected by the Article 10 of the GDPR; or studies in the field of genetics, based on the processing of genetic data, also considered as a special category of personal data, subject to additional safeguards.

The processing of personal data can also be an integral part of scientific research projects not related to medicine: for example, research projects in sociology that involve surveys; or research projects in the economic field that require the profiling of possible consumers, etc.

---

[136] See: Section 4.2 "Legal issues of the Open Research Data".
[137] See: Section 4.3 "Ethical Issues of the Open Research Data".
[138] PwC EU Services, *Cost of not having FAIR research data. Cost-Benefit analysis for FAIR research data*, Luxembourg: Publications Office of the European Union, 2018, p. 4, doi: 10.2777/02999.
[139] For this reason, Chapter 4 is dedicated to the topic of "Open research data".
[140] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, *on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ L 119, 4.5.2016, 1–88, ELI: http://data.europa.eu/eli/reg/2016/679/oj.

In all these cases a personal data processing is carried out and, therefore, it falls under the application of the GDPR, which, according to the Article 2 dedicated to the "material scope", states that "[T]his regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system"[141].

The situation becomes even more complex when such processing involves cloud computing services or HPC platforms: the flexibility of the GDPR does not limit its applicability in new technological scenarios, but nevertheless they require further and more in-depth considerations.

Processing of personal data for "archiving purposes in the public interest, scientific or historical research purposes or statistical purposes"[142] are covered by specific – and partially derogatory – rules, with the aim of promoting as much as possible the development of research.

Given, on the one hand, the openness and sharing driven by the Open Science paradigm (albeit not blindly or without limits) and, on the other hand, the specific European data protection rules, it is crucial to deepen the relations between the two topics[143]. This need is real considering both the gap in literature about the protection of personal data in the context of Open Science, and the increasing number of problems in this field arising in everyday research.

2.3.3 Copyright and Ownership of Data

Another key issue related to scientific research is the ownership: this refers to research results, i.e., publications, and to research data, whether created or collected. The issue of ownership is closely related to the European disciplines on copyright.

---

[141] The Article 2(1), GDPR.

[142] The Article 89, GDPR.

[143] For this reason, Chapter 5 is dedicated to the topic of the "Data protection issues in Open Science".

This domain is complex from a legal perspective, due to a rather consistent stratification of provisions, at European and international level. In addition, the innovation introduced by technology increases the complexity in identifying the subject of protection under the copyright legal framework.

In the field of copyright, the first landmarks should refer to the international sources of law, starting from the Berne Convention of 1866, then moving on to the complex European framework[144] – which has been developed around the international one[145] – and, finally, to the national provisions[146].

The subject of copyright protection, i.e., the "work of authorship", is specifically defined, either in the international or in the European framework, as a product of human creativity, protecting the expressive form, not the mere idea at the basis. In this context, the concept of originality becomes crucial. Yet, the minimum level of the necessary originality criterion can be identified by the jurisprudence of the European Court of Justice (ECJ)[147].

---

[144] At European level, consider: Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001, *on the harmonisation of certain aspects of copyright and related rights in the information society*, OJ L 167, 22.6.2001, 10–19, ELI: http://data.europa.eu/eli/dir/2001/29/oj; Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004, *on the enforcement of intellectual property rights*, OJ L 157, 30.4.2004, 45–86, ELI: http://data.europa.eu/eli/dir/2004/48/oj; Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019, *on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*, OJ L 130, 17.5.2019, 92–125, ELI: http://data.europa.eu/eli/dir/2019/790/oj; Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996, *on the legal protection of databases*, OJ L 77, 27.3.1996, 20–28, ELI: http://data.europa.eu/eli/dir/1996/9/oj; Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009, *on the legal protection of computer programs*, OJ L 111, 5.5.2009, 16–22, ELI: http://data.europa.eu/eli/dir/2009/24/oj; Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012, *on certain permitted uses of orphan works*, OJ L 299, 27.10.2012, 5–12, ELI: http://data.europa.eu/eli/dir/2012/28/oj. On the European Copyright legislation, see: PAOLO GUARDA, "Research paper on Case Study (iii): Barriers that citizens face regarding their intellectual property rights", *BEUCitizen, University of Trento* (2016), pp. 10-30.

[145] LUCIE GUIBAULT, ANDREAS WIEBE, *Safe to be open: Study on the protection of research data and recommendations for access and usage*. (Göttingen: Universitätsverlag Göttingen, 2013), p. 20.

[146] Since the complex European copyright framework is mainly composed of directives, it is essential to take into account the national dimension, i.e., the transposition acts of the various Member States, which can also differ considerably, making the discipline extremely fragmented.

[147] Among others, see: Judgment of the Court (Fourth Chamber) of 16 July 2009, *Infopaq International*

Considering the publications, understood as the result of scientific research, i.e., the work produced by researchers as the outcome of their own project, the protection provided by copyright is in force[148].

The situation is more complex and debated in the case in which the subjects of protection are research data, understood as set of data, organised in datasets, which have been collected, eventually processed, and analysed, in the realisation of a scientific research project[149].

Since the protection envisaged by copyright for intellectual works must possess the requirement of originality, it may not always be easy to protect datasets containing research data[150]. By contrast, sometimes an over-protection may arise, hindering data sharing[151].

In a scenario in which the policies of European institutions seem strongly directed to the promotion of sharing and reuse of research data, the situation – at first glance, even contradictory – is considerably complicated[152].

A/S v Danske Dagblades Forening. C-5/08 - Infopaq International, ECLI:EU:C:2009:465.

[148] However, this field is not free of problems. On the concept of "academic copyright" see: EBERHARD FEESS, MARC SCHEUFEN, "Academic copyright in the publishing game: A contest perspective." *European journal of law and economics* 42.2 (2016): 263-294; and, also, ROBERTO CASO, *La rivoluzione incompiuta: La scienza aperta tra diritto d'autore e proprietà intellettuale.* (Milano: Ledizioni, 2019), p. 127. A relevant issue which is a matter of debate in this field is represented by the imposition by publishers of article-processing charges (APCs) on authors, to allow publications in Open Access. A post published in June 2020 by some researchers at the University of Leeds (UK) explained how they had to renounce to publish a paper in the prestigious journal "International Journal of Environmental Research and Health" just because of such APCs (https://wash.leeds.ac.uk/what-the-f-how-we-failed-to-publish-a-journal-special-issue-on-failures/). As claimed in the blog of the London School of Economics: "[...] the APC model, while deconstructing the paywall blocking access for readers has merely erected a new paywall at the other end of the pipeline, blocking access to publication for less-privileged authors" (https://blogs.lse.ac.uk/impactofsocialsciences/2020/10/23/open-science-who-is-left-behind/).

[149] Regarding this aspect, it is worth noting the aforementioned Database Directive, Directive 96/9/EC. This Directive, however, does not dispel all the doubts related to data sharing and reuse.

[150] LUCIE GUIBAULT, ANDREAS WIEBE. *Safe to be open, op. cit.,* p. 21.

[151] This aspect will be clarified in Section 4.2.3.3, examining "Licenses for Research Data Sharing and Reuse".

[152] Towards a direction of greater openness and sharing, however, the exception contemplated for text and data mining in the Copyright Directive (Directive (EU) 2019/790) should be remarked. Regarding text and data mining techniques, Recital 8 states: "Such technologies benefit universities and other research organisations, as well as cultural heritage institutions since they could also carry out research in the context of their main activities. However, in the Union, such organisations and institutions are

2.3.4 New Proposed Regulations

Looking forward in the regulatory framework of scientific research, consider the package of legislative proposals, recently released by the European Commission: the proposed Regulation on European data governance (Data Governance Act)[153]; the proposed Regulation on a Single Market for Digital Services (Digital Services Act)[154]; the proposed Regulation on contestable and fair markets in the digital sector (Digital Markets Act)[155].

The Digital Services Act (DSA) aims to implement the legal framework on E-commerce based on the Directive 2001/29/EC, identifying a set of rules on the liability of intermediary service providers. The Digital Markets Act (DMA) intends to regulate the position of gatekeepers in the market, i.e., those who stand between business users and end users. The DSA and the DMA do not seem to have a direct impact on the field of scientific research and on the implementation of the European projects of Open Science.

confronted with legal uncertainty as to the extent to which they can perform text and data mining of content. In certain instances, text and data mining can involve acts protected by copyright, by the sui generis database right or by both, in particular, the reproduction of works or other subject matter, the extraction of contents from a database or both which occur for example when the data are normalised in the process of text and data mining. Where no exception or limitation applies, an authorisation to undertake such acts is required from right holders".

[153] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on European data governance (Data Governance Act)*, COM/2020/767, 25.11.2020, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767.

[154] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, COM/2020/825 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

[155] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on contestable and fair markets in the digital sector (Digital Markets Act)*, COM/2020/842 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN. To this evolving regulatory framework must be added the new Artificial Intelligence Act. See: European Commission, Proposal for a Regulation of the European Parliament and of the Council, *laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM/2021/206, 21.04.2021 ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. In this dissertation, however, this discipline – and the related debate – will not be taken into account.

The situation is different for the Digital Governance Act (DGA). For instance, consider the proposed introduction of a mechanism of data altruism contained in the DGA: it is defined as "[...] the consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking a reward, for purposes of general interest"[156].

Note that the Article 2(10) of the DGA, indicates as an example of data processing for purposes of general interest – i.e., processing for which the mechanism of data altruism can be triggered – explicitly the purposes of scientific research.

Based on this consent, then, the European Commission disciplines a specific mechanism, providing for the conditions, limits and even a registration process[157].

A further examination on the link between the Open Science policies and the data altruism mechanism should be conducted, since there is the possibility that it will become part of a European Regulation: as such, a normative act of general scope, binding in all its elements and directly applicable in the Member States' legal systems, according to the Article 288 of the TFEU[158].

In light of the overview of the legislative frameworks at European level[159], next section illustrates a summary of possible approaches to

[156] The Article 16, DGA.

[157] See Chapter IV of the Data Governance Act (DGA), entitled "Data altruism". The registration mechanism allows entities to be defined as "recognized data altruism organizations".

[158] On the impact of the DGA on scientific research, see: Section 4.4.

[159] It is well known that the variety of factual realities can certainly overwhelm the law, creating unique and unprecedented scenarios. It is possible to identify unknown scenarios involving other legal frameworks. For instance, there may be situations involving other aspects of intellectual property: think about the ongoing debate about vaccines and patents. See: BENJAMIN THAM, MARK JAMES FINDLAY, "COVID-19 Vaccine Research, Development, Regulation and Access." *SMU Centre for AI & Data Governance Research*, 3 (2020). (Besides, in Italy, the AISA Association sent an open letter to the Italian government, February 18, 2021 asking for a COVID-19 vaccine open and public, see: https://aisa.sp.unipi.it/wp-content/uploads/2021/02/AISA_lettera_aperta_covid.pdf). Furthermore, there may also be aspects related to consumer law (e.g., a scenario in which citizens participate in research projects involving private entities). Concerning research data, see, for instance, PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica.* (Trento: Collana della facoltà di Giurisprudenza dell'Università di Trento, 2021), pp. ff. 220. The Author, adopting an holistic approach,

the Open Science governance. After the examination of two extreme and opposite approaches, an alternative, represented by a third way, is proposed.

2.4 Approaches to Open Science Governance

Given the European policies designed to support the Open Science (Section 2.2) and the existing European legal framework in which these policies are placed (Section 2.3), now attention should be drawn to possible approaches to Open Science governance.

Next sections will summarise the various positions in two opposing approaches: on the one hand, those who, having acknowledged the emergence of the Open Science phenomenon, argue that an *ad hoc* law to regulate its different facets should be introduced (Section 2.4.1); and on the other hand, those who, starting from the same premise, i.e. the emergence of Open Science, simply declare its incompatibility with the current legal framework (Section 2.4.2).

Between these polarised positions of supporters and opponents, an alternative third way, represented by legal coordination, is proposed (Section 2.4.3).

2.4.1 Supporters: A Regulation for Open Science

Especially in the early days of Open Science, many scholars approached the topic as genuine supporters, opposing the traditional way of doing science[160]. The process of establishing Open Science was not free of difficulties, especially from a legal point of view.

In light of these tensions, some suggested the identification of an *ad hoc* discipline, arguing that "[A] possible alternative which would

takes into account all legal disciplines involved in the processing of research data, such as competition law. Our study, however, intends to limit its focus to the analysis of the legal issues of these four legal frameworks, considering them the most directly impacting on the governance aspects of scientific research.

[160] See: Section 3.2, in which an all-encompassing definition of the Open Science paradigm is proposed, to take into account the many facets of the issue, as well as its development over time (and the difficulties faced in emerging).

enable us to avoid a double pitfall of a utopian democracy in the digital arena making knowledge freely accessible on the one hand and a society subject to the dangers of sousveillance or improper capturing on the other hand would be a law on open science inscribed in legislation as a fundamental right"[161].

Although the needs on which this proposal is based can be endorsed[162], this solution does not seem feasible, mainly for two reasons: a legal reason and an economic one.

First, from a legal point of view, the emergence of the paradigm of Open Science, as a new way of conducting science, does not arise from scratch: it must be included in a context extremely tangled of provisions and the main ones have been examined in Section 2.3. Think about the major legal issues, e.g., the protection of personal data in the context of Open Science or copyright issues: it seems fair to admit that identifying an *ad hoc* law to deal with these issues would represent a duplication of disciplines. The risk is that one of the fundamental principles of law, the principle of legal certainty, would be affected and the effectiveness of the provision itself would be undermined.

In addition, the identification of a specific law, understood in terms of hard law, to solve and clarify several of legal issues related to a given phenomenon runs the risk of being too detailed and anchored to a precise historical time. The timeframe of the introduction of the legislation often differs considerably from the timing of technological

---

[161] RENAUD FABRE, *New challenges for knowledge: Digital dynamics to access and sharing*. (Hoboken, NJ: John Wiley & Sons, 2016), 142. The term "sousveillance", used by the Author, indicates: "[…] the uncontrolled reuse of information through, among others means, social media", p. 142, from: DOMINIQUE QUESSADA, "De la Sousveillance." *Multitudes*, 1/40 (2010): 54-59.

[162] The Author, Renaud Fabre, the then director of the "*Direction de l'information scientifique et technique*" of the CNRS (*Centre National de la Recherche Scientifique*), i.e. the largest public research organization in France, bases this assumption on the French experience and, in particular, on the important presence of platforms in the field of science, which he claims are assuming an increasingly persuasive role. He states, in fact, that "Science in France is thus currently subject to intellectual property law: the platforms control all data, including its reuse", RENAUD FABRE, *New challenges, op. cit.*, p. 146.

development: there is a risk of making the law completely outdated and obsolete in a short span of time[163].

Second, consider the economic reason. There is a total lack of political intention in identifying a law of Open Science, given the costs that must necessarily be taken into account by establishing a new discipline[164]. The identification of a new piece of legislation, especially in terms of hard law, necessarily requires the allocation of investments to support its implementation. Despite the fact that European institutions are investing – and have invested, in the past years[165] – considerable amounts of money in the promotion of Open Science, no investment has ever been proposed for the release of an Open Science law *per se*.

2.4.2 Deniers of Open Science Feasibility

While some scholars, fully supporting Open Science, claim that an *ad hoc* law is needed to solve all legal problems in a unified approach, on the opposite side there are those who simply argue that the Open Science paradigm is incompatible with the current European legal framework. This incompatibility has been specifically discussed with regard to the European data protection framework. Phillis and

[163] Ugo Pagallo gives a clear example, which is that of the 2000 European Electronic Money Regulation. The release of the Regulation just preceded the launch of Paypal, which, from both a technological and a legal point of view, completely transformed the scenario in the field of electronic purchases, making the then recent European directive worthless paper. See: UGO PAGALLO, "From automation to autonomous systems: A legal phenomenology with problems of accountability." *26th International Joint Conference on Artificial Intelligence, IJCAI 2017* (2017): 17-23.

[164] There is a strand of research on this subject, in the field of economics, which is specifically concerned with investigating the economic impact of the introduction of new derogatory regulations. See: ERIK VAN DER MAREL, *et al.*, "A methodology to estimate the costs of data regulations." *International Economics* 146 (2016): 12-39.; and also reports provided by the European Centre for International Political Economy (ECIPE), such as: HOSUK LEE-MAKIYAMA BADRI NARAYANAN GOPALAKRISHNAN, "Economic cost of ex ante Regulations", ECPIE, 2020, https://ecipe.org/publications/ex-ante/.

[165] Consider that at European level, investments in Open Science are always anchored at the framework programmes for investment in research and innovation: "Under Horizon 2020, the EU Research and Innovation programme for the 2014-2020 period, €600 million was allocated to setting up EOSC. In the post-2020 period, EOSC will be supported by the Horizon Europe programme, as the principle of open science will become the modus operandi of Horizon Europe.", see: https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud.

Knoppers[166], for instance, argue that "[D]espite the appearance that the GDPR strikes the proper balance between accommodating scientific research and securing individual rights and dignity, the tension between open science and data protection goes to the very core of the two movements"[167], adopting an interpretation of "data protection as a legal limit of open science"[168]. Although it is deemed that, even here, there are some acceptable premises[169] and some valid suggestions[170], it is not believed there exists a total divergence between the architecture of the European data protection discipline and the paradigm of Open Science, mainly for two reasons.

First, if apparently the aspects of sharing and openness of this paradigm of science seems to conflict with the intent to protect the personal data of individuals, as enshrined in the Article 8 of the Charter of Fundamental Rights of the EU, a further examination shows that these two aspects tend to converge. In fact, consider the Article 1 of the GDPR, concerning the purposes of the Regulation: the second paragraph provides the aim of protecting the fundamental rights and freedoms of individuals, with particular regard to the right to the protection of personal data; but the next paragraph emphasises

---

[166] MARK PHILLIPS, BARTHA M. KNOPPERS, "Whose Commons? Data Protection as a Legal Limit of Open Science." *The Journal of Law, Medicine & Ethics* 47.1 (2019): 106-111.

[167] MARK PHILLIPS, BARTHA M. KNOPPERS, "*Whose Commons?", op. cit.*, p. 109.

[168] MARK PHILLIPS, BARTHA M. KNOPPERS, "*Whose Commons?", op. cit.*, p. 108.

[169] For instance, attention is raised to two risks of Open Science such as the risk of exacerbating inequities within science ("The potential pitfalls of open science include exacerbating existing inequalities, by supporting the development of expensive new diagnostics and treatments that are practically available only to the stratum of the population who can afford them, while putting already-disadvantaged individuals and groups at risk of harms, such as discrimination and stigmatization.", p. 107) and the risk of falling into vicious circles of data privatization ("A related risk is the de facto privatization of personal data, by organizing data in a manner that benefits only those who possess sufficient resources to allow them to usefully analyze them, thus transforming public funding of open science into an indirect subsidy to private industry.", p. 107), which must surely be taken into account when analyzing the developments of European projects and their practical implementation. On this aspect, see: Section 5.3.1, but also Section 3.4.2.

[170] Indeed, the Authors suggest that more interplay is needed between the field of data protection and Open Science, identifying an attempt in that direction in the drafting of codes of conduct for medical research. MARK PHILLIPS, BARTHA M. KNOPPERS, "*Whose Commons?", op. cit.*, p. 110.

that "the free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data"[171]. The European discipline on data protection attempts to strike a fair balance between the need for protection of personal data and the freedom of movement of data, in accordance with the already mentioned fifth European freedom, i.e., the freedom of free flow of knowledge[172]. Similarly, the Open Science aims to affirm a scientific research based on the openness of each phase of the research cycle, but always respecting the formula "as open as possible, as closed as necessary": it emerges the same effort to balance the different interests at stake.

Second, although the implementation of Open Science projects has been underway for some years now, as seen in Section 2.2, the development is still in the early stages. The institutions (at all levels) and the other stakeholders (universities, associations, researchers, private entities and citizens) have the duty to participate in the design of these projects and their developments. In other words, nothing is crystallised and it is up to the actors involved to decide how to shape this new way of conducting science in a manner that is respectful of European values and principles.

It is important to avoid the conceptual trap of considering technology as something that creates itself, with an unpredictable impact. Although the impact of ICTs in specific fields may determine the emergence of new problems, this does not mean that there can be no control over its development. Technological development is still a product of human action, following a human vision or idea. As Juan Carlos De Martin, the co-founder of the Nexa center for Internet and society[173], argues, "[I]t is essential, however, that from the very

[171] The Article 1(3), GDPR.
[172] In particular, see: Section 2.2.1.
[173] https://nexa.polito.it/about-en.

beginning of this philosophical and political debate technology is conceived in a new way: no longer a mere set of tools with variable effectiveness, but an immanent component of humanity"[174]. The Open Science intends to encourage greater openness in scientific research using ICTs and there is a duty of the actors involved to guide its development.

The same need was expressed by the Nobel Prize winner for physics Arno Penzias, who, in 1989, discussing on the impact of technology, claimed: "[M]achines need direction from human minds, and human minds need inspiration from human leaders"[175]. As a result, it is now time to investigate the possibility to choose a third alternative approach.

2.4.3 A Third Way: Legal Coordination

The alternative model of Open Science governance that is proposed here starts from the premises examined at the beginning of this chapter, in Section 2.1: on the one hand, science is changing due to the impact of technology and this change is considered in terms of Open Science; on the other hand, European institutions assume the commitment of promoting this new paradigm of science.

Then, consider the European legal framework in which these premises are placed: a *plethora* of provisions and prescriptions that sometimes risk overlapping[176].

---

[174] "E' essenziale, però, che fin dall'inizio di questa riflessione filosofica e politica si pensi alla tecnologia in modo nuovo: non più mero insieme di strumenti dall'efficacia variabile, ma componente immanente dell'umanità.", in JUAN CARLOS DE MARTIN, Opening speech of the first edition of "Biennale Tecnologia", Turin, 2020, https://www.youtube.com/watch?v=KmZVKhmiGyw. [Translation from the Italian original text].

[175] ARNO PENZIAS, *Ideas and information: Managing in a high-tech world* (New York: WW Norton & Co., Inc., 1989), p. 219.

[176] On this aspect, see: UGO PAGALLO, "La grande trasformazione: datificazione della società, tutela dell'ambiente e rischi e opportunità dell'innovazione digitale." in MASSIMO DURANTE, UGO PAGALLO, (eds.), *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società.* (Milano-Udine: Mimesis, 2022), pp. 123-140.

Rather than proposing a specific regulatory regime for Open Science, or simply deeming European strategies incompatible with the current legislative provisions, it is suggested to adopt a perspective of coordination between different regulatory frameworks[177], with the necessary adaptation to the specificities determined by the Open Science scenario.

The need to conceive coordination stems from the complexity of the scenario under consideration. In particular, this complexity can be structured in three dimensions, explored below: (i) levels; (ii) systems; and (iii) actors.

2.4.3.1 Levels

The first dimension of the complexity of the scenario under investigation is represented by the multiplicity of levels of regulation that coexist. The European projects on Open Science cannot be placed in a normative vacuum[178]. In addition to the European legal disciplines explored in Section 2.3, national dimensions must also necessarily be taken into account. For instance, think about the transposition of directives by the different European Member States, or even the autonomous domestic regulation resulting from national policies.

Then, the role of the Courts also adds complexity to the scenario. Different levels can also be found in the field of jurisprudence, from the rulings of national courts to those issued by the ECJ or the European Court on Human Rights (ECHR)[179].

Finally, the relationship between the primary rules and the secondary rules of the legal system should be taken into account. As analysed by Pagallo, "[…] the aim of the primary rules is to directly govern human and social behaviour either through techno-regulation,

[177] This need, however, appears to be very recurring, considering that the European Commission, in the proposed DGA, in Recital 25, reiterates the need for harmonisation, at European level, of the provisions on data sharing and reuse.

[178] Ugo Pagallo, "Algo-rhythms and the beat of the legal drum." *Philosophy & Technology* 31.4 (2018): p. 512, doi: 10.1007/s13347-017-0277-z.

[179] See, for instance, Section 4.2.4.2.

e.g., some variants of the principle of privacy by design, or the manifold means of law as a meta-technology, such as achieving particular effects with hard laws […]; administrative regulation (e.g., ISO standards); or soft law […]"[180]. In parallel, secondary rules can also play a key role in dealing with technological development[181].

2.4.3.2 Systems

The second dimension that determines a complex scenario is the coexistence of different regulatory systems, e.g., European and Member States' legislations; social norms of scientific community; self-regulation of private actors involved in science, etc.

Considering that the origins and sources of power of the regulatory systems are different, the risk of unfolding tensions between them is high. Lessig describes this situation in terms of competition between different normative systems[182]. On this aspect, Pagallo explains that: "[T]he relation between law and technology should […] be grasped as the interaction between competing regulatory systems that not only may reinforce or contend against each other, but against further regulatory systems, such as the forces of the market and of social norms. Every regulatory system claims to govern social behavior by its own means and can even render the claim of another regulatory system superfluous"[183].

[180] UGO PAGALLO, "The legal challenges of big data: putting secondary rules first in the field of EU data protection." *Eur. Data Prot. L. Rev.* 3 (2017): pp. 39-40, doi: 10.21552/edpl/2017/1/7.

[181] For instance, considering the secondary rules of change: "The aim of the secondary rules of change is to allow the creation, modification, and suppression of the primary rules. This aim can either concern the substitution of a given regulation, […] or they can concern mechanisms of legal flexibility", in UGO PAGALLO, *"The legal challenges of big data", op. cit.*, p. 40.

[182] In Lessig's interpretation, in particular, considering cyberspace, the competition between different regulatory systems is outlined, in the first place, with the code: "[...] there is regulation of behavior in cyberspace, but that regulation is imposed primarily through code. What distinguishes different parts of cyberspace are the differences in the regulations effected through code.", in LAWRENCE LESSIG, *Code and other laws of cyberspace.* (New York: Basic Books, 1999), p. 20.

[183] UGO PAGALLO, "LegalAIze: Tackling the Normative Challenges of Artificial Intelligence and Robotics Through the Secondary Rules of Law." In: MARCELO CORRALES, *et al.,* (eds) *New Technology, Big Data and the Law. Perspectives in Law, Business and Innovation* (Singapore: Springer 2017): p. 285, doi: 10.1007/978-981-10-5038-1_11.

These general considerations on the interplay between law and technology can be applied to the specificity of this study. The governance of scientific research should take into account the different regulatory systems that it entails. Consequently, according to Pagallo, law should not be seen as the only regulatory system. This is particularly relevant when enacting hard law in the field of scientific research[184].

This second dimension of complexity, related to the systems, is closely linked to the third dimension, represented by the actors involved.

2.4.3.3 Actors

The third dimension of the coordination is about the multitude of actors involved. In the context of Open Science there are different levels of actors and different typologies of them.

The different levels of actors refer to the international, European and national ones. If the European level has already been described in Section 2.2, with regard to the international level, instead, think about the joint effort of UNESCO, WHO, CERN and the Office of the United Nations High Commissioner for Human Rights, for the drafting of the first recommendation on Open Science[185].

However, at international level there are many other actors. See, for instance, the Research Data Alliance (RDA): it is a community, established in 2013, jointly by the European Commission, the United States Government's National Science Foundation, the National Institute of Standards and Technology, and the Australian Government's Department of Innovation, pursuing the goal of "[...] building the social and technical infrastructure to enable open sharing

---

[184] In this regard, the issue of the risk of lack of enforceability of a provision contained in the Open Data Directive will be discussed below, precisely in light of the multiplicity of regulatory systems. See: Section 4.2.3.2.

[185] See: Section 2.1.

and re-use of data"[186]. The RDA outputs and recommendations have a significant impact on the development of projects of European or national institutions around the world.

There is also a further level of actors of Open Science, as relevant as the first two, i.e., the national actors. The role of Member States is crucial in the implementation of the European policies, as seen for example in relation to the Recommendation n. 790/2018 of the European Commission[187]; or in relation to the transposition of European directives in the field of copyright; or, also, in the implementation of the regulation of the processing of personal data for scientific research purposes, given the wide discretion granted to Member States by the Article 89 of the GDPR[188].

In this third dimension of the complexity, represented by the multiplicity of actors, there are not only different levels of actors, but also different typologies. Alongside the traditional actors of research, such as professors, researchers, students, public funding bodies and scientific publishers, several new actors appear in the context of Open Science. Even in relation to science, as in many other areas of our lives marked by the digital revolution, intermediaries have a new and fundamental importance. In the field of scientific research, the intermediaries are the providers of services for the collection, analysis, processing and storage of research data and companies that manage databases of publications.

In addition, very often research projects can involve, in various ways, private entities (companies, private research centers, etc.), which represent a further type of actors[189].

Furthermore, a fundamental role is played by associations. For example, in Italy, the AISA (Italian Association for Open Science) plays

[186] https://www.rd-alliance.org/about-rda.
[187] See: Section 2.2.4.
[188] On this aspect, see: Section 5.2.2 and Section 5.5.1.
[189] On the participation of private actors in the scientific research process, see: Section 5.1; but also, Section 6.3.

a very active role in the debate on science at the national level, as well as a very effective function at the political level[190].

Considering this complex scenario, it is essential to ensure a high level of coordination, between different levels of regulation (first dimension), different normative systems (second dimension) and, finally, different actors (third dimension).

2.6 Conclusive Remarks

This chapter has investigated the governance of the European scientific research, in order to describe the scenario in which this investigation is set. At the basis, there are two main premises: a methodological one and an institutional one.

The first premise is represented by the impact of ICTs on scientific research and the consequent significant changes. This premise on how the methodology of science is changing has been represented by the emergence of the paradigm of Open Science: a new way of conducting scientific research, in all its aspects, which, due to the technology, is becoming more open, collaborative, and global.

The second premise, the institutional one, represents the reaction of the European institutions to this transformation of science: the European Union seems committed to promote and foster the Open Science paradigm.

The investigation of these two premises clarified the scenario in terms of technology, science, and law: while the first premise connects technology and science, the second premise connects science and law.

The investigation proceeded by mapping initiatives and interventions that the European institutions put in place in the field of

[190] In February 2021, the Italian Ministry of University and Research (MIUR) published the PNR 2021-2027, "*Programma Nazionale della Ricerca*" (National Research Program), guiding national policies on scientific research for the following years. AISA has represented an essential contribution – as an expression of voluntary associationism – in the drafting of PNR, with specific reference to the section of Open Science.

Open Science. The complex *corpus* of European documentation has been divided into different phases, in order to follow the evolution of the Open Science at the European level.

Three main considerations emerged from the systematisation of the European institutions' interventions on Open Science.

First, although the European institutions started to put in place comprehensive interventions in the field of Open Science in 2015, the concept of open and collaborative science is believed to date back much earlier in the European architecture. The concept of open and shared science was born, in the European context, with the establishment of the European Research Area (ERA) in the early 2000s. The ERA is the European area of free exchange of knowledge and free movement of researchers. Although the formal introduction of the Open Science concept within the architecture of the European Union is considered to be in 2015 with the Digital Single Market initiative, it is argued that its substantial origin lies in the establishment of the ERA in the early 2000s and the identification of the fifth European freedom of movement, i.e., the free circulation of knowledge. In other words, the concept of Open Science appears strongly embedded in the foundations of the EU architecture.

The second consideration that emerges from this mapping effort is that recently the EU context of the Open Science has been reshaped: on the one hand, the year 2020, for certain reasons[191], and on the other hand the establishment of the Horizon Europe programme, lead to a proper recognition of the Open Science phenomenon by the European institutions[192].

The third consideration is related to the European Open Science Cloud (EOSC), a project that represents the cornerstone of the current European Open Science policies. The dual nature of the project is

[191] The identified reasons are: (i) the concretisation of what it is identified as the "von der Leyen Doctrine"; (ii) the COVID-19 pandemic; (iii) the developments on the governance of EOSC project; and, finally, (iv) the reform of the European Research Area (ERA). See Section 2.2.4.

[192] This aspect is further analysed in Section 3.2.

stressed. On the one hand, it seems in accordance with the process of technological convergence, since it focuses on research data, but also on research infrastructures and the computational power required in science[193]. On the other hand, especially in the most recent stages of its implementation, it seems to embody what has been defined as the "von der Leyen Doctrine". The EOSC, in fact, seems to have become a means of unlocking the economic value of research data, for the benefit not only of researchers but also of companies and the public sector.

After investigating the scenario, and systematising the European interventions on Open Science, attention was drawn to the European legal frameworks involved: (i) the discipline of Open Data; (ii) the discipline of data protection; (iii) the set of rules on copyright and the ownership of data; (iv) the set of Regulation proposals recently presented by the European Commission.

This analysis was functional to identify a sound approach to the EU Open Science governance. In this regard, an approach aimed to the cooperation has been proposed, designed to make the components of this complex scenario interact, i.e., the different levels, systems, and actors.

Chapter 2 investigated the European scientific research governance, analysing the evolution of the Open Science policies in EU. So far, in this dissertation, the Open Science has been identified as the openness of every phase of the research cycle, generated by the ICTs, leading to a more open, collaborative, and global science.

However, beyond this general description, many dimensions of Open Science have emerged from the analysis. For this reason, it is fair to wonder: what is currently meant by the Open Science paradigm?

Next chapter intends exactly to answer this question.

[193] On the relationship between Open Science, e-Infrastructures, and computational power, see: Section 6.2.

Chapter 3

The Open Science Paradigm

Chapter 3 discusses the evolution of the Open Science concept, in order to provide an interpretation of the current paradigm of scientific research.

In Chapter 2, the analysis of the governance of scientific research in the European Union showed a complex scenario, characterised by the co-existence of multiples levels of interventions, regulatory systems, and actors. In addition, the evolution of the EU Open Science policies revealed the existence of many and different dimensions of the phenomenon.

In order to address the main legal challenges of the Open Science projects developed at European level, it is essential to clarify some notions. Several of the concepts underlying the Open Science paradigm are relatively recent, others are very broad: this can lead to misrepresentation. This potential lack of clarity in drawing out the terms may affect the initiatives.

Initially, Section 3.1 describes the notion of Open Science. The term Open Science represents an "umbrella concept" and encompasses many different dimensions: it is essential to see how scholars and institutions have dealt with these various facets so far. There is a real need to define the Open Science paradigm, and its related concepts. The complexity of this task is widely recognised. The EOSC Strategic implementation plan acknowledges this definitional complexity: "There is a parable of the blind men and the elephant, which originated in ancient India. It is the story of a group of blind men who have never come across an elephant before and who are to describe the elephant by respectively touching one – only one – different part of the elephant. Each blind man feels a different part of the elephant's body, such as the tail, the trunk, one leg. They describe the elephant based on their different experiences and of course, the descriptions are entirely different from one another. If you ask a room full of people what EOSC is, you'll get a room full of different answers. It's such a large-scale initiative and ambitious mission that it's like an elephant – everyone sees a different part and few see the big picture"[1]. The metaphor proposed for the EOSC also fits well with the more general concept of Open Science and its many facets. For this reason, it is relevant to dwell on the meaning(s) that can be attributed to Open Science.

Then, Section 3.2 proposes an interpretation of the current scientific research paradigm, conceiving Open Science as a process. Adopting Floridi's method of Levels of Abstraction (LoA), the intention is to provide a comprehensive representation of the phenomenon: this goal is attained by identifying the constitutive elements of the Open Science paradigm, embodied in five observables (Resources; Actors; Methods; Tools; Benefits).

This representation of Open Science as "Open Scientific Research Process" enables to achieve a twofold result: first, by outlining the Open Science paradigm in its complexity, it is possible to move beyond the taxonomies and classifications proposed so far, which focuses only on certain aspects of the phenomenon; second, the identification of the constitutive elements, which can be modulated through regulation, generates a considerable impact on the entire scenario of scientific research.

Section 3.3 concerns the analysis of the framework of fundamental rights related to science, at international and European level. The aim is to establish the legal basis of the Open Science: the foundations are set in the framework of fundamental and human rights related to science.

---

[1] Sarah Jones, Jean-François Abramatic, "European Open Science Cloud (EOSC) Strategic Implementation Plan." *Publications Office of the European Union* (2019), p. 4, doi: 10.2777/202370.

The detailed disciplines of the different issues related to the Open Science implementation may undergo changes and renovations according to the needs of varying circumstances and short-term events. On the contrary, the fundamental and human rights remain unchanged pillars of the legal and social architecture: a constant guide and a safeguard for every individual.

The Open Science paradigm, as an expression of the human and fundamental rights linked to the sphere of science and knowledge, can become the means by which the European Union can emerge and be a leading player in the field of scientific research[2].

In this regard, the Italian jurist Stefano Rodotà stated: "[B]y following the path of rights, we can really discover another Europe, very different from the dominant economic Europe and the evanescent political Europe. It is precisely the Europe of rights, too often neglected and pushed into the shadows. A Europe that is annoying for those who want to reduce everything to the market dimension and that, instead, should be evaluated just when the winds of anti-Europeanism are blowing strong, showing citizens that it is precisely in the field of rights that the European Union can offer them an 'added value', and therefore a very different picture from the unwelcome and unacceptable one that identifies it with the continuous imposition of sacrifices"[3].

Afterwards, Section 3.4 deals with the implementation of the Open Science paradigm.

The analysis of the Open Science concept is structured in two stages: the development stage and the deployment stage[4].

---

[2] On this aspect, see: Section 1.1.2 "The institutional Premise". There, the concept of the so-called "Brussels effect" is evoked.

[3] STEFANO RODOTÀ, *Il diritto di avere diritti.* (Roma-Bari: Laterza & Figli Spa, 2012), p. 38. [Translation from the Italian original text].

[4] Once again, in adopting the distinction between development and implementation stages, the reference goes to Luciano Floridi. The philosopher, in fact, uses this distinction in the investigation of Artificial Intelligence (AI) progress. See, in particular: LUCIANO FLORIDI, "What the near future of artificial intelligence could be." The 2019 Yearbook of the Digital Ethics Lab (Cham: Springer, 2020), 127-142, doi: 10.1007/s13347-019-00345-y; JESSICA MORLEY, LUCIANO FLORIDI, LIBBY KINSEY, ANAT ELHALAL, "From what to how: an initial review of publicly available AI ethics tools, methods and

The former, the stage of development, can be represented by the analysis of (i) the evolution of the Open Science concept, illustrated in Section 3.1; (ii) the proposed interpretation in terms of "Open Scientific Research Process", presented in Section 3.2; and (iii) the foundation in the human and fundamental rights framework, described in Section 3.3.

The second stage, i.e., the deployment, concerns the Open Science implementation. Investigating it, Section 3.4 focuses on the role of the scientific community and on the possible pitfalls that may emerge in this delicate phase.

Finally, the chapter ends with some conclusive remarks, in Section 3.5.

3.1 The Concept of Open Science

The term Open Science does not have a commonly accepted definition. Besides, even the concept of science itself has been extremely debated for centuries. This chapter does not aim to review centuries of epistemology. The intention, instead, is to clarify the boundaries of the phenomenon of Open Science, as a representation of the new paradigm of science that is emerging in these decades, i.e., a new way of conducting science, shaped by the impact of digital ICTs.

This analysis fulfils a dual function: first, it serves to address the legal issues involved in the implementation of the EU policies; second, it fills a gap in the literature, in which a comprehensive mapping of the Open Science evolution is missing so far[5].

As a preliminary consideration, notice that, in this study, the word "science" – whether or not associated with the adjective "open" – refers

research to translate principles into practices." *Science and engineering ethics* 26.4 (2020): 2141-2168, doi: 10.1007/s11948-019-00165-5.

[5] There are a lot of partial analysis of the concept, but, so far, a comprehensive study of the evolution of the concept of Open Science is not found, although the expression is widely used. Moreover, what is lacking is an all-encompassing mapping aimed at identifying the constituent elements of the phenomenon, which in this dissertation assume the role of observables. See, the following Section 3.2.

to all the domains of knowledge: it includes the so-called STEM, i.e., science, technology, engineering and mathematics, as well as the so-called HSS, i.e., humanities and social sciences.

Next section explores the origin of the term Open Science with the intention of tracing its evolution. Afterwards, the most well-known definitions are presented in order to identify the most characteristic features of the phenomenon.

3.1.1 Origins: Modern Science

Historically, the origin of the term Open Science can be traced back to the 17th century, with the emergence of Modern Science. At that time, however, the concept of Open Science was partially different from the current one: science was defined open as opposed to the traditional closed science, understood as secret or occult science[6].

This transformation, from a secret and occult science towards the openness, can be understood through the analysis of five elements, explored below: (i) publicity; (ii) need of verifiability; (iii) technology; (iv) new system of publication; (v) copyright.

The identification of these five features will be instrumental to grasp the interpretation of the current paradigm of Open Science, proposed in Section 3.2.

(i) Publicity. As stated by the historian of science Paolo Rossi, who extensively studied the birth of Modern Science in Europe, the intention to impose the openness of science emerged starting from the first scientific revolution: it happened when the secrecy became a non-value for scientists[7]. This process to obtain a science characterised by openness was made possible by the publicity: science became a matter of public interest[8].

---

[6] Roberto Caso, *La società della mercificazione e della sorveglianza: dalla persona ai dati: casi e problemi di diritto civile.* (Milano: Ledizioni, 2021), p. 309.
[7] Paolo Rossi, *La nascita della scienza moderna in Europa.* (Roma-Bari: Laterza, 1999), p. 34.
[8] "Publicity is one of the basic features of Modern Science: research cannot be secret any longer, because

It is significant that, even today, although much has changed since the science of modern age, the concept of "public interest" still plays a central role in this domain: in fact, one of the features still persisting in the current interpretation of Open Science is the publicity[9].

The historical origin of Open Science in the modern age has also been investigated by the historian of economics Paul David, which similarly brings out this aspect: "[A]n essential, defining feature of Modern Science thus is found in its public, collective character, and its commitment to cooperative inquiry and free sharing of knowledge. While to most of us the idea of science as the pursuit of 'public knowledge' seems a natural, indeed a primitive conceptualization, it is actually a social contrivance; and by historical standards, a comparatively recent innovation at that, having taken form only as recently as the sixteenth and seventeenth centuries. [...] the late sixteenth and early seventeenth centuries also witnessed a transition from the previously dominant *ethos* of secrecy in the pursuit of Nature's Secrets, which gave way to a new set of norms, incentives and organizational structures"[10]. As a result, according to David, "[T]hese institutional transformations reinforced scientific researchers' commitments to rapid disclosure and wider dissemination of their new discovers and inventions"[11].

(ii) Need of Verifiability. Another fundamental aspect of Modern Science was the search for a severe linguistic correctness and terminological clarity. This feature of linguistic accuracy was not

---

it needs the community of scientists to 'provide for the social validation of scientific work'.", see: MARIA CHIARA PIEVATOLO, "Open science: human emancipation or bureaucratic serfdom?". *SCIRES-it* (2019): p. 36, doi: 10.2423/i22394303v10Sp35.

[9] See: Section 3.2. In addition, in the field of data protection, the European lawmaker, in accordance with the major literature, has grounded the processing of personal data for scientific research purposes on the legal basis of public interest (Article 6(1)e of the GDPR). On this aspect, see: Section 5.3.1.

[10] PAUL A. DAVID, "The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution." *Capitalism and Society* 3.2 (2008): pp. 10-11, doi: 10.2202/1932-0213.1040.

[11] *Ibid*.

required merely for the sake of rigour but was instrumental: it served to make scientific theories fully communicable and scientific experiments replicable, and thus verifiable[12].

A further comparison with today occurs. In the modern age, accuracy was the means by which ensuring correctness and verifiability: the new science of that time generated a need for verifiability, which was satisfied through the technology of writing. Today we have a similar problem of verifiability in scientific research based on the processing of huge amounts of data, the Big Data[13]. In order to guarantee the reproducibility of the results[14] – and thus the correctness of the research – we again rely on a technological standard: the identification of set of rules, such as the FAIR Data principles[15], meet the same need for accuracy that existed centuries ago.

A related interesting aspect, on which David dwelt on, is the link between the public reputation of the scientist and the publicity of knowledge[16]. In modern society, characterised by multiple sources of power, both religious and civil, where the press disseminated knowledge more easily[17], it became important to find a system of validation of results. This was put in place by the scientific community itself in order to guarantee the validity of the producers of knowledge, and thus the validity of knowledge itself[18]. Alongside the need for verifiability of scientific results – or, more correctly, as part of it – there emerged a need for evaluation of the scientists themselves.

---

[12] PAOLO ROSSI, *La nascita della scienza moderna*, *op. cit.*, p. 27.

[13] Consider that this aspect is further illustrated in Section 3.2, analysing the observable "Methods".

[14] On the relevance of the reproducibility and the related issues (e.g., the so-called "crisis of reproducibility") see: Section 4.1.1. There, the reproducibility is identified as one of the constitutive elements of the notion of "research data".

[15] On the concept of FAIR Data Principles, see: Section 2.2.3 and also, in particular, Section 4.1.2.

[16] PAUL A. DAVID, *"The Historical Origins of 'Open Science': An Essay", op. cit.*, pp. 69-72.

[17] The dissemination of knowledge was easier than in the previous historical period, but it must be stressed that at that time it was related to a small circle of people: knowledge was still relevant only to the *elite* of societies.

[18] MARIA CHIARA PIEVATOLO, "L'università e le sue crisi: una riflessione storica". *Bollettino telematico di filosofia politica* (2012).

(iii) Technology. In Modern Science, the technological means by which it was possible to bridge the gap between the old traditional, closed, secret science and the new open, public science was the printing press, invented in 1455. The consequent and increasing dissemination of books was precisely the element that enabled the shift from the traditional to the emerging science.

In the profound changes that Modern Science faced, the technological development was a fundamental aspect: printing had an incredible impact on the way science was done and the way its results were shared.

This technological advancement, represented by printing, is closely linked to the feature analysed below: the emergence of a new system of publication for scientific results.

(iv) New System of Publication. In this scenario, characterised by the emergence of Modern Science, a further fundamental element should be noted: the emergence of a new system of publication. In 1665, the first ever scientific-academic journal was published for the first time: "The Philosophical Transactions", edited by Henry Oldenburg, as an initiative of the London's academic association, the Royal Society[19].

The aim was to provide a coherent collection of scientific articles containing the most relevant innovations in the philosophy of nature or the major scientific debates of the time. As made clear by Roberto Caso, this was the period in which the concept of publicity became a major tenet of science: "[T]he scientific journal is from then on a public record of the priority of scientific discovery"[20].

---

[19] AILEEN FYFE, "The production, circulation, consumption and ownership of scientific knowledge: historical perspectives." *CREATe Working Paper* 4 (2020): p. 5-7. doi:10.5281/zenodo.3859492. The Author proposes an interesting historical investigation, starting from the foundation of the first scientific journal "The Philosophical Transactions", which shows that the private system of publication management is not an immutable reality, which has always existed. On the contrary, the Author focuses on the period when the control of scientific publications was decentralised.

[20] ROBERTO CASO, *La società della mercificazione, op. cit.*, p. 310. [Translation from the Italian original text].

This event is remarkable because, albeit without knowing it, the publication of "The philosophical transactions" set the course for the future development of scientific publications. *Mutatis mutandis*, this model of disseminating the results of scientific research has characterised the following centuries, without particularly varying.

(v) The Emergence of Copyright. The last crucial feature in understanding the emergence of the concept of Open Science occurred later. On 10 April 1710, the Copyright Act, promulgated by Queen Anne of England in 1709, came into force. It was the first comprehensive copyright statute[21]. The relevant aspect is that, from this date, the rights to a work copy passed from the licensee to the author of the work[22]. It was the first real protection of the outcome of intellectual work. The emergence of copyright law, in addition, is an example of how a change in the legislative framework can radically transform the social structure of an entire category, or more precisely, of an entire society.

The greatest credit of this legislative intervention was certainly that it understood the needs of society and converted them into legally binding rules. In other words, the first copyright law was not an imposition of a top-down approach, without taking into account the demands of the actors involved. Rather, the normative system of law, in this instance, was able to express a need of its time, i.e., the authors' need for recognition.

---

[21] MARCO RICOLFI, "Intellectual Property Rights and Legal Order". *Global Jurist* 2.1 (2002): p. 2, doi: 10.2202/1535-1661.1052; ROBERTO CASO, "Alle origini del copyright e del diritto d'autore: spunti in chiave di diritto e tecnologia." in UMBERTO IZZO (ed.), *Alle origini del copyright e del diritto d'autore: tecnologia, interessi e cambiamento giuridico*, (Roma: Carocci, 2010), p. 249.

[22] From 1662, with the Press Act, it was established that the Royal Society held the privilege of licensing print texts. As Francesca Di Donato observes, before Queen Anna's Statute was passed, the role of the licensee was more important than that of the author. See: FRANCESCA DI DONATO, *La scienza e la rete: l'uso pubblico della ragione nell'età del web*. (Firenze: Firenze university press, 2009), p. 18.

In this section we shed light on how and when the term Open Science arose[23]. This historical overview of the original concept of Open Science has been developed on the basis of five key elements: publicity; need of verifiability; technology; new system of publication; and establishment of copyright. These various elements, identified as underpinning the emerging Open Science in the age of Modern Science, undergo several changes that will result in the current Open Science paradigm.

Next section investigates one of these major changes, which triggers the current definition of Open Science: the Open Access movement.

3.1.2 Beginning: The Open Access Initiatives

For a comprehensive theorisation of the concept of Open Science as it is understood today it is necessary to make a considerable leap forward, to the early 2000s, when the digital revolution was beginning to emerge considerably. At the beginning of the new century, the Open Access movement appeared: it was the first step towards the establishment of the current Open Science paradigm[24].

---

[23] Consider that this meaning of Open Science as opposed to an occult or secret science still persists until the second half of the 20th century. Chubin in 1985, starting with a critique of Merton, argues that the needs of science (but perhaps rather his vision of scientists) do not coincide with the needs of society, to the prejudice of openness: "In the process of weighing the costs and the benefits of observing certain norms and ignoring others, the tradeoffs are clarified. A measure of autonomy is compromised to heighten accountability, candor is abridged to preserve confidentiality, and dissemination of research results delayed by the prospect of commercial profitability", see: DARYL E. CHUBIN, "Open science and closed science: Tradeoffs in a democracy." *Science, Technology, & Human Values* 10.2 (1985): p. 79.

[24] It should be noted that some Scholars underlined the role of the "Free Software movement" in the development of the Open Science concept: "In connection with developments in software technology, a growing push for the sharing and collaborative improvement of software source code emerged. Computer programmers and developers started sharing software with the aim to mutually learn and improve computing. Examples of software released freely were the TeX typesetting system, the Netscape Communicator Internet suite (which subsequently led to Mozilla Firefox) and Linux. In order to encourage and support this open source movement, in 1998 – shortly after the Netscape source code was released – the Open Source Initiative (OSI) was founded, as an educational, advocacy, and stewardship organisation. The ideal behind the initiative was that a participatory approach to creating and improving software technology would ultimately benefit the progress of computing. This ideal highly resonates with the values underlying the Open Science transition […].": in ANNA BERTI SUMAN, ROBIN PIERCE, "Challenges for citizen science and the EU open science agenda under the GDPR." *European Data Protection Law Review* 4 (2018): 286, doi: 10.21552/edpl/2018/3/7. In this dissertation, the relevant role of the free software movement in shaping the current configuration of Open Science is recognised, especially with regard to the basic principles of transparency, collaboration and sharing.

The Open Access movement promoted the Open Access literature.

The philosopher of law, researcher at the Berkman Klein Center for Internet & Society, Peter Suber, defined the Open Access literature as follow:

> Open Access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions.[25]

Between 2002 and 2003 three fundamental declarations on the topic of openness of the scientific results were announced: the Budapest Open Access Declaration (2002)[26]; the Bethesda Statement on Open Access Publishing (2003)[27]; and the Berlin Declaration on Open Access to Knowledge in Sciences and Humanities (2003)[28]. These three statements represented the foundations on which the Open Access movement was built, striving for an open, digital, and easily sharable scientific literature.

The Open Access movement stemmed from a specific need: to overcome the lock-in mechanism of scientific research results achieved through the so-called "pay-wall", imposed by scientific publishers. In other words, with the "pay-wall", the access to scientific resources is only possible by the payment of prices required by scientific publishers.

The results of research projects, often publicly funded, which took the form of scientific articles, are traditionally published by a few private scientific publishers[29]. This oligopoly charges high prices both

---

However, it is intended to bring out the original dimension of the Open Science movement connected primarily with the intention of promoting free access to scientific publications.

[25] Peter Suber, *Open Access*. (Cambridge, Massachusetts: The MIT Press, 2012), p. 4.

[26] https://www.budapestopenaccessinitiative.org/read.

[27] https://osc.universityofcalifornia.edu/2003/04/bethesda-statement-on-open-access-publishing/.

[28] https://openaccess.mpg.de/Berlin-Declaration. At this point, a distinction has been made between "gold" and "green" Open Access. The former involves the publication of a scientific paper in an academic journal that is directly open access, i.e., directly accessible from everyone, without paying; the latter, i.e., the "green" open access option, involves the author depositing (self-archiving) the paper published by scientific publishers, not in the format provided by the publisher but in the latest version prior to publication.

[29] The small number of scientific publishers has been described as "an oligarchic power", in Enrico Francesconi, Ginevra Peruginelli, "An open access policy for legal informatics dissemination and

to researchers (i.e., the authors of those papers) for access to individual articles, and to university libraries for subscriptions to scientific journals or series containing them[30]. This mechanism imposes a double payment of public money: in the first instance, public money is used to fund research projects; in the second instance, the fee is allocated to pay for subscriptions to scientific journals, so that the results of those research projects could be shared[31].

This mechanism is primarily rooted in two of the Modern Science features described in the previous section: (iv) the model of publishing research results in scientific journals, starting with "The Philosophical Transactions" of 1665, which was meant to report on scientific novelties and relevant debates; and (v) the model of copyright, as developed from the Statute of Anne, in 1710, which had the purpose of guaranteeing the author of works.

Although for centuries this mechanism had been the basis of the concept of openness of science, helping the scientific community as a whole and academic author individually, in the early 2000s it became the hallmark of the degeneration of the system[32]. The historian of science Jean-Claud Guédon defined this degeneration in terms of "serial pricing crisis", pointing to the steep increase in the price of subscriptions to scientific journals that must be covered by

---

sharing." *An Open Access Policy for Legal Informatics Dissemination and Sharing* (Berlin-Heidelberg: Springer, 2011): p. 170, doi: 10.1007%252F978-3-642-35731-2.

[30] The costs incurred by universities for access to scientific publications are in the range of millions of euros per year. See: MARK C. WILSON, "Universities spend millions of accessing results of publicly funded research." *The Conversation* (2017), https://theconversation.com/universities-spend-millions-on-accessing-results-of-publicly-funded-research-88392; .

[31] Here, the mechanism is simplified. Sometimes, in fact, the second payment, related to the sharing of research content, is not paid by the user, but still by the author (or her institution) through the so-called "article processing charges" (APCs). For the year 2022, Nature Neuroscience declared the ACP amounts to EUR 9500 per article, see: "Nature Neuroscience offers open access publishing." *Nature Neuroscience* 25, 1 (2022), doi: 10.1038/s41593-021-00995-2.

[32] On how the weight of choices developed in the past has resulted in a malfunctioning system today, see: MARIA CHIARA PIEVATOLO, "Open Access/Accesso Aperto." *Le parole dell'innovazione*, (2012) https://archiviomarini.sp.unipi.it/437/.

universities, to the benefit of commercial scientific publishers[33]. This degeneration was also exposed by the jurist Lawrence Lessig, in 2011, at a conference held at CERN in Geneva, with a speech evocatively entitled "the architecture of access to scientific knowledge: just how badly we have messed this up"[34].

The need to avoid the paradox of double payment and the emergence of the Internet and ICTs generated this first occurrence of the Open Science concept, namely the Open Access. Two aspects must be considered.

First, at the beginning of the 2000s the concept of openness was still strongly connected to printing and publishing. The expression Open Access means, in fact, free and immediate access, without cost, to scientific publications, i.e., scientific articles, written by scholars and published by scientific publishers, with the aim of disseminating and sharing the results of their research as much as possible. It still lacked any reference to the openness and sharing of research data or other aspects that later will characterise the Open Science paradigm.

Second, consider the nature of the Open Access initiatives. If today the Open Science appears to be an emerging paradigm supported by institutions, its first occurrence, i.e., the Open Access, had a very different nature: it stemmed from a bottom-up approach[35]. Those who signed the Berlin Declaration, or the Budapest Declaration, were professors, directors of research centres, representatives of public

---

[33] JEAN-CLAUDE GUÉDON, "In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing." In *Creating the Digital Future : Association of Research Libraries 138th Annual Meeting* (Toronto, Canada , 2001), http://eprints.rclis.org/6375/1/ARL_Proceedings_138_In_Oldenburg%27s_Long_Shadow%2C_by_Guedon.htm.

[34] LAWRENCE LESSIG, "The architecture of access to scientific knowledge: just how badly we have messed this up", *CERN Colloquium and Library Science Talk*, 18 April 2011, video, http://cdsweb.cern.ch/record/1345337.

[35] On this aspect, in particular on the legal domain, see, also: PAOLO GUARDA, "Open Access to Legal Scholarship and Open Archives: Towards a Better Future?." *From Information to Knowledge*. IOS Press (2011): 143-151.

universities or publishers: the claims of the Open Access movement were brought forward by the scientific community.

During the emergence of Web 2.0 and the subsequent explosion of active participation and community-based initiatives, which have been successful in different ways[36], Open Access evolved, becoming part of a real paradigm shift in contemporary science. Bearing in mind the origin of the phenomenon is essential to analyse its development and especially its current shape.

3.1.3 Evolution: Shaping the Open Science

The historian Paolo Rossi, in his study of the emergence of Modern Science in Europe, wrote:

> We are so familiar with that individual activity, which takes place in silence or isolation, of reading books, that it is difficult for us to realise that the familiar object we have in our hands could have appeared as a shocking novelty, something that not only spread ideas and knowledge in a previously inconceivable way, but also replaced the previously predominantly collective reading, probably conducted aloud, of texts without punctuation […].[37]

The interesting aspect of this excerpt is that the Author assumes that the reader is holding in her hands the "familiar object", the book. Only a few years after the publication of these words, however, it is unlikely to imagine a reader – especially if it is scientific and academic literature – on paper. It is much more likely that the reader is accessing papers and academic literature through a digital medium. Today, we are experiencing a new technical breakthrough: the digital revolution, which has characterised every aspect of our lives. The impact of the use

---

[36] On the ability of communities to self-regulation, see: MASSIMO DURANTE, *Il futuro del Web: etica, diritto, decentramento. Dalla sussidiarietà digitale all'economia dell'informazione in rete*. (Torino: Giappichelli Editore, 2007), 197-203. Specifically, three cases characterising the period of emergence of Web 2.0 were analysed: Wikipedia, Google, and Slashdot. Similarly, YOCHAI BENKLER, *The wealth of networks: how social production transforms markets and freedom*. (New Haven: Yale University Press, 2008).

[37] PAOLO ROSSI, *La nascita della scienza moderna*, *op. cit.*, p. 55. [translated from the Italian original text].

of ICTs in so many areas of our lives is so significant on almost every human being that it has been defined as an "everyday revolution"[38]. This major transformation has also concerned scientific research, as illustrated in Section 2.1.1, and it is from the first decade of the 2000s that it begins to be discussed in terms of Open Science.

In order to explore the evolution of the Open Science paradigm, the major perspectives expressed so far by scholars and institutions are analysed below: (i) the idea of a networked science; (ii) the schools of thought of Open Science; (iii) the so-called integrated definition of Open Science; (iv) the OECD stance; (v) the outcome of the FOSTER project; and finally (vi) the point of view of the European Union.

(i) The Networked Science. One of the first scholars to deal with the matter of Open Science organically has been the quantum theory expert Michael Nielsen. Nielsen proposes a formulation of a new way of doing science called "networked science"[39]. The starting assumption is that, in just two decades, between the 1990s and 2000s, access to knowledge expanded by thousands of times compared to the previous period. The idea of an open, interconnected science, made up of blogs[40], free access to resources and societal engagement[41] clearly emerges from his writing.

---

[38] MASSIMO DURANTE, *Computational Power: The Impact of ICT on Law, Society and Knowledge.* (New York: Routledge, 2021), p. 2.

[39] MICHAEL NIELSEN, *Reinventing discovery. The new era of networked science.* (Princeton: Princeton University Press, 2012), p. 89.

[40] MICHAEL NIELSEN, *Reinventing discovery, op. cit.*, pp. 165-169. The Author proposes an in-depth examination of the concept of the "science blogging", starting with the 2008 case concerning the chiropractic profession. A Guardian journalist, Simon Singh, published an article criticising the British Chiropratic Association (BCA), claiming that there was no scientific evidence to support the so-called medical treatments carried out by them. The association sued the journalist and published a document setting out the scientific evidence for chiropractic treatments: "What happened next was unexpected. Almost immediately, the evidence released by the BCA was investigated and torn apart by an *ad hoc* group of science bloggers, acting on their own initiative.", p. 166. The intervention of the scientist-bloggers was crucial in settling the matter.

[41] MICHAEL NIELSEN, *Reinventing discovery, op. cit.,* pp. 148-155. The Author analyses the role of citizens in science, represented by the phenomenon of "citizens science", stressing how technology changed the role of social engagement in science: "Citizen science is not an invention of the internet era. Many of the earliest scientists were amateurs, often pursuing science as a hobby alongside some more lucrative

The weakness of this reasoning, however, is the absence of a comprehensive definition of Open Science. Although a clear idea based on many features (openness, sharing, use of technology, etc.) is evoked, and although the positive aspects and limits of this new emerging science are clarified, the defining aspect is missing. Nielsen refers to the "Open Science imperative", arguing that the openness of science must be embraced by the scientific community as a whole, in order to be effective. With the aim of pursuing effectiveness, Nielsen identifies the need to encourage Open Science indicating a series of "practical steps"[42]. Besides the insightful considerations about the nature of the Open Science problems, also from a cultural point of view[43], the lack of a full definition of the phenomenon is relevant for setting the limits within which any institutional action should be envisaged. Supporting Open Science as an imperative, a "must be", and thus demanding the intervention of institutions to foster it, necessarily requires the identification of a definition, in order to determine the boundaries of the phenomenon, from which to operate effectively.

(ii) The Open Science Schools of Thought. By the middle of the second decade of the 2000s, there was already a well-established debate about the impact of the digital revolution on knowledge and, specifically, on scientific knowledge. The concept of Open Science was taking shape in its complexity, but there was still a lack of clarity. Crucial in this context was the publication "Opening science", by Sönke Bartling and Sascha Friesike, published by Springer and released in Open Access[44]. By mapping almost all aspects of the Open Science

profession […]. But even after science was professionalized, amateurs continued to dominate some parts of science. […] Although citizen science is not new, online tools are enabling far more people to participate […] and also expanding the range of scientific work those people can do.", pp. 148-149.

[42] MICHAEL NIELSEN, *Reinventing discovery, op. cit.,* pp. 203-206.

[43] "Today, creating an open scientific culture seems to require an impossible change in how scientists work. But by taking small steps we can gradually cause a major cultural change." in MICHAEL NIELSEN, *Reinventing discovery, op. cit.,* p. 206.

[44] SÖNKE BARTLING, SASCHA FRIESIKE, *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publis*hing. (Cham: Springer Nature, 2014),

phenomenon[45], this work brings order to the debate with great insight. In particular, with the aim of providing "[…] a comprehensible overview of the predominant thought patterns in the current Open Science discourse and point towards new directions in research regarding Open Science"[46], a classification of Open Science was proposed, organised according to five different "schools of thought":

> The infrastructure school (which is concerned with the technological architecture), the public school (which is concerned with the accessibility of knowledge creation), the measurement school (which is concerned with alternative impact measurement), the democratic school (which is concerned with access to knowledge) and the pragmatic school (which is concerned with collaborative research). [47]

According to this classification, the term "Open Science" can be represented through five different schools of thought, as summarised in Figure 3.1, and further analysed below.

- The infrastructure school is focused on the technological side, i.e., the use of platforms, tools and services that are openly available to researchers. This aspect of Open Science is closely related to the Open Source movement[48].

- The public school is perhaps the closest to the original concept of Open Science, analysed in Section 3.1.1, as opposed to secret science, held by a few. The aim is to foster the maximum inclusion of individuals in science, through direct participation (i.e., citizen science[49]), or indirect participation (e.g., crowdfunding of science or blogging).

---

doi:10.1007/978-3-319-00026-8.

[45] Nearly all of them, because the analysis of the purely legal aspects and, more specifically, the issues of personal data protection is not extensively covered.

[46] SÖNKE BARTLING, SASCHA FRIESIKE, *Opening science, op. cit.*, p. 18.

[47] SÖNKE BARTLING, SASCHA FRIESIKE, *Opening science, op. cit.,* p. 17.

[48] On the concept of Open Source, see: CHRIS DIBONA, SAM OCKMAN, *Open sources: Voices from the open source revolution*. (Sebastopol, California: O'Reilly Media Inc. 1999); and also MARIELLA BERRA, ANGELO RAFFAELE MEO, *Libertà di software, hardware e conoscenza. Informatica solidale 2,* (Torino: Bollati Boringhieri, 2006).

[49] On the concept of citizen science and the societal engagement in science, see, chiefly, the first relevant study of this topic: ALAN IRWIN, *Citizen Science: A Study of People, Expertise and Sustainable*

- The measurement school understands Open Science as a new method of evaluating science. This interpretation is related to the positions of the original Open Access movements: in fact, as well as criticising the concept of the so-called "pay-wall" mentioned above[50], it rejects the traditional system of evaluating scientific results as established since the 1930s[51], synthesised by the formula "publish or perish"[52]. The aim is to identify and build up alternative models of research evaluation.

- The democratic school interprets the concept of openness in terms of accessibility to contents. This openness, however, does not only concern scientific publications, i.e., the results of scientific research. It also covers, more generally, the data on which research projects have been developed (this dimension of Open Science is close to the field of Open Data[53]); but also, any other types of dissemination and communication of scientific results should be taken into account (e.g., teaching activities[54]).

Development. (Abingdon: Routledge, 1995). Then, also consider: CHRISTOPHER KULLENBERG, DICK KASPEROWSKI, "What is citizen science?–A scientometric meta-analysis." *PloS one* 11.1 (2016): e0147152, doi: 10.1371/journal.pone.0147152; RICK BONNEY, *et al.*, "Next steps for citizen science." *Science* 343.6178 (2014): 1436–1437, doi: 10.1126/science.1251554; PANOS PETRIDIS, *et. al.*, "The role of science in sustainability transitions: citizen science, transformative research, and experiences from Samothraki Island Greece." *ISJ* 12.1, (2017): 115–134, doi: 10.24043/isj.8; but see, also, LEA DEN BROEDER, *et al.,* "Citizen Science for public health." *Health promotion international* 33.3 (2018): 505-514, doi: 10.1093/heapro/daw086. The phenomenon of "citizen science" has already been mentioned in Section 2.2.6, related to the establishment of the new Horizon Europe funding programme.

[50] See: Section 3.1.2.

[51] FRANCESCA DI DONATO, *La scienza e la rete, op. cit.*, pp. 23-24. This traditional system of evaluating scientific results, in any case, is grounded in the model developed during Modern Science, formalised around 1930 and degenerated at the end of 1990th, as claimed by the Open Access movement.

[52] This formula indicates a trend that links the growth of a career in scientific research solely to the constant publication of articles. The negative aspect of the system, which this formula is intended to emphasise, is that the quantity of publications prevails over their quality. See: TENNANT, JONATHAN P., *et al.*, "Ten hot topics around scholarly publishing." *Publications* 7.2 (2019): p. 10, doi: 10.3390/publications7020034.

[53] On the relation between Open Science and Open Data, see: Section 2.3.1; and also, Chapter 4.

[54] On this aspect, see: the observable O5 "Benefits", in Section 3.2.

- The pragmatic school focuses on methodologies: openness is understood in terms of sharing the processes of creating scientific knowledge, based on maximum collaboration between peers[55].

This taxonomy represents a valuable contribution to the shaping of the Open Science paradigm, able to underline the multifold nature of the phenomenon. What does not seem entirely convincing, however, is the idea that Open Science can be represented as a mere aggregation of "schools of thought".

However, consider that this classification depicts the phenomenon at the beginning of its current configuration when Open Science was an individual's choice or a personal vision of science. Today, this scenario no longer seems to match reality, due to the commitment of institutions, especially the European ones, in the context of the Open Science[56].

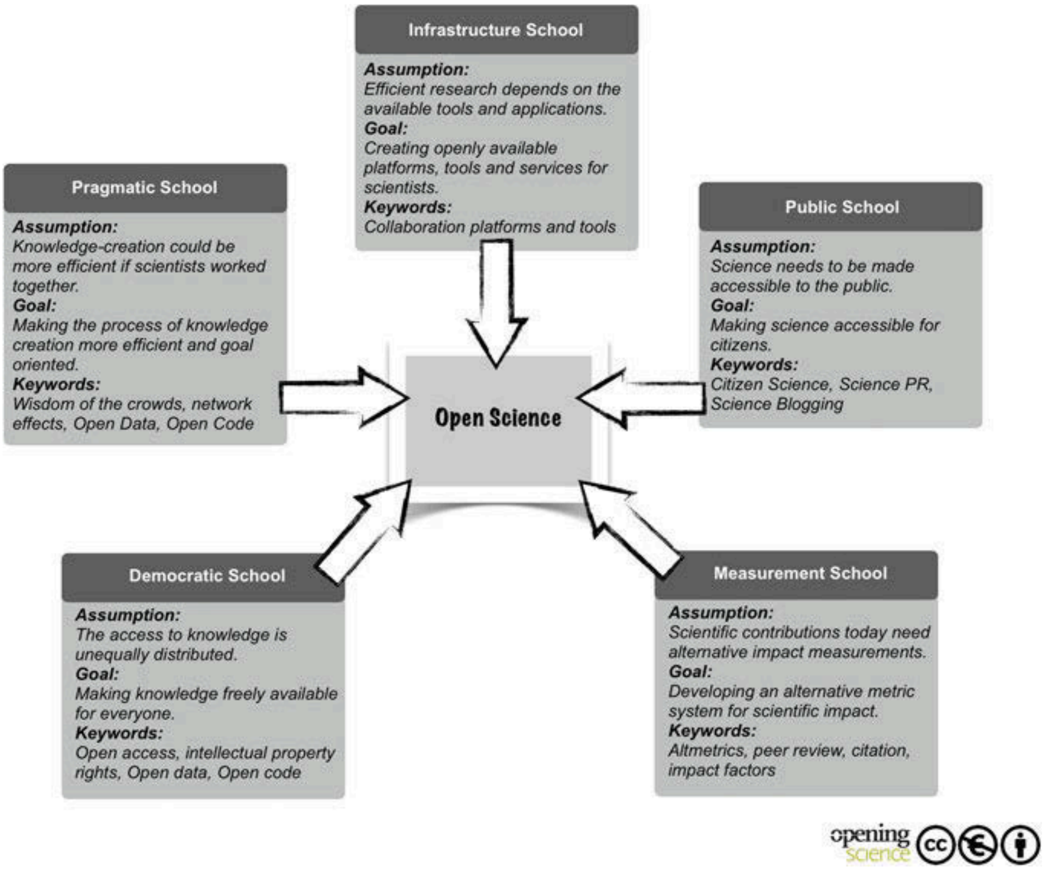
Certainly, the taxonomy of Bartling and Friesike still represent an important analysis due to the fact that it provides a coherent vision of a phenomenon that in 2014 had already been discretely investigated, but always with partial analyses, not comprehensive. Yet, according to this classification, the Open Science represents many different dimensions, not communicating with each other: for some, openness was only related to the aspect of scholarly communication; for others, only to the technological infrastructures; for others it only meant a broader participation in scientific projects, in terms of actors involved, etc. This classification brings all these different (and persisting) dimensions of the phenomenon under the same umbrella of Open Science. What still seems to be missing, however, is a link between

[55] On the concept of peer collaboration in science, see: JULIE S. HUI, ELIZABETH GERBER, "Crowdfunding science: sharing research with an extended audience." *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, New York* (2015): 31- 43, doi: 10.1145/2675133.2675188.

[56] This commitment has been represented as the "institutional premise" of this dissertation, exposed in Section 2.1.2 and further developed in Section 2.2, mapping the evolution of the EU Open Science policies.

these different dimensions, represented as separate and non-communicating silos.

Figure 3.1: "Open Science: One Term, Five Schools of Thought" in Bartling and Friesike (2014)[57]



(iii) An Integrated Definition. The need for greater clarity in the identification of the concept of Open Science persists in following years. A 2018 study[58], in fact, proposes a further mapping of the existing

[57] SÖNKE BARTLING, SASCHA FRIESIKE, *Opening science, op. cit.*, p. 19.
[58] RUBÉN VICENTE-SÁEZ, CLARA MARTÍNEZ-FUENTES, "Open Science now: A systematic literature review for an integrated definition." *Journal of business research* 88 (2018): 428-436, doi: 10.1016/j.jbusres.2017.12.043.

definitions and interpretations of the concept of Open Science, starting from the assumption that "[…] there is a lack of awareness about what Open Science is, mainly due to the fact that there is no formal definition of Open Science"[59]. This analysis concludes with the proposal of an integrated definition of Open Science:

> Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks.[60]

This definition fulfills the aim of providing a general idea of the phenomenon, encompassing the central features of transparency, access to knowledge and network collaboration. However, it is very broad, not providing much information on the actual nature of the phenomenon. This integrated definition has the merit of moving further away from the Modern Science definition of Open Science, and closer current needs of the scientific research. Openness ceases to be in opposition only to the old occult and secret science. As the palaeontologist and Open Science advocate Jon Tennant said on 11 January 2018: "[T]he opposite of 'open science' isn't 'closed science' – it's bad science"[61].

(iv) OECD Stance. Alongside the interventions of scholars, some fundamental steps forward in identifying the phenomenon of Open Science have been taken by different institutions.

The first relevant example is provided by the Organization for Economic Co-operation and Development (OECD), which released the report "Making Open Science a reality", in 2015[62].

Regarding the definition of Open Science, it says:

> There is no formal definition of open science. In this report, the term refers to efforts by researchers, governments, research funding

[59] RUBÉN VICENTE-SÁEZ, CLARA MARTÍNEZ-FUENTES, "Open Science now", *op. cit.*, p. 429.
[60] RUBÉN VICENTE-SÁEZ, CLARA MARTÍNEZ-FUENTES, "Open Science now", *op. cit.*, p. 435.
[61] https://twitter.com/protohedgehog/status/951413580167110656.
[62] OECD, "Making Open Science a Reality", *OECD Science, Technology and Industry Policy Papers*, n. 25 (Paris: OECD Publishing, 2015), http://dx.doi.org/10.1787/5jrs2f963zs1-en.

> agencies or the scientific community itself to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction as a means for accelerating research; these efforts are in the interest of enhancing transparency and collaboration, and fostering innovation.

The mapping of Open Science conducted by the OECD report is the first relevant analysis of the phenomenon resulting from the collaboration between institutions and Academia[63]. It represents the output of the activities realised by the OECD's Working Party on Innovation and Technology Policy (TIP) of the Committee for Scientific and Technology Policy (CSTP). The study involved some members of the OECD Secretariat, the TIP and some academic experts in the field[64].

The overview provided by the OECD report is rather exhaustive, taking into account both access to publications and access to data, and adopts a comprehensive view of the topic: "Open science commonly refers to efforts to make the output of publicly funded research more widely accessible in digital format to the scientific community, the business sector, or society more generally. Open science is the encounter between the age-old tradition of openness in science and the tools of information and communications technologies (ICTs) that have

---

[63] Rather than being strictly institutional, the nature of the report is hybrid. Although it was issued by the OECD, it was written by a group of experts from many different academic backgrounds. The OECD has played a very active role in this direction. As illustrated by Nielsen: "[…] in 2007 the Organization for Economic Co-operation and Development (OECD) recommended that member countries make publicly funded research data openly accessible. Such recommendations take time to filter down, but over time they can have an impact.", in MICHAEL NIELSEN, *Reinventing discovery, op. cit.,* p. 191.

[64] First, Lucie Guibault and Thomas Margoni (University of Amsterdam) who worked on the background paper to this report, containing detailed analysis of the legal aspects of Open Science and Open Data. Second, the report has been enriched by suggestions and comments on the state of policy in the different countries involved, from several academics (e.g., Prof. Juan Carlos De Martin, Politecnico di Torino) and institutional actors (e.g., a representative of the European Commission, or the Turkish Minister of Science).

reshaped the scientific enterprise and require a critical look from policy makers seeking to promote long-term research as well as innovation”[65].

Although the report was published only a few years ago, one aspect has changed considerably in the meantime: the role of the private sector in science. In the OECD report, the role of the private actors in scientific research was limited to certain areas of action. Today the boundaries between public and private sector in the scientific research process are, in certain circumstances, blurred, adding further complexity, as discussed later[66].

(v) Outcome of the FOSTER Project. The most comprehensive taxonomy currently available is the one resulting from the project “Fostering the practical implementation of Open Science in Horizon 2020 and beyond”, known as “FOSTER”, a project funded by the European Union, which intends “[…] to contribute to a real and lasting shift in the behaviour of European researchers to ensure that Open Science (OS) becomes the norm”[67], carried out by eleven partners from different Member States.

The FOSTER taxonomy has the advantage of bringing together the various aspects of the Open Science, providing a clear and coherent view of the phenomenon.

FOSTER defines the Open Science as follow:

> Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.[68]

---

[65] OECD, *“Making Open Science a Reality”, op. cit.*, p. 9.

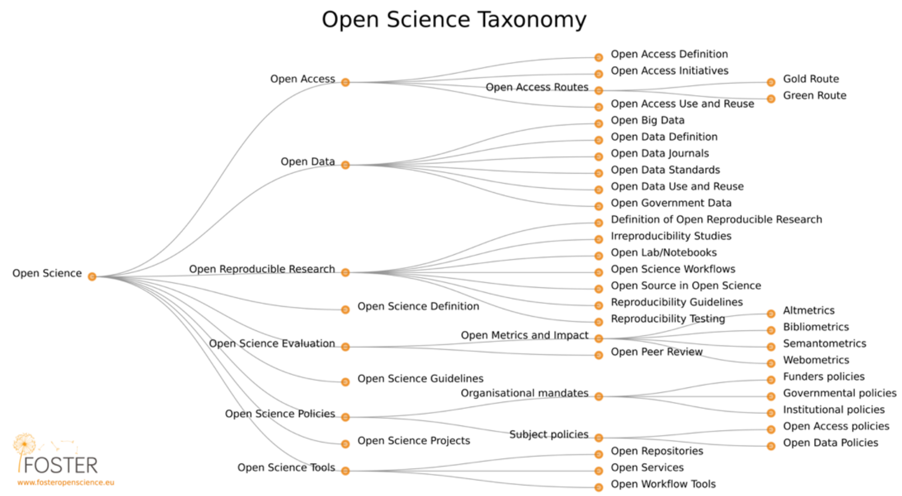
[66] See: Section 3.2. With regard to the analysis of the Observable “Actors”, this transformation is also described in its qualitative and quantitative dimensions in Section 6.3.

[67] https://www.fosteropenscience.eu/.

[68] https://book.fosteropenscience.eu/en/02OpenScienceBasics/01OpenConceptsAndPrinciples.html. The Open Science definition provided by FOSTER, see: https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition.

The definition provided by FOSTER combines the various dimensions of Open Science, making the separate silos, mentioned above, communicate, as represented in the following Figure 3.2.

Figure 3.2: Open Science Taxonomy by FOSTER (2018)



The FOSTER's taxonomy is useful in providing a mapping of the most debated expressions of the Open Science paradigm. For instance, it is very specific in relation to open research data or in relation to publications. However, it does not point out clearly the constitutive elements of the phenomenon, which can be found in all occurrences of the Open Science paradigm. In other words, this definition lacks an interface that makes the analysis of the phenomenon fruitful[69].

In addition, it does not include, for example, any mention to the Citizen Science or the Open Education, which are relevant strands of the Open Science field. At least, it is believed that these dimensions should be taken into consideration if the intention is to propose an all-encompassing taxonomy of the phenomenon.

[69] This interface is described in terms of “Level of Abstraction”, in the following Section 3.2.

(vi) EU Perspective. The expression Open Science, mentioned a few times in the official acts of the European Union from 2010 to 2015, becomes recurring from 2016 onwards. Under the boost of the European Commission, the European Union has developed since that period a set of policies, recommendations, roadmaps, and initiatives with the aim to promote Open Science, which are still ongoing and have been analysed and mapped in the previous Section 2.2.

In 2019 the official Open Science.eu[70] webpage was created, and the following definition was provided:

> Open Science is a system change allowing for better science through open and collaborative ways of producing and sharing knowledge and data, as early as possible in the research process, and for communicating and sharing results. This new approach affects research institutions and science practices by bringing about new ways of funding, evaluating and rewarding researchers. Open Science increases the quality and impact of science by fostering reproducibility and interdisciplinarity. It makes science more efficient through better sharing of resources, more reliable through better verification and more responsive to society's needs.[71]

Although this definition is not particularly analytical, it can be considered a synthesis of the previous developments of the phenomenon. It focuses on the relevant aspects for a policy-making entity. In addition, recently, the EU institutions provided a legal definition of the Open Science, in the Article 2 of the EU Regulation 2021/695, establishing the Horizon Europe programme[72]:

---

[70] The European Commission's webpage dedicated to Open Science is the following: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science_en.

[71] https://ec.europa.eu/info/sites/default/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_factsheet-open-science_2019.pdf.

[72] Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021, *establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013*, OJ L 170, 12.5.2021, p. 1–68, ELI: http://data.europa.eu/eli/reg/2021/695/oj.

> 'open science' means an approach to the scientific process based on open cooperative work, tools and diffusing knowledge, and includes the elements listed in Article 14.[73]

It should be noted that the Article 14 mentioned is entirely devoted to Open Science and has already been discussed above[74]. However, what is important here is to stress the fact that for the first time Open Science has been envisaged and defined in a hard law legislative text: the emergence of the new paradigm of science is formally recognised by the European institutions.

After the illustration of the most relevant perspectives of the Open Science paradigm, representing the evolution of the concept, it is now time to present our interpretation of the Open Science, in terms of Open Scientific Research Process.

3.2 The Definition of Open Science as a Process

In light of the current state-of-the-art on the Open Science notion just analysed, here the intention is to propose a definition of Open Science able to identify the constitutive elements of the phenomenon. The purpose is to pinpoint the pivotal factors of this new paradigm of science, represented by the encounter between the pillars of openness, sharing and collaboration of science, with the capabilities of digital ICTs: in doing so, institutions acting on one or more of these kernels can trigger an effective improvement of the governance of the entire scientific research in Europe. The intention is to suggest an interpretation of Open Science as an "Open Scientific Research Process".

So far, the analysis showed that the definitions of the Open Science phenomenon can be multiple, as multiple are the facets of this umbrella concept. In order to sidestep the hurdles encountered by previous definitions and taxonomies proposed by scholars and institutions, the

[73] The Article 2(5), Regulation EU 2021/695.
[74] See: Section 2.2.6.

definition of Open Science as a process adopts the method of levels of abstraction (LoA), developed by Floridi[75]. The preliminary consideration is that every aspect of reality can be investigated from different perspectives or points of view: the same applies for Open Science.

The LoA method enables to identify a specific point of view, from which to develop the investigation of the concept under examination. As stated by Pagallo, "[…] the chosen level of abstraction can therefore be understood as the interface that makes the analysis of the system possible"[76]: in other words, the lens through which we intend to observe a given phenomenon.

Once the level of abstraction is determined, the investigation moves on to assess the features of each LoA which, according to Floridi's terminology, are referred to as "observables". Finally, the investigation, proceeding from the general to the particular, can be further specified, through the indication of the multiple features, typical of each observable, called "variables".

In our case, the system under investigation is the Open Science. The level of abstraction chosen is the process level: Open Science is understood as an Open Scientific Research Process.

This process can therefore be represented through observables. These observables, in our model, must be considered as the constitutive elements of the Open Scientific Research Process. The identification of these observables plays a crucial role in the development of the governance of scientific research: by intervening on each one of them, it is possible to modify the entire system.

---

[75] LUCIANO FLORIDI, "The method of levels of abstraction." *Minds and machines* 18.3 (2008): 303-329, doi: 10.1007/s11023-008-9113-7; LUCIANO FLORIDI, "Levels of abstraction and the Turing test." *Kybernetes*, 39.3 (2010): 423-440, doi: 10.1108/03684921011036150. On the multiple definitions of the concepts that constitutes the LoA model (e.g., observables, variables, etc.), see: LUCIANO FLORIDI, *The philosophy of Information*. (Oxford: OUP 2011), pp. 46-79.

[76] UGO PAGALLO, *Il diritto nell'età dell'informazione: il riposizionamento tecnologico degli ordinamenti giuridici tra complessità sociale, lotta per il potere e tutela dei diritti*. (Torino: Giappichelli Editore, 2014), p. 18. [Translation from the Italian original text].

For the LoA "Open Scientific Research Process" five observables have been identified, which are detailed below: (O_1) Resources; (O_2) Actors; (O_3) Methods; (O_4) Tools; (O_5) Benefits.

(O_1) Resources. Every process necessarily requires an input to trigger it. In the case of the Open Scientific Research Process, the observable (O_1) "resources" is the input of the process and can be represented by two variables: ($V_{1.1}$) research data; ($V_{1.2}$) economic resources.

Data ($V_{1.1}$) are becoming increasingly important in every field, first and foremost in scientific research[77]. The use of data in the field of science implies a series of choices (e.g., the methodologies through which they are collected; the purposes one wants to achieve through their processing; the techniques by which data are stored and possibly reused, etc.). These choices have a great impact on the entire research project: what Leonelli calls "data journey"[78] becomes fundamental[79].

Similarly, economic resources ($V_{1.2}$) underpin the research process. This process is based on the allocation of funding, which may be public or private grants. Economic resources represent the input of the scientific process since they ensure the funding of researchers, equipment, and technology, and generally cover all the related expenses.

---

[77] This aspect has been discussed in more detail in Chapter 2, Section 2.1.1 "Methodological premise". Consider that the role of data is so important in scientific research that it has been referred to as "data-driven science". On this aspect, see: YASUHIKO IGARASHI, *et al.*, "Three levels of data-driven science." *Journal of Physics: Conference Series* IOP Publishing, 699.1 (2016): 1-13, doi: 10.1088/1742-6596/699/1/012001. In this paper, the Authors deemed "[…] that any problem of data analysis should be discussed at different three levels: computational theory, modeling, and representation/algorithm". Then, against the supposed neutrality of numbers, see: FULVIO MAZZOCCHI, "Could Big Data be the end of theory in science? A few remarks on the epistemology of data-driven science." *EMBO reports,* 16.10 (2015): 1250-1255, doi: 10.15252/embr.201541001.
[78] SABINA LEONELLI, "What difference does quantity make? On the epistemology of Big Data in biology." *Big data & society* 1.1 (2014): 1-11, doi: 10.1177/2053951714534395.
[79] On the role of research data and the related issues, see Chapter 4.

By acting respectively on the management of data ($V_{1.1}$), or on the allocation of economic resources ($V_{1.2}$), the whole process is influenced[80].

(O_2) Actors. Another constitutive element of the Open Scientific Research Process is represented by the actors. As described in Section 2.4.3.3, the actors of scientific research are experiencing a considerable transformation. In order to be more precise, here, this change can be described by considering two trends: on the one hand, there is a quantitative transformation; on the other, there is a qualitative transformation.

The quantitative transformation is due to the fact that the Open Science scenario involves the interaction of a great number of actors, which is considerably larger than in the traditional mechanism of science. For instance, consider the role of society: the new Horizon Europe funding programme repeatedly calls for the engagement of society in science[81]. This increased role of civil participation is a new feature of the emerging science paradigm.

The qualitative transformation relates to a change in the quality of actors, understood as a variation of the properties characterising some of the traditional actors. An example is given by that strand of research which is investigating the transformation of traditional scientific publishers into data analysis businesses[82].

[80] A strand of research has for years been investigating the impact of grant allocation in scientific research and its consequences, from multiple perspectives. Among others, see: NEIL VINER, PHILIP POWELL, ROD GREEN, "Institutionalized biases in the award of research grants: a preliminary analysis revisiting the principle of accumulative advantage." Research Policy 33.3 (2004): 443-454, doi: 10.1016/j.respol.2003.09.005; HANNA HOTTENROTT, CORNELIA LAWSON, "Research grants, sources of ideas and the effects on academic research." *Economics of Innovation and New Technology* 23.2 (2014): 109-133, doi: 10.1080/10438599.2013.814425.
[81] See: Section 2.2.6.
[82] A report exploring the transformation of the major publishers in the field of scientific research has recently been published, see: Committee on Scientific Library Services and Information Systems of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), "Data tracking in research: aggregation and use or sale of usage data by academic publishers", June 2021, available at:

(O3) Methods. The observable O3 "Methods" concerns the whole set of methodologies adopted in carrying out the Open Scientific Research Process. It has already been pointed out several times that a debate concerning the validation of results is currently ongoing. This debate can be articulated under two headings: evaluation (V3.1); and verifiability (V3.2).

First, V3.1 represents the validation in terms of evaluation of the results of science: evaluation of the performance of researchers, citation indices, productivity, impact, etc[83].

Second, V3.2 refers to validation in terms of the verifiability of the results obtained: here the emphasis is on the reproducibility of the experiments underlying the research projects[84]. This second dimension

---

https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_en.pdf. On this aspect, see: Section 5.1. The consequences are being investigated by many scholars: "At a time when humanity has at its disposal the most powerful technology for dialogue (the Internet), the dissemination of scientific publications is artificially restricted in order to benefit the commercial interests of a few oligopolists", in ROBERTO CASO, *La società della mercificazione, op. cit.*, p. 314. [Translation from the Italian original text].

[83] The debate is ongoing. For an overview about the Academia's metric, see, among others: ASHOK AGARWAL, *et al.*, "Bibliometrics: tracking research impact by selecting the appropriate metrics." *Asian journal of andrology* 18.2 (2016): 296-309, doi: 10.4103/1008-682X.171582; DIANA HICKS, *et al.*, "Bibliometrics: the Leiden Manifesto for research metrics." *Nature News* 520.7548 (2015): 429-431, doi: 10.1038/520429a. On the field of the evaluation in the law domain, see: GINEVRA PERUGINELLI, SEBASTIANO FARO, "Research quality evaluation: The case of legal studies." *The evaluation of research in social sciences and humanities* (Cham: Springer, 2018): 103-129, doi: 10.1007/978-3-319-68554-0_5. This dimension of the Open Science is related to the so-called "Open peer-review" as "[…] an umbrella term for a number of overlapping ways that peer review models can be adapted in line with the aims of Open Science, including making reviewer and author identities open, publishing review reports and enabling greater participation in the peer review process." in TONY ROSS-HELLAUER, "What is open peer review? A systematic review." F1000Research 6 (2017), p. 1, doi: 10.12688/f1000research.11369.2.

[84] Here a clarification is needed (which is, then, further analysed in Section 4.1.1), on the distinction between reproducibility and replicability: "Here 'repeatability' means that, given a certain scientific result, this must be able to be obtained or re-obtainable (within a plausible range of error) by the same researcher, in the same laboratory and using the same set of tools. 'Reproducibility', on the other hand, means that, given a certain scientific result, this result must be able to be obtained or re-obtained (within a plausible range of error) in any laboratory, at any time, by any researcher, even using different tools and different empirical methods" in GIOVANNI BONIOLO, *Molti. Discorso sulle identità plurime*. (Torino: Bollati Boringhieri, 2021), p. 100. The Author develops this clarification, dealing with the issue of the interaction of different branches of knowledge, elaborating as a consequence a remark on the scientific type of representation: "Everything that is not hypothetical, logically coherent and empirically repeatable and reproducible is not a scientific representation, but something else (which, depending on the case, can be magic, myth, metaphysics, religion, fiction, nonsense and so on).", p. 100. [Translation

of the debate is particularly relevant: for instance, think about the research projects carried out on the basis of unsupervised Machine Learning (ML) algorithms, which makes it difficult to reproduce the machine's process exactly. There is a strand of research that encompasses all issues relating to the difficulties of reproduction – and therefore verifiability – of research in the expression "reproducibility crisis"[85]. The causes of this reproducibility crisis may be manifold: the absence of raw data[86], or the lack of sharing the source code of the algorithm used to perform the search[87], etc[88].

In any case, intervening at any level on methods implies changes embedded in the whole process of the open scientific research.

(O4) Tools. The role that tools play in this context was already previously discussed, investigating what has been called the methodological premise of this dissertation[89]. This premise concerned precisely the impact of new technological tools on the traditional scientific research. Applying Moore's interpretation of computer ethics to this study, two aspects of the impact of technology on scientific research have been explored: on the one hand, how the way of doing

from the Italian original text].

[85] Among others, see: MONYA BAKER, "Reproducibility crisis." *Nature* 533.26 (2016): 353-366, doi: 10.1038/nature17990; DANIELE FANELLI, "Opinion: Is science really facing a reproducibility crisis, and do we need it to?" *Proceedings of the National Academy of Sciences* 115.11 (2018): 2628-2631, doi: 10.1073/pnas.1708272114; NIELS G. MEDE, *et al*. "The 'replication crisis' in the public eye: Germans' awareness and perceptions of the (ir) reproducibility of scientific research." *Public Understanding of Science* 30.1 (2021): 91-102, doi: 10.1177/0963662520954370. This topic will be further examined in Section 4.1.1.

[86] TSUYOSHI MIYAKAWA, "No raw data, no science: another possible source of the reproducibility crisis." *Molecular Brain*, 13 (2020): 13-24, doi: 10.1186/s13041-020-0552-2.

[87] MATTHEW HUTSON, "Artificial intelligence faces reproducibility crisis." *Science* 359.6377 (2018): 725-726, doi: 10.1126/science.359.6377.725.

[88] Some reasons of the reproducibility crisis are not related to the technologies involved. For instance, quantum physics is currently facing a considerable problem concerning not only the verifiability, but the falsifiability of string theories, see: SOPHIE RITSON, KRISTIAN CAMILLERI, "Contested boundaries: The string theory debates and ideologies of science." *Perspectives on Science* 23.2 (2015): 192-227, doi:10.1162/POSC_a_00168; and BRIAN GREENE, *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory* (New York: W.W. Norton, 1999).

[89] See, Section 2.1.1.

science has changed; on the other hand, the way of wondering about science.

The introduction of new technologies and new tools in science concerns different aspects: the realisation of a research project; the archiving of results; the reuse of research data; the dissemination of research results, etc. The case study illustrated in Chapter 6 casts light on this constitutive element of the Open Scientific Research Process, i.e., the role of technologies and e-infrastructures in the scientific research scenario.

(O_5) Benefits. The observable O_5 "Benefits" represents the output and the final stage of the Open Scientific Research Process. The variables may be multiple. The output of the Open Scientific Research Process can in fact be represented by the publication of the results of the research project; by teaching; by the application of the research finding to the industry sector; by the dissemination of the contents with civil society; by the realisation of reports that can be exploited in public deliberation[90], etc.

As regards O_5 "Benefits", an interesting aspect is represented by the link between Open Science and Open Education. In general terms, the formula "Open Education" identifies the open and free access to educational resources of various kinds. For a while, taxonomies or classifications of the Open Science phenomenon did not include the educational aspects, somehow reducing the role of Open Education to the so-called MOOC, i.e., Massive Open Online Courses[91].

---

[90] On the role of public deliberation and the deliberative democracy, see: GIOVANNI BONIOLO, *The art of deliberating: democracy, deliberation and the life sciences between history and theory*. (Berlin, Heidelberg: Springer Science & Business Media, 2012): p. 4, doi: 10.1007/978-3-642-31954-9. In particular: "[…] deliberative democracy […] puts emphasis on the role of the reasons behind a given choice; namely, it stresses the need for and the importance of the offered justifications. In this case, what really matters is not just the final moment of the actual choice, but the relevant process that leads to the choosing, and such a process is always collective, since individuals with diverging positions should rationally dialogue with each other in order to achieve a common result".

[91] On the multifaceted concept of Open Education, see: FABIO NASCIMBENI, *Open Education*. (Milano: Franco Angeli Editore, 2020), pp. 17-28.

Open Education is much more than that and embraces many aspects of teaching, which also acquires a new meaning as part of the Open Scientific Research Process[92].

Once again, it is emphasised how, by affecting one observable (or one of its variables), a repercussion on the entire Open Scientific Research Process (LoA) is generated. Consider, as an illustration, the so-called "Plan s", an initiative launched in 2018 by the so-called "Coalition S"[93], a *consortium* of national research agencies from various European countries. This initiative has set the goal of affirming the practice of Open Access publications, also through an abrupt change (the "S" stands for "shock"): the aim is to require the publication in Open Access of the results of publicly funded research projects, as a mandatory condition[94].

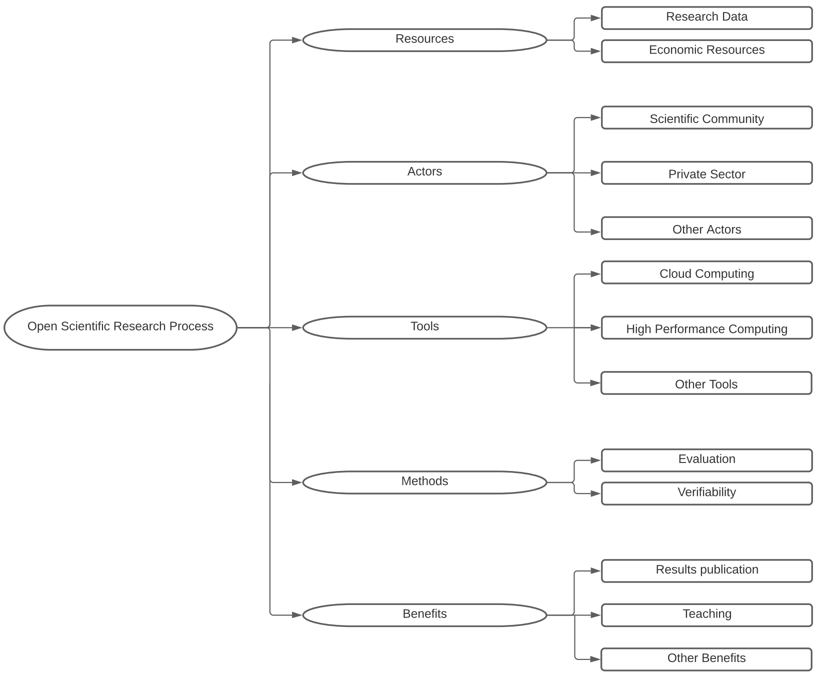
However, this initiative imposes a requirement that is hardly enforceable, given the very nature of the Coalition S (and despite the more or less explicit support of the European institutions). In spite of the controversial nature (and effectiveness) of this initiative[95], the debate has indeed been remarkable. By considering the debate generated by Plan S, a greater impact of a regulatory intervention on any of the several variables of this observable O_5 might be expected.

To summarise the interpretation of the Open Science as an "Open Scientific Research Process" (LoA), defined by its five constituent elements (observables), I propose the following definition:

[92] A new perspective on the Open Higher Education as part of the Open Science paradigm, see: DANIEL BURGOS, *Radical Solutions and Open Science: An Open Approach to Boost Higher Education.* (Singapore: Springer Nature, 2020), doi: 10.1007/978-981-15-4276-3.
[93] https://www.coalition-s.org/about/.
[94] https://www.coalition-s.org/.
[95] TOMASZ J.GUZIK, AMRITA AHLUWALIA, "Plan S: in Service or Disservice to Society? The controversial plan for scientific research publications to be published in compliant Open Access Journals or on compliant Open Access Platforms, is discussed.", *European Heart Journal*, 40.12 (2019): 949-952, doi: 10.1093/eurheartj/ehz065.

the Open Science describes the Open Scientific Research Process in which openness regards (i) the inputs, i.e., the underlying resources, such as data and economic funding; (ii) actors, including public at large; (iii) methods, in terms of evaluation and verifiability; (iv) tools; and (v) benefits deriving from this process, including the dissemination of the results, through publication, teaching activities or more.[96]

Figure 3.3: LoA Open Scientific Research Process



The purpose of this definition is to approach the issue dynamically, pointing out constitutive elements of the Open Science paradigm.

[96] It is intended to clarify that while "Open Science" is the *definiendum*, the *definiens* is the "Open Scientific Research Process", identified in its constituent elements, i.e., inputs, actors, methods, tools, and benefits. The reference to the notion of openness could be misleading: consider that in order to sidestep the risk of *circulus in definiendo*, the notion of openness will be specified from time to time in this dissertation, in the meaning it will adopt in relation to the various constitutive elements investigated. For instance, about the notion of openness related to the V1.1 "Research Data", see: Chapter 4. On the identifying aspects of the definition, see: GIOVANNI BONIOLO, PAOLO VIVALDI, *Strumenti per ragionare. Logica e teoria dell'argomentazione*. (Milano: Mondadori, 2011), pp. 89-97.

Through the modulation of these constitutive elements, summarised in Figure 3.3, the European institutions and all the other actors involved can shape the entire context: by co-designing the necessary regulations, incentives and disincentives, they may provide a flexible but robust governance of the scientific research.

In light of this proposed interpretation of Open Science, focused on its constitutive elements (i.e., resources; actors; methods; tools; benefits), next section explores the framework of the human and fundamental rights in which the Open Science paradigm is grounded.

3.3 Human and Fundamental Rights Framework of Open Science

Adopting a legal approach to analyse the Open Science phenomenon requires an examination of the sources of law that characterise its foundation. In the previous Chapter, a mapping of European policies on Open Science was proposed[97]: now, it is time to consider the framework of fundamental rights to which these initiatives are rooted. As regards the rights connected to the field of science, it is first presented the international dimension and then the European one[98].

---

[97] See: Section 2.2.

[98] It should be stressed that the rights relating to science, in its various meanings (access to science, availability of scientific results, up to academic and dissemination freedoms), also have a national dimension and are generally covered by Constitutions. For instance, the Italian Constitution mentions freedom of art and science in the Article 33 ("*L'arte e la scienza sono libere e libero ne è l'insegnamento*"); the Spanish Constitution of 1978 devotes the Article 44 specifically to access to culture and science: "*Los poderes públicos promoverán y tutelarán el acceso a la cultura, a la que todos tienen derecho. Los poderes públicos promoverán la ciencia y la investigación científica y técnica en beneficio del interés general.*"; the Bulgarian Constitution enshrined in the Article 23 the right to science, stating: "The State shall establish conditions conducive to the free development of science, education and the arts, and shall assist that development."; the *Grudgesetz* (the German Fundamental Law of 1949) enshrines the freedom of art and science in the Article 5, which is dedicated to freedom of expression. On the consitutional foundations of the right to scientific research in France, see: STÉPAHN MOUTON, *Les fondementes consitututionnels de la liberté de la recherche* (Toulouse: Presses de l'Université de Toulouse, 2005): 93-111. In particular: "*La première base constitutionnelle de la liberté de la recherche est constituée de la liberté de conscience et d'opinion d'une part et de la liberté de communication d'autre part*", underling the role of the "Conseil Constitutionnel": Conseil Constitutionnel, 23 November 1977,

3.3.1 International Dimension

The emergence of the Open Science paradigm should be framed in a wider debate related to the so-called "right to science"[99]. This expression refers to the right enshrined in the Article 27 of the 1948 Universal Declaration of Human Rights (the Declaration, hereinafter)[100], which states:

> 1. Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.
> 2. Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.

This Article encompasses multiple facets. The first paragraph takes into account the aspect of the transmission of knowledge: both the individual right to participate in the enjoyment of art and scientific developments, and the collective right to take an active role within the community in the development of the cultural life.

The second paragraph, on the other hand, refers to a different aspect, namely the production of knowledge: the individual right of the author to protect her own interests – moral or material – is enshrined. This second paragraph can be identified as the legal basis for the different forms of intellectual property protection[101], albeit differently articulated in the various legal systems of reference[102].

n. 77-87 sur la liberté d'enseignement, https://www.conseil-constitutionnel.fr/decision/1977/7787DC.htm; Conseil Constitutionnel, 20 janvier 1984, n. 83-165 sur le libertés universitaires, https://www.conseil-constitutionnel.fr/decision/1984/83165DC.htm.

[99] On an overview of the right to science, its origins and how it should be re-identified today see: HELLE PORSDAM, SEBASTIAN PORSDAM MANN (eds.) *The Right to Science. Then and Now*. (Cambridge: CUP, 2021).

[100] United Nations, Universal Declaration of Human Rights, 1948, https://www.un.org/en/about-us/universal-declaration-of-human-rights.

[101] The close connection between the right to science and the field of intellectual property is explored in more detail in: LEA SHAVER, "The right to science and culture." *Wisconsin Law Review* (2010): 121-184.

[102] On this aspect, see: Section 2.3.3 "Copyright and Ownership of Data".

This right, in 1966, was incorporated into the International Covenant on Economic, Social and Cultural Rights (the Covenant, hereinafter)[103], issued by the Economic and Social Council of the United Nations, which came into force in 1976, voluntarily joined by 169 States so far. The Article 15 of the Covenant incorporates the main assumptions of the Article 27 of the Declaration. The first paragraph, in fact, states:

> The States Parties to the present Covenant recognize the right of everyone:
> (a) To take part in cultural life;
> (b) To enjoy the benefits of scientific progress and its applications;
> (c) To benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he is the author.

The following three paragraphs of the Article 15 of the Covenant enhances the content of the right enshrined in the Article 27 of the Declaration. The third paragraph affirms that "[T]he States Parties to the present Covenant undertake to respect the freedom indispensable for scientific research and creative activity" and brings the so-called "academic freedom" under the right to science. Finally, the fourth paragraph recognises the benefits deriving from international cooperation in the scientific and cultural fields, emphasising the collaborative nature of science.

In light of both, the Article 27 of the Declaration and the Article 15 of the Covenant, the right to science has been recognised in international law.

Despite this recognition, for a long time this right appeared unclear and blurred, in two main aspects. First, it received little

[103] United Nations, International Covenant on Economic, Social and Cultural Rights, 1966 https://www.ohchr.org/en/professionalinterest/pages/cescr.aspx.

consideration[104], perceived more as an ideal to strive for[105] or, at best, as a less relevant right. Second, although the right to science is intuitively easy to identify, it has never been given a precise legal definition[106].

The definition of the right content helps to make it effectively operational and enforceable[107]. For this reason, in October 2018, the UN Committee on Economic, Social and Cultural Rights undertook an initiative with the aim to define the right to science, which would be all-encompassing of its various dimensions and meanings. These efforts led to the release of the General Comment No. 25 on science and economic, social, and cultural rights (the Comment, hereinafter), on 30 April 2020[108]. The Comment frames the right to science but, at the

---

[104] AUDREY CHAPMAN, JESSICA WYNDHAM, "A human right to science." *Science* 340.6138 (2013): 1291-1291, doi: 1126/science.1233319; VALENTINA ZAMBRANO, "Il «diritto umano alla scienza» e l'emergenza da CoViD-19." *BioLaw Journal-Rivista di BioDiritto* 20.1 (2020): 259-267, doi: 10.15168/2284-4503-584.

[105] This is not to say that the right to science has not been the subject of international interest. In this regard, in fact, there are some UN resolutions, collected by the American Association for the Advancement of Science, see: https://www.aaas.org/resources/Article-15/resources-international-human-rights-documents. The nature of these interventions for a long time, however, was very general, with basic principles. Consider the UN General Assembly resolution, Human rights and scientific and technological progress, UN Doc. A/RES/48/140, 7 March 1994. The resolution stressed the fact that "[…] many advances in scientific knowledge and technology in health, education, housing and other social spheres should be readily available to the populations as the heritage of humanity […]". The role of the Member States in adopting measures to improve the quality of their services was also highlighted "[…] to ensure that the results of science and technology are used only for the benefit of the human being and do not lead to the disturbance of the ecological environment, that is, *inter alia*, measures against the illicit dumping of toxic and dangerous products and waste", introducing the issue of environmental protection in scientific research.

[106] "[…] the right to science has never been legally defined and is often ignored in practice by the governments bound to implement it." in JESSICA M. WYNDHAM, MARGARET WEIGERS VITULLO, "Define the human right to science." *Science* 362.6418 (2018): 975-975, doi: 10.1126/science.aaw1467.

[107] Some scholars have argued that some challenges arise from the absence of a definition of scientific research under international law, also highlighting another consequence: "The lack of a clear and precise delimitation of the boundaries within which the concept of scientific research under international law extends thus leaves room for abusive behaviour.", in CHIARA RAGNI, *Scienza, diritto e giustizia internazionale.* (Milano: Giuffrè, 2020), p. 46. [Translation from the Italian original text].

[108] It should be considered also the previous General Comment No. 17 to the ICESCR, "The right of everyone to benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic productions of which he or she is the author," (Article 15, paragraph 1(c), of the Covenant), 12 January 2006. That Comment, however, delved more deeply into aspects related to Intellectual Property, mainly analysing paragraph 1(c) of Article 15. Specifically, this Comment stressed the distinction between the human right of protection of the author as creator and the legal

same time, deepens the relationship between science and economic, social and cultural rights. If this right has been considered for decades as a not enforceable right, or as an ideal to be attained, the Comment intends to lay the foundations for an action committed to the substantial implementation of the right itself: "States parties must take steps, to the maximum of their available resources, for the full realisation of the right to participate in and to enjoy the benefits of scientific progress and its applications. While full realisation of the right may be achieved progressively, steps towards it must be taken immediately or within a reasonably short period of time"[109]. The Comment also pointed out that the States initiatives, in order to become concrete, must start from the definition of "budgetary measures" or, in other words, from the allocation of investments.

Even though this Comment is the result of a process that began in 2018, it was published at the middle of the COVID-19 pandemic. Once again, the year 2020 is identified as a turning point for the field of scientific research[110]: science became a widely debated topic for the public at large, and also, the discussion has been reignited among experts at international level. Many aspects overlooked for decades immediately acquired new relevance.

The Comment, dealing with "the right of everyone to enjoy the benefits of scientific progress and its applications", represents a real milestone and foundation for Open Science. Yet, the Comment only twice mentions the expression Open Science, unfortunately without defining this expression.

The Paragraph 16 of the Comment focuses on the obligation of States Parties to take an active role in the preservation, development and

---

entitlements protected in the different legal systems by the intellectual property disciplines.

[109] General comment No. 25 to the ICESCR, "On science and economic, social and cultural rights," (Article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights), 30 April 2020. In particular, paragraph 23.

[110] On this aspect, see: Section 2.2.5 "The Year 2020: A Tipping Point for Open Science".

dissemination of science[111], stating that: "[…] States should promote open science and open source publication of research. Research findings and research data funded by States should be accessible to the public"[112].

The second mention of the term Open Science occurs in paragraph 49, where the primary role of States in promoting citizen participation in science is stressed: "[B]asic knowledge of science, its methods and results, has become an essential element for being an empowered citizen and for the exercise of other rights, such as access to decent work. States must exert every effort to ensure equitable and open access to scientific literature, data and content, including by removing barriers to publishing, sharing and archiving scientific outputs. However, open science cannot be achieved by the State alone"[113]. This last sentence of the UN Comment underlines a need that in this dissertation has been argued as a matter of legal coordination of systems, levels and actors[114].

The Open Science paradigm in the Comment, however, is not only strictly referred to these two statements, namely to the literal mentions. In fact, the Open Science claim hinges on the entire *corpus* of the Comment. These two mentions of the term do not evoke the complete and multifaceted phenomenon described in our interpretation of the Open Science paradigm as the Open Scientific Research Process. It appears, rather, to refer only to certain aspects of the phenomenon, which can be related to the so-called "democratic school", exposed in Figure 3.1, which identifies Open Science as the means by which to

[111] In accordance with the relevant "UNESCO Recommendation on Science and Scientific Researchers", adopted on 13 November 2017 (Annex II), where paragraphs 25-26 state: "Member States should develop policies for the protection and preservation of research objects, scientific infrastructure and scientific archives, including in instances of conflict. Member States should establish as a norm for any scientific publishing, including publishing in open access journals, that peer review based on established quality standards for science is essential". Available at: https://unesdoc.unesco.org/ark:/48223/pf0000260889.page=116.
[112] General comment No. 25 to the ICESCR, par. 16.
[113] General comment No. 25 to the ICESCR, par. 49.
[114] See: Section 2.4.3.

strengthen the democratic and informed participation of citizens within society.

Rather, the backbone of the Comment appears profoundly characterised by the tenets of Open Science. Consider that the core of the debate concerns the intervention of the States, in order to put in place concrete and effective steps to make the right to science operational. The concept of "benefits" of scientific advancement is analysed in its threefold dimension:

> The term "benefits" refers first to the material results of the applications of scientific research, such as vaccinations, fertilizers, technological instruments and the like. Secondly, benefits refer to the scientific knowledge and information directly deriving from scientific activity, as science provides benefits through the development and dissemination of the knowledge itself. Lastly, benefits refer also to the role of science in forming critical and responsible citizens who are able to participate fully in a democratic society.

It is precisely the achievement of these three benefits that institutions at all levels should think about when identifying Open Science policies. In other words, the Open Science as the Open Scientific Research Process is interpreted – or should be interpreted – as the set of policies enforcing the right to science as enshrined in the Article 27 of the Declaration, reinforced by the Article 15 of the Covenant and developed in the Comment.

As an illustration of this implicit correlation between the right to science and the Open Science paradigm within the Comment, consider paragraph 21, concerning the limitations of the right to science. These constraints on the right to science can be understood as a more precise representation of the formula "as open as possible as closed as necessary", at the basis of the Open Science paradigm: "[…] first, limitations have to be determined by law; second, they must promote 'the general welfare in a democratic society'; and third, any restriction must be compatible with the nature of the right restricted"[115]. The limitations to the right to science, therefore, are particularly restrictive. As a result: "[…] where there are several means reasonably capable of achieving the legitimate aim of the limitation, the one that is least restrictive to economic, social and cultural rights must be selected, and the burdens imposed on the enjoyment of the right should not outweigh the benefits of the limitation"[116].

In other words, it seems fair to admit that the right to science, enshrined in the Article 27 of the Declaration, and complemented by the Article 15 of the Covenant, is not only at the basis of current Open Science policies, but can even be considered as an enforceable right to *open* science[117]. In fact, it has been argued in many contexts, among scholars[118] and by institutions[119], that over time, when the paradigm of Open Science will be fully established and implemented, it will lose the connotation "open" to become normally what is considered science.

This perspective, based on the international framework of the right to science, determines the direction that Open Science policies must – or should – take.

As an illustration of this need consider "the shared appeal on Open Science by UNESCO, WHO, CERN and the Office of the United Nations High Commissioner for Human Rights" launched on 27

---

[115] General comment No. 25 to the ICESCR, par. 21.

[116] *Ibid.*

[117] Similar interpretations are exposed in: FEDERICO BINDA, ROBERTO CASO, "Il Diritto Umano alla Scienza Aperta." *Trento Law and Technology Research Group Research Paper Series*, 41 (2020): 1-6; and also in EFFY VAYENA, JOHN TASIOULAS, "We the Scientists: A Human Right to Citizen Science." *Philosophy & Technology* 28 (2015): 479-485, doi: 10.1007/s13347-015-0204-0.

[118] For instance, JEAN-CLAUDE BURGELMAN, *et al.*, "Open science, open data and open scholarship: European policies to make science fit for the 21st century." *Frontiers in Big Data* 2.43 (2019): 1-6. doi: 10.3389/fdata.2019.00043.

[119] For instance, the European Research and Innovation Days, the annual policy event of the European Commission, in the 2019 edition, held on 24-26 September 2019 in Brussels, Belgium, hosted a session entitled "Open Science is the New Normal", arguing that: "Open Science is becoming the new normal. This leads to a revolution in research practices, where our unprecedented ability to access, combine and process heterogeneous information transforms the research landscape and accelerates the pace of innovation.", available at: https://www.eoscsecretariat.eu/events/european-research-and-innovation-days.

October 2020. The purpose has been to issue the first UNESCO recommendation on Open Science, in November 2021[120]. This Recommendation refers many times to the concept of Open Science as outlined by the European institutions in recent years[121].

It is suggested to consider the Recommendation as an effective and explicit convergence between the right to science and the paradigm of Open Science, interpreted as the Open Scientific Research Process, described in the previous Section.

However, beyond this general intention, the main purpose of the Recommendation is "[…] to provide an international framework for Open Science policy and practice that recognises regional differences in Open Science perspectives". In other words, the intention is rather to strike a balance between the differences in the advancement of Open Science among distinct geographical areas of the world. Considerable importance is given to the regional and local aspects of the advancement of Open Science, and to their specificities, defining a set of common principles and values, helping to harmonise different situations, without undermining them[122].

In light of the analysis on the human rights framework in the field of science at international level, next section addresses the European dimension.

3.3.2 European Dimension

Traditionally, the four pillars of the European Union are the free movement of goods, services, persons, and capital. Since the Ljubljana

[120] UNESCO, "UNESCO Recommendation on Open Science." (2021), https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en.

[121] In Section 2.1.2 and Section 2.4.3.3, this first draft of the UNESCO recommendation on Open Science was mentioned precisely as an example of the leading position of European policies at the international level.

[122] One of the aspects that emerges clearly from the Recommendation, linked to the enhancement of local specificities, is the value of multilingualism as a founding principle of the Open Science paradigm. On this topic and its relevance, in particular in the European context, see: GINEVRA PERUGINELLI, *Multilinguismo e sistemi di accesso all'informazione giuridica* (Milano: Giuffrè Editore, 2009), pp. 11-26.

Council of 2008, a fifth freedom of movement must be included in the set of freedoms of the European Union: the free movement of knowledge and researchers[123]. In the Conclusions issued by the EU Council, on the definition of a "2020 Vision for the European Research Area"[124], it is stated that the key purpose is to enable and foster the circulation of both researchers and knowledge: "[P]layers are able to access, manage and share knowledge (including via open access) across the ERA using interoperable high-performance information systems. European research institutions provide attractive working conditions for researchers from all parts of the world, both men and women, in the framework of a single labour market which enables mobility between countries and sectors with minimal financial or administrative obstacles"[125].

The fifth freedom of movement of the EU, the freedom of knowledge circulation, is built on the Articles 179-180 of the TFEU[126]. The Article 179 TFEU, besides stating that the Union's purpose is to create a European Research Area[127], emphasises the need for cooperation among researchers. This is confirmed by the Article 180 of the TFEU which, in addition to setting the aim of "[…] dissemination and optimisation of the results of activities in Union research, technological development and demonstration"[128], it also pointed out the need to "[…] promotion of cooperation in the field of Union research, technological development and demonstration"[129].

Another fundamental aspect that shapes the freedom of circulation of knowledge is the so-called "academic freedom", enshrined in the

---

[123] See: Section 2.2.1.

[124] Conclusions of the Council on the definition of a 2020 vision for the European research area OJ C 25, 2009/C 25/01, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52009XG0131(01).

[125] Conclusion of the Council, 2009/C 25/01, p. 4.

[126] Consolidated version of the Treaty on the Functioning of the European Union (TFEU), OJ C 326, 26.10.2012, p. 47–390, http://data.europa.eu/eli/treaty/tfeu_2012/oj.

[127] See: Section 2.2.1 and Section 2.2.5, where the establishment of the European Research Area, ERA, is discussed in detail.

[128] Article 180(1), lett. *c*, TFEU.

[129] Article 180(1), lett. *b*, TFEU.

Article 13 of the Charter of Fundamental Rights of the European Union (the Charter, hereinafter), entitled "Freedom of the arts and sciences"[130]. It states that "[T]he arts and scientific research shall be free of constraint. Academic freedom shall be respected".

The academic freedom becomes a fundamental condition for the full realisation of the freedom of movement of knowledge[131]. The reason is that academic freedom encompasses a threefold nature: first, it includes (i) the right of academics to do research; second, (ii) the right to disseminate the results of their research; and, finally, (iii) the right to do education, whether to students or to civil society[132].

This framework of European rights has been reinforced over time by case law. The European Court of Justice (ECJ) has clarified the outlines of this freedom of circulation of knowledge – albeit indirectly – with specific reference to academic freedom.

---

[130] In this dissertation, the specific aim is to link academic freedom and the creation of the ERA to the fifth European freedom, the so-called circulation of knowledge. It is pointed out that, in general, even if one does not accept the vision of such a fifth freedom of movement, referring to the traditional four European freedoms of movement, the identification of an open research area in Europe, such as the ERA, is embedded in the Article 45 TFEU, concerning the freedom of movement of workers – who in this case are research workers: "[…] is the system of the European Union, enabling the universities of Europe to be configured as a 'network system' based on the mobility guaranteed to research workers (Article 45 TFEU) […]."; and also "The citizens of the Union are recognised as having academic freedom in accordance with the provisions of the internal market, first and foremost as a 'corollary' of freedom of movement, in compliance with a model of establishing rights based on the purely economic connotation of the origins of the economic system.", in BARBARA GAGLIARDI, *La tutela amministrativa della libertà accademica.* (Milano: CEDAM, 2018), pp. 12, 101. [Translation from the Italian original text].

[131] In addition, it has also been defined as a proper subjective right, as argued in: GIANMARIA AJANI, ROBERTO CAVALLO PERIN, BARBARA GAGLIARDI, "L'Università: un'amministrazione pubblica particolare." *Federalismi.it* 14 (2017): 2-16. On the boundaries of the academic freedom, see: ROBERTO CAVALLO PERIN, "Il Contributo Italiano alla Libertà di Scienza nel Sistema delle Libertà Costituzionali" *Diritto Amministrativo*, 3 (2021): 587-620.

[132] See the Declaration of Principles on Academic Freedom and Academic Tenure, provided by the American Association of University Professors (AAUP), in 1915: "Academic freedom in this sense comprises three elements: freedom of inquiry and research; freedom of teaching within the university or college; and freedom of extramural utterance and action", available at: http://www.aaup-ui.org/Documents/Principles/Gen_Dec_Princ.pdf.

As an illustration, consider the recent case *European Commission v. Hungary*[133]. The case involved the violation of the EU Law and the GATS Agreements[134] by a Hungarian law of 2017, which amended the Hungary's Higher Education Act of 2011, restricting the freedom of establishment of higher education institutions on Hungarian territory. The ECJ had the opportunity to emphasise that: "[…] academic freedom in research and in teaching should guarantee freedom of expression and of action, freedom to disseminate information and freedom to conduct research and to distribute knowledge and truth without restriction". In the ruling, the freedom to establish higher education institutions in any EU Member State is linked to the academic freedom enshrined in the Article 13 of the Charter. The Grand Chamber's reasoning thus becomes an emblematic example of the role of the Article 13 of the Charter as means by which the fifth European freedom of movement can be realised.

Against this backdrop, the Open Science paradigm, understood as the Open Scientific Research Process, can be interpreted as an attempt to implement in practice the rights enshrined at the international level by the right to science and at the European level by the fifth freedom of circulation of knowledge.

In addition, this interpretation endows science with a specific role: science can be considered a field capable of giving an unprecedented and renewed boost to the process of European integration; but it can also represent the ground on which to build an effective international cooperation.

The analysis of the origin and evolution of the Open Science concept (Section 3.1), of its constitutive elements (Section 3.2) and of its fundamental and human rights ground (Section 3.3) represent the investigation of the development phase of the phenomenon.

[133] Case C-66/18 *European Commission v Hungary* (2020) ECJ, ECLI: EU:C:2020:792.
[134] The General Agreement on Trade in Services (GATS), the treaty of the World Trade Organization (WTO).

This investigation has enabled the conceptual foundations for the next phase of implementation of the paradigm, i.e., how the Open Science works in practice, which is illustrated in the next section.

3.4 Open Science at Work

Albeit the idea of the openness of the scientific research has ancient roots, the Open Science phenomenon under investigation is very recent: it started to acquire its current shape in the first decade of the 2000s, to become fully recognised and recognisable around 2015.

Since there isn't a unique and commonly accepted definition of the expression, this study focused on many different contributions: from literature (e.g., Vincente-Sáez and Martínez-Fuentes or Bartling and Friesike); entities (e.g., OECD); and projects (e.g., FOSTER).

In light of this analysis, then, a new definition of the Open Science has been proposed, adopting the Floirdi's method of LoA. The Open Science paradigm has been identified in terms of Open Scientific Research Process, characterised by five observables (i.e., resources; actors; methods; tools; benefits) that represent the constitutive elements of the phenomenon.

The implementation is the subsequent stage: clarified origin, evolution, definition and legal basis of the Open Science, the deployment stage is focused on the implementation of the paradigm. The phenomenon of Open Science, which originally stemmed from the scientific community[135], now acquires a more comprehensive and structured shape, through the involvement of institutions. This formalisation has been achieved (and is still ongoing) by the identification of a set of strategies and initiatives designed to strengthen all the various dimensions of a heterogeneous phenomenon such as the Open Science. Due to the interventions of institutions, the claims that just a few years ago were made by the most innovative

[135] See: Section 3.1.2 "Beginning: The Open Access Initiatives".

academics, now become formalised as an integral part of the European governance of scientific research[136]. This phase definitively marks the overcoming of the vision of scientific and academic freedom as a negative freedom, limited to the sphere of non-interference[137].

Open science is therefore the result of the bottom-up efforts by the scientific community and the top-down commitments, by institutions. Specifically, Section 3.4.1 sheds light on the role of the scientific community in this implementation phase.

However, some problems still need to be solved and some undeniable contradictions still persist. For this reason, Section 3.4.2 illustrates the possible pitfalls into which the implementation of Open Science risks falling.

3.4.1 Scientific Community in Open Science Implementation

One of the fundamental actors of the implementation stage, understood as a turning point of the Open Science, is the scientific community.

First, as repeatedly stressed, the Open Science paradigm arose within the scientific community[138], based on a bottom-up approach[139].

---

[136] For instance, think about the impact of Open Science on the new funding programme 2021-2027, i.e., the Horizon Europe programme, see: Section 2.2.6.

[137] Traditionally, freedom of scientific research or academic freedom has been excluded from the list of social rights because of its connection with negative freedoms. The limits of interpretation are currently evident, see: BARBARA GAGLIARDI, *La tutela amministrativa della libertà accademica, op cit.,* 98. Going further, in this dissertation it is believed that with the recognition and support of European institutions of this new emerging paradigm of science, i.e., the Open Science paradigm, this traditional interpretation might be considered definitively overtaken.

[138] Consider the analysis of the evolution of the concept, in Section 3.1, and the role of the Open Access movement at the beginning of the 2000s.

[139] Open Science expert Jean-Claude Burgelman underlines the bottom-up approach of the EOSC too, stressing that: "[…] the 'midwives' were the 100ds of scientists and science actors consulted along the way in 2014 – 2015 whose ideas on the future of European science were crystallized into a European policy. Without these 'midwives' – and the supportive Director General – we would never have been confident to defend – quite often or most of the time against several odds – EOSC. These "midwives" allowed us to say that EOSC was not an invention of some bureaucrats not knowing what to do, but responded to the desire of a large community about the future of European science (and not technologies). That is why from then on we inserted in all our presentation a standard slide – showing the grand place of Brussels – with the message: EOSC: not a cloud invented in Brussels.", in JEAN-CLAUDE BURGELMAN, "Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story)." *Data Intelligence* 3.1 (2021): 5-19, doi: 10.1162/dint_a_00069.

Thus, in the deployment stage, the scientific community should play an important role in the Open Science governance[140].

Second, the scientific community itself triggered the transformation and the consequent interventions of institutions, in order to put in practice and empower this emerging vision of science, which necessarily was conflicting with the traditional one[141].

Finally, bear in mind that this new phase of the deployment of Open Science is achieved in accordance with and – above all – *for* the scientific community.

Although there are many actors involved (service providers, industries, etc.)[142] and many corresponding interests at stake, the scientific community is the most important player. In other words, the focus should be on the community of scientists who, according to the philosopher Gaston Bachelard, are engaged in the scientific development and in the definition of the *connaissance scientifique*[143].

For all these reasons, the scientific community should play a fundamental role in the governance of scientific research. European institutions have the opportunity to play a pro-active and coordinating role that is essential to support the paradigm shift that science is facing[144]. However, this will only be possible if the European Union

---

[140] As an illustration, consider the Lindau Nobel Laureate Meetings. This is a meeting that has been held annually in the city of Lindau, Bavaria, since 1951, with the aim of gathering together Nobel Prize winners and young researchers to stimulate the exchange of ideas between different generations and scientific domains. At the 2018 Lindau Nobel Laureate Meeting, the Nobel laureate in physiology and medicine, Elizabeth Blackburn, set out a list of major goals for science in the so-called "Lindau Declaration", which have been released at the Lindau Nobel Laureate Meeting 2021: https://www.lindau-nobel.org/wp-content/uploads/2021/11/2021_Annual-Report_Web.pdf. The Lindau Declaration once again refers to the need for renewal and is, once again, the expression of a request formulated from the scientific community. Among the various purposes, there is the need to publish results in Open Access and the will to strengthen communication with society, as well as the intention to share more knowledge. See: https://www.lindau-nobel.org.

[141] Among others, see: FRANCESCA DI DONATO, *La scienza e la rete, op. cit.*, p. 106; MICHAEL NIELSEN, *Reinventing discovery, op. cit.,* p. 188.

[142] On the actors involved in the Open Science scenario, see: Section 2.4.3.

[143] GASTON BACHELARD, *Le rationalisme appliqué.* (Paris: Presses universitaires de France, 1949), pp. 124-142.

[144] The EOSC is seen, by the same bodies managing its implementation, as fundamental for the same

proves to be able to overcome certain hurdles, described in the next section.

3.4.2 Deployment Stage Pitfalls

The deployment stage of Open Science must avoid three risks: (i) the distortion of the original purposes; (ii) the bureaucratic degeneration; and (iii) the overlapping of Open Science with Open Innovation.

(i) The first risk is the possible distortion of the original purposes of Open Science, due to the institutional intervention, leading to a limitation of the phenomenon, rather than encouraging it. The intervention of institutions may distort Open Science insofar the original objectives pursued by the scientific community, with a bottom-up approach, are ignored, explicitly or implicitly. This risk may damage the core of the project. This occurs if other types of interests, different from those pursued by the scientific community, prevail (e.g., economic interests or lobby pressure)[145].

(ii) The second pitfall is well represented by a question posed by the political philosopher Maria Chiara Pievatolo: should Open Science be considered "human emancipation or bureaucratic serfdom"[146]? Starting from the assumption that Academia and scientific research are managed by administrative entities, Pievatolo believes that mandating

reasons: "EOSC will stimulate the cultural changes in the entire research ecosystem. Open Science, which is realised with the help of EOSC, is striving for better horizontal and vertical links between scientists, scientific institutions, research and data infrastructures, and interconnecting scientific disciplines. It equilibrates the traditional research outputs, such as publications, patents, etc., with other forms of research outputs, including, for example, data, software, including models, simulations and methodologies. Making these outputs as findable, accessible, interoperable and reusable (FAIR) as possible is therefore a key requirement in measuring and rewarding the contribution of research.", SRIA, pp. 142-143.

[145] An example of this distortion is the Editorial of the "Nature Neuroscience" already mentioned before, in which the prestigious scientific journal declared the intention to opt for Open Access, setting the price for publishing each article in open access at €9800. This distorting operation is done by stating that "Of particular note, our transformative model is compliant with Plan S, the mandate for open access publishing laid out by cOAlition S, a large group of governmental and philanthropic funders.", see: "Nature Neuroscience offers open access publishing." *Nature Neuroscience* 25, 1 (2022): p. 1, doi: 10.1038/s41593-021-00995-2.

[146] MARIA CHIARA PIEVATOLO, "Open science: human emancipation or bureaucratic serfdom?." *SCIRES-it* (2019): 1-25.

Open Science to institutions, in some ways, risks reducing it "[…] to a bureaucratic, commodified enterprise whose horizon is not the advancement of learning – or discoveries and revolutions yet to do – but the production of information and data whose goal is not determined by the will to knowledge any longer, but by economic and political powers"[147].

The risk expressed by the philosopher is real and concrete: this is the reason why the governance of scientific research in Europe is so relevant. It is essential, in fact, that the action of institutions is always led by the original claims of the bottom-up approach, supported by the scientific community.

(iii) Finally, the third drawback is closely related to what has been defined in Section 2.2.5, as the "von der Leyen Doctrine"[148]. This expression refers to the political vision being pursued by the President of the European Commission Ursula von der Leyen: this vision, focused on the data-driven economy, supports the sharing and reuse of data, including scientific data, to enable their economic value to emerge, in support of European companies[149]. Although this aspect is essential and should be supported, on the other hand there is a risk that Open Science will be limited to this economic dimension: a reduction of the many and varied facets of Open Science to the so-called Open Innovation[150].

Bear in mind that the link between science and the market must necessarily be a component of these strategies. As pointed out by Nielsen, scientific research without a related market, fails: "[T]he

---

147 MARIA CHIARA PIEVATOLO*, "Open science", op. cit.*, p. 2.

148 On the "von der Leyen Doctrine", see: Section 2.2.5.

149 Jean-Claude Burgelman refers to this vision using the expression "science industry", in JEAN-CLAUDE BURGELMAN, *"Politics and Open Science", op. cit.*, p. 7.

150 Among many different definitions: "Open innovation is defined as systematically performing knowledge exploration, retention, and exploitation inside and outside an organisation's boundaries throughout the innovation process", in: ULRICH LICHTENTHALER, "Open innovation: Past research, current debates, and future directions." *Academy of management perspectives* 25.1 (2011): p. 77, doi: 10.5465/amp.25.1.75.

importance of the market to the role of science is vividly illustrated by what happened when the market was suppressed in the Soviet Union. Although the Soviet Union had one of the best scientific research systems in the world, without a market system it was mostly unable to make scientific innovations available to its citizens"[151]. Although the economic dimension is fundamental, extremising this vision risks undermining the other dimensions, such as those recognised by the framework of the human and fundamental rights of science[152].

In light of these risks, the interventions of institutions in the Open Science field should be understood as:

- the process through which to support the emergence of the new paradigm of science of our time, i.e., the Open Science paradigm;
- a method to jointly consider all the multiple dimensions and facets of the Open Science;
- an expression both of the "von der Leyen Doctrine" directed towards Open Innovation, but also as a manner of enforcing the international and European framework of human and fundamental rights connected to science.

As Nielsen reminds us: "[…] science isn't just an interest group. It's a way of understanding the world. Ideally, our institutions for governance would incorporate in public policy the knowledge gained by science – as imperfect, uncertain, and provisional as that knowledge is – as well as possible"[153].

---

151 MICHAEL NIELSEN, *Reinventing discovery, op. cit.,* p. 158.

152 On this aspect, Pievatolo explains "[…] the Humboldtian purpose of education, *Bildung*, conceived as an 'active process of appropriating the world' by developing the inner freedom of persons whose sense cannot be wholly determined by society and the market", in MARIA CHIARA PIEVATOLO, "Open science", *op. cit.*, p. 12. On the related topic of the interplay between public and private sector in science, see: Section 6.3.

153 MICHAEL NIELSEN, *Reinventing discovery, op. cit.,* p. 157. On this aspect, see: JOE R. BIDEN, "Memorandum on Restoring Trust in Government Through Scientific Integrity and Evidence-Based Policymaking", The White House, January 27, 2020. https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/memorandum-on-restoring-trust-in-government-through-scientific-integrity-and-evidence-based-policymaking/. The aim pursued by the president of the United

The proposed interpretation of the Open Science paradigm as an Open Scientific Research Process, supported by institutions, has once again underlined the complexity of the governance of scientific research (in terms of levels, systems and actors involved[154]). The current situation characterised by the implementation of the European Open Science strategies is extremely sensitive considering the choices that will be made and the investments that will be allocated.

Making the deployment stage concrete and effective will depend on the ability of the actors involved to modulate the constitutive elements of the Open Science as the Open Scientific Research Process, in order to design the best possible governance[155].

3.5 Conclusive Remarks

Chapter 3 of this study, entitled "The Open Science paradigm", answered the RQ3:

> What should be understood by the Open Science paradigm? What are its foundations? Which are its different dimensions and how do they interact with each other?

Therefore, the chapter pursued four aims: (i) the first aim was to analyse the concept of Open Science, the origins of the expression and its evolution; (ii) the second aim was to propose a definition of Open Science, adopting the Floridi's method of LoA, in order to identify the constitutive elements of the phenomenon under investigation; (iii) the third goal was to ground the Open Science within the framework of human and fundamental rights; and, finally, (iv) the last aim was to discuss about the implementation phase of the Open Science paradigm, bringing out possible pitfalls.

States in this Memorandum was analysed at the beginning of Chapter 1.

[154] See: Section 2.4.3.

[155] Here the concept of co-ordination is implicitly evoked, which is exposed in Section 2.4.3 and in Chapter 7.

It is suggested to adopt the following definition of Open Science:

> the Open Science describes the Open Scientific Research Process in which openness regards (i) the inputs, i.e., the underlying resources, such as data and economic funding; (ii) actors, including public at large; (iii) methods, in terms of evaluation and verifiability; (iv) tools; and (v) benefits deriving from this process, including the dissemination of the results, through publication, teaching activities or more.

The Open Science paradigm is rooted in the human and fundamental rights framework. For this reason, the Open Science policies should be considered as the set of interventions and strategies effectively enforcing the right to science as enshrined in the Article 27 of the Declaration of Human Rights, reinforced by the Article 15 of the Covenant on Economic, Social and Cultural Rights and developed in the UN Comment n. 25 of 2020. Going further, the right to science is not only at the basis of the current Open Science policies but may be interpreted as a real "right to *open* science".

Therefore, during the implementation phase of the Open Science, the approach of the EU institutions should be understood as:

- the process through which to support the emergence of the new paradigm of science of our time;
- a method to jointly consider all the multiple dimensions and facets of the Open Science;
- an expression both of the "von der Leyen Doctrine" directed towards Open Innovation, but also as a manner of enforcing the international and European framework of human and fundamental rights connected to science.

The issue at stake, therefore, is to assess whether and how this identified European attitude is applied in dealing with the different constitutive elements of the phenomenon. Specifically, next chapter assesses the variable $V_{1.1}$ of the observable (O_1) "Resources" as input of the Open Scientific Research Process: research data.

Chapter 4

Open Research Data

A profound transformation is currently taking place in the scientific domain due to the impact of new technologies. The new paradigm of science emerging from this notable change has been described in terms of "Open Scientific Research Process".

The concept of openness necessarily has different meanings, in relation to each constitutive elements of the Open Science paradigm, i.e., (O_1) Resources; (O_2) Actors; (O_3) Methods; (O_4) Tools; and (O_5) Benefits[1].

This chapter intends to investigate what is meant by openness in the context of research data, representing one of the resources needed to trigger the research process.

Section 4.1, initially, frames the role of data within the context of scientific research and identifies the specific features of the openness of research data, in terms of sharing and reuse. This latter analysis is carried out through an in-depth examination of one of the key concepts of the Open Science paradigm: the FAIR Data Principles[2]. After the investigation on the basic concepts of openness and FAIRness of data, the intention is to shed light on the distinction between the legal and the ethical level of the issues underlying the sharing and reuse of research data.

Afterwards, Section 4.2 intends to address the first set of issues at the basis of the openness of research data, namely the legal issues. This analysis starts by redefining the vague notion of "legal interoperability": this expression is generally used in the

[1] For an overview of the LoA "Open Scientific Research Process" and its observables, see: Section 3.2 and Figure 3.3.

[2] The FAIR Data Principles are introduced in Section 2.2.3 and in Section 2.3.1.

implementation of the EOSC project to summarise the legal and cultural challenges to be faced, in order to support the openness of research data.

Then, an overview of the European legal framework for Open Data is presented. Proceeding from the overall picture to the specifics, subsequently an in-depth analysis of the Article 10 of the Open Data Directive (ODD)[3], entitled "research data", is proposed. Consider that the Members States had time until the 17 July 2021 to transpose the Directive: the ODD entails considerable transformations in the field of sharing and reuse of both public sector data and research data.

The critical interpretation of the Article 10 of the ODD lays the groundwork to focus on the legal challenges posed by the openness of research data. Specifically, the analysis concerns: (i) the hurdles arising from the need for national transposition of the ODD; (ii) the limits of the enforceability of the openness "by default"; and, finally, (iii) the issues related to the choice of licences through which the openness of research data can be practically achieved.

Next, Section 4.3 addresses the other set of challenges related to the openness of research data, i.e., the ethical issues. Attention is drawn, first, to the concept of quality of research data: how is it identified? Who exercises control over the quality of the research data, whose sharing is intended to be promoted?

After answering these questions, the intention is to clarify the relationship between FAIRness and quality of research data, adopting Floridi's distinction between infraethics and ethics.

Finally, having shed light on the legal and ethical issues, Section 4.4 looks a little further into future developments, investigating the European Commission's new proposed regulation of the Data Governance Act (DGA). Specifically, attention is paid to the impact that

[3] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019, *on open data and the re-use of public sector information*, OJ L 172, 26.6.2019, 56–83, ELI: http://data.europa.eu/eli/dir/2019/1024/oj.

the DGA will have on the European Open Data legal framework, focusing on the new "data altruism" mechanism that the DGA is expected to introduce.

Section 4.5 concludes the chapter, summarising the main insights of the analysis on the open research data.

4.1 Data Processing within the Scientific Research Field

Data are a very relevant aspect of the scientific research since they can be one of the possible triggers of the research process itself. But what exactly are research data? How does this typology of data differ from any other data?

In order to answer these questions, first, the intention is to examine what is meant by research data (Section 4.1.1).

Second, a concept that underpins the processing of data in scientific research under the Open Science paradigm is introduced: the FAIR Data Principles (Section 4.1.2).

Once the terminology will be clarified, attention should be drawn to the management of research data within the Open Science scenario. In this regard, a preliminary issue is pointed out: often the ethical and legal challenges of research data tend to overlap, making it necessary to untangle these two different strands of investigation (Section 4.1.3).

4.1.1 Research Data

Research data are increasingly mentioned in any discussion about scientific research, or about the impact of digital ICTs on science. For this reason, it is necessary starting the analysis identifying what is meant by research data.

The use of data in scientific research is not new: science has always relied on data, by registering reality numerically, in order to support its theories, to refute others, or to trigger an investigation. What is new, rather, is the amount of data now available to science, i.e., the so-called

Big Data[4]. ICTs enables, first, to collect, record or create huge quantities of data and, second, to process them, in order to detect correlations, clusters, statistics, etc., up to identifying predictions on future trends and developments.

The entry "Scientific Research and Big Data" in the Stanford Encyclopedia of Philosophy states: "[T]he last few decades have witnessed the creation of novel ways to produce, store, and analyse data, culminating in the emergence of the field of data science, which brings together computational, algorithmic, statistical and mathematical techniques towards extrapolating knowledge from big data"[5].

There are many definitions of research data, from scholars, reports of institutions, and also from legal texts, but uniformity is lacking. Guibault and Wiebe, in their "Study on the protection of research data and recommendations for access and usage", report precisely the complexity of identifying a single definition of research data that can be used in different scientific contexts: "[F]rom a legal point of view, one of the very basic questions of this study is which kind of potentially protected data we are dealing with in the context of e-infrastructures for publications and research data [...]. The term 'research data' in this context does not seem to be very helpful, since there is no common definition of what research data basically is. It seems rather that every author or research study in this context uses its own definition of the

---

[4] On the definition of the concept of Big Data, among many, see: CLAUS BEISBART, "How Can Computer Simulations Produce New Knowledge?.", *European Journal for Philosophy of Science*, 2.3 (2012): 395-434. doi: 10.1007/s13194-012-0049-7; ROB KITCHIN, GAVIN MCARDLE, "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets", *Big Data & Society*, 3.1 (2016): 1-10. doi: 10.1177/2053951716631130; LUCIANO FLORIDI, "Big data and their epistemological challenge." *Philosophy & Technology* 25.4 (2012): 435-437, doi: 10.1007/s13347-012-0093-4. Specifically, Big Data has been characterised by the so-called "v-words": volume; velocity; variety; veracity; validity; volatility; value; among others, see: SABINA LEONELLI, "Global Data Quality Assessment and the Situated Nature of 'Best' Research Practices in Biology." *Data Science Journal*, 16.32 (2017): 1-11, doi: 10.5334/dsj-2017-032; SABINA LEONELLI, *La ricerca scientifica nell'era dei big data. Cinque modi in cui i Big Data danneggiano la scienza, e come salvarla*. (Milano: Meltemi Editore, 2018), pp. 8-15.

[5] SABINA LEONELLI, "Scientific Research and Big Data", *The Stanford Encyclopedia of Philosophy* (2020), https://plato.stanford.edu/entries/science-big-data/.

term. Therefore, the term 'research data' will not be strictly defined, but will include any kind of data produced in the course of scientific research, such as databases of raw data, tables, graphics, pictures or whatever else"[6].

Despite the difficulties of a common and unique definition, an attempt to define research data is carried out by the Open Data Directive (ODD)[7]. The Article 2(9) of the ODD defines research data as follows:

> 'research data' means documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results.

Similarly, the Organization for Economic Co-operation and Development (OECD), in its report "OECD Principles and Guidelines for Access to Research Data from Public Funding"[8], released in 2007, defines research data as follows:

> "research data" are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.

In light of the many definitions and interpretations of research data, certain common features characterising this type of data stand out. It may be worth exploring these characteristics, in order to identify a definition of research data, which will be beneficial for this

[6] LUCIE GUIBAULT, ANDREAS WIEBE, *Safe to be open: Study on the protection of research data and recommendations for access and usage*. (Göttingen: Universitätsverlag Göttingen, 2013), p. 17.

[7] Directive (EU) 2019/1024, *On open data and the re-use of public sector information, cit.*, ODD hereinafter.

[8] OECD, *OECD Principles and Guidelines for Access to Research Data from Public Fun*ding (Paris: OECD Publishing, 2007), p. 13, doi: 10.1787/9789264034020-en-fr.

dissertation. These elements, discussed below, are: (i) purpose; (ii) method; and (iii) reproducibility.

(i) Purpose. Research data do not differ from data collected and processed in other contexts, from the point of view of their nature: in any case they can be interpreted as factual records[9]. What primarily marks the difference is their purpose: research data are those data processed for the purpose of scientific research, broadly understood, whether carried out by public or private research centres. This does not mean that data can be defined as research data if and only if originally collected for that purpose. Research data may in fact also be data collected for other purposes and only then processed for scientific research purposes: the data can become research data through reuse or the so-called "secondary use" of data collected for a different purposes.

(ii) The method. Closely related to the purpose of the scientific research is the method by which the processing of the research data is carried out. Research data are part of a process of analysis traditionally called scientific method[10]. Even the expression "scientific method" can generate some ambiguity, given the considerable differences that exist in any field of knowledge. In general, however, for centuries the scientific method has been that method which permits the distinction between scientific and non-scientific or pseudoscientific activities[11].

---

[9] These factual records can vary in their nature: think about the distinction between data and metadata; or between raw and processed data, etc. Many data classifications are proposed by scholars, for example: JESSICA PARLAND-VON ESSEN, *et al*. "Supporting FAIR data: categorization of research data as a tool in data management." *Informaatiotutkimus* 37.4 (2018): pp. 135-17, doi: 10.23978/inf.77419. Here, a distinction is made between operational data; generic research data; and research data publication. Within this dissertation, research data are not identified on the basis of their type. Any mention of research data generally refers to the whole array of data possessing the three characteristics (purpose of scientific research; scientific method, and reproducibility) investigated in this Section 4.1.1.
[10] On the concept of scientific method, see: BRIAN HEPBURN, HANNE ANDERSEN, "Scientific method", *The Stanford Encyclopedia of Philosophy* (2021), https://plato.stanford.edu/entries/scientific-method/.
[11] The distinction between science and pseudoscience is also a topic that has been investigated for centuries. In this dissertation, the reference goes to the entry in the Stanford encyclopedia of

(iii) Reproducibility. The last fundamental feature is the reproducibility requirement: the processes of collection, elaboration, and analysis, as well as the assumptions based on these research data, must have undergone the critical review of the scientific community. In order to pass such a review, the experiments performed through the analysis and processing of the data should be reproducible[12] and, consequently, the research data should be available, for a long time[13]. Reproducibility can be considered as a fundamental element of the definition of research data, based on the assumption that "[S]cience should be available for evaluation by other scientists and for public scrutiny, just as it has been since Galileo's time. It should not be heading for epistemological suicide as a result of vested interests or a creeping loss of awareness of the theory of knowledge"[14].

Therefore, in light of the explored definitions of research data, and at the basis of the identified common features characterising this type

---

philosophy, in order to get an overview of the topic: SVEN OVE HANSSON, "Science and Pseudo-Science", *The Stanford Encyclopedia of Philosophy* (2021), https://plato.stanford.edu/entries/pseudo-science/; for a sociological analysis of the topic, see: GIUSEPPE TIPALDO, *La società della pseudoscienza.* (Bologna, Il mulino: 2019), 93-187. On an analysis of the criteria of scientificity, intended to determine the boundary between science and non-science, see: GIOVANNI BONIOLO, *Metodo e rappresentazione. Per un'altra filosofia della scienza.* (Milano: Mondadori, 1999), pp. 17-60.

[12] It should be noted that here the notion of "reproducibility" is used in a broad sense. On this topic, see the studies of Brian Nosek, specifically on the diversity between "reproducibility", understood as the process that allows the reproduction of the experiment with the same data, carried out with the same analyses; "robustness", understood as the procedure based on the same data, but carried out with different analyses; "replicability", understood as the procedure that proves the result of the experiment based on different data; and finally, "repeatability", as the general term that indicates all the previous control processes of science. See, for example: BRIAN A. NOSEK, *et al.* "Replicability, robustness, and reproducibility in psychological science." *Annual Review of Psychology*, 73, (2022): 1-94, doi: 10.1146/annurev-psych-020821-114157. See, also, GIOVANNI BONIOLO, *Il virus dell'idiozia. Sette scrittini su Covid-19, scienza, intellettuali e cittadini.* (Milano-Udine: Mimesis, 2021), p. 71, in which the Author clearly express the concept: "[…] a result, in order to be scientific, must be at least inter-subjectively verifiable (i.e., it must be replicable and reproducible) and therefore the theoretical context and the experimental method, that made it possible to find, it must always be made evident and transparent." [Translation from the Italian original text].

[13] The issue of reproducibility of the results and processes of science, in the Open Science paradigm, has already been mentioned in Section 3.2: specifically investigating the Observable O_3 Methods ($V_{3.2}$, Verifiability).

[14] GIOVANNI BONIOLO, THOMAS VACCARI, "Alarming shift away from sharing results." *Nature* 488.7410 (2012): p. 157.

of data, the following operational definition of research data can be adopted:

> Research data are factual records, (i) used for research purposes; (ii) as part of a process of analysis traditionally called scientific method; (iii) enabling the reproducibility of experiments, in order to ensure the review of the research output by the scientific community.

Once the notion of research data adopted in this investigation has been identified (i.e., in terms of data processed for scientific research purposes; adopting the scientific method; available in a way that guarantees reproducibility), it is now time to turn the attention to one of the milestones of the Open Science paradigm: the FAIR Data Principles.

4.1.2 FAIR Data Principles

The FAIR Data Principles indicate a set of guidelines to standardise the management of data – or more generally of digital objects[15] – in scientific research, from a technical point of view. FAIR is an acronym that stands for Findable, Accessible, Interoperable and Reusable: the features that represent the FAIR Data Principles.

In 2014, at the FAIRPORT Conference in Leiden, the concept of FAIR Data Principles began to take shape as a result of a discussion between scholars, scientific publishers and representatives of research centres. The conference report explains that the aim was to identify "[…] minimal but strict conventions that enable a wide variety of services and applications needed to realize computer-friendly as well as human-friendly data interoperability, stewardship and compliance against data and metadata standards, policies, and practices"[16]. The management of research data is therefore starting to become crucial.

---

[15] The concept of digital object is broader than the concept of data: "Central to the realisation of FAIR are FAIR Digital Objects, which may represent data, software or other research resources", in SANDRA COLLINS, *et al.*, "Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data." *Publications Office of the European Union* (2018), p. 8, doi: 10.2777/54599.
[16] "Jointly designing a DATA FAIRPORT", Conference Report Workshop at Lorentz Center @ Snellius, Leiden, 13-16 January 2014, 7, http://www.lorentzcenter.nl/lc/web/2014/602/info.php3?wsid=602. A

Then, the paper by Wilkinson *et al.*, published in 2016, represented the first formal publication on FAIR Data Principles[17]. The paper, besides describing the characters of the term FAIR, underlines two fundamental aspects: first, the nature of the FAIR Data Principles; and second, the goal pursued.

As regards the nature, it is made clear that the FAIR Data Principles are – precisely – principles and not standards: "These high-level FAIR Guiding Principles precede implementation choices, and do not suggest any specific technology, standard, or implementation-solution; moreover, the Principles are not, themselves, a standard or a specification"[18]. The idea is rather to have standards in data management set in the different research domains, which are inspired and guided by the FAIR Data Principles[19]. In fact, it is important to clarify one of the most common misunderstandings about the concept of FAIR Data Principles: these principles do not point to making open

---

fundamental role in the identification of the FAIR Data Principles was also played by the FORCE11 association, "[…] a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing. Individually and collectively, we aim to bring about a change in modern scholarly communications through the effective use of information technology", which also in 2014 issued a Joint Declaration: MARYANN MARTONE (ed.), *Data Citation Synthesis Group of FORCE11. Joint Declaration of Data Citation Principles*. (San Diego, CA: FORCE11, 2014), doi: 10.25490/a97f-egyk. Furthermore, the first steps towards the identification of data reuse policies were taken in 2016 at the G7 in Japan, where data reuse was defined as a priority, and at the G20 in China in the same year, where FAIR Data was specifically mentioned, see: BAREND MONS, *et al.*, "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." *Information Services & Use* 37.1 (2017): p. 49, doi: 10.3233/ISU-170824.

[17] MARK D WILKINSON, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1 (2016): 1-9, doi: 10.1038/sdata.2016.18.

[18] MARK D WILKINSON, *et al.*, *"The FAIR Guiding Principles"*, *op. cit.*, p. 5. Here it is evoked the concept of neutrality, already discussed in Section 2.2, applied to the EU Open Science policies.

[19] "Standards are prescriptive, while guidelines are permissive. We suggest that a variety of valuable standards can and should be developed, each of which is guided by the FAIR Principles.", in BAREND MONS, *et al.*, *"Cloudy, increasingly FAIR"*, *op. cit.*, p. 51.

research data[20], but rather aim to achieve a good data management, allowing for "long term care"[21].

Second, as regards the goals of the FAIR Data Principles, one of the most relevant – if not the most relevant – goal is that these principles aim to make research data readable both by humans and, above all, by machines. The huge amounts of Big Data are in fact analysed by machines, algorithms and AI systems: "This necessitates machines to be capable of autonomously and appropriately acting when faced with the wide range of types, formats, and access-mechanisms/protocols that will be encountered during their self-guided exploration of the global data ecosystem. It also necessitates that the machines keep an exquisite record of provenance such that the data they are collecting can be accurately and adequately cited"[22]. The possibility of having machine-readable data is a fundamental condition for allowing them to be open. Crucial is the definition of machine readability: "[…] we use the phrase 'machine actionable' to indicate a continuum of possible states wherein a digital object provides increasingly more detailed information to an autonomously acting, computational data explorer. This information enables the agent – to a degree dependent on the amount of detail provided – to have the capacity, when faced with a digital object never encountered before, to: a) identify the type of object (with respect to both structure and intent), b) determine if it is useful within the context of the agent's current task by interrogating metadata and/or data elements, c) determine if it is usable, with respect to license, consent, or other accessibility or use constraints, and

[20] "Although the FAIR principles apply to data regardless of their public availability and specifically do not require that data should be Open, this report considers what is needed to make data FAIR in the context of the EOSC and global drive towards Open Science. In that context, the implementation of FAIR data needs to go hand-in-hand with the principle that data created by publicly-funded research must be as Open as possible and as closed as necessary.", in SANDRA COLLINS, *et al.*, *"Turning FAIR into reality", op. cit.*, op. 10.

[21] It has been clarified that: "[…] science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments.", in MARK D WILKINSON, *et al.*, *"The FAIR Guiding Principles", op. cit.*, p. 1.

[22] MARK D WILKINSON, *et al.*, *"The FAIR Guiding Principles", op. cit.*, p. 3.

d) take appropriate action, in much the same manner that a human would"[23].

Specifically, what do the FAIR Data Principles indicate? In order to be findable, data must, for instance, be described by rich metadata, allowing third parties to effectively understand it; and furthermore, it must have a globally unique and persistent identifier. To be accessible, they must effectively be traceable through their identifier, using a standardised communication protocol, and the associated metadata must remain accessible even if the data are no longer available[24]. Interoperability refers to the language in which such data are released: primarily, it must be formal, accessible, and shared. Finally, the characteristic of reusability, much debated amongst experts[25], indicates a set of conditions intended to allow an effective reuse by third parties. Thus, data, in order to be defined reusable, must be released under a clear and easily accessible licence; have a clear description of its provenance; and comply with the standards of the relevant scientific community. In this way, a third party who finds and accesses data, understands it as interoperable, and clearly know what she has the power to do with it, e.g., by consulting its licence and investigating its provenance. Figure 4.1, taken from the Wilkinson *et al.* publication[26], provides a summary of the elements characterising each factor.

---

[23] *Ibid.*

[24] A clarification: research data must have this character of accessibility, in terms of traceability or, *potential* accessibility, even if not open. The fact that research data are accessible does not mean that they are open. On this topic, see also: Section 5.3.2.1.

[25] On the concept of reusability, see, among others: IGNASI LABASTIDA, THOMAS MARGONI, "Licensing FAIR data for reuse." *Data Intelligence* 2.1-2 (2020): 199-207, doi: 10.1162/dint_a_00042; MATTHEW WOLF, *et al.*, "Reusability first: toward FAIR workflows." *2021 IEEE International Conference on Cluster Computing CLUSTER* (2021): 444-455, doi: 10.1109/Cluster48925.2021.00053; ANGELA P. MURILLO, "An examination of scientific data repositories, data reusability, and the incorporation of FAIR." *Proceedings of the Association for Information Science and Technology* 57.1 (2020):1-3, doi: 10.1002/pra2.386.

[26] MARK D WILKINSON, *et al.*, *"The FAIR Guiding Principles", op. cit.*, p. 4.

Since 2016, there have been significant developments related to the FAIR Data Principles. In particular, the concept has been widely embraced by European institutions: already in 2016, the European Commission adopted the FAIR Data Principles when defining the guidelines for data management within the Horizon 2020 project[27]. Later, in 2018, the final report of the European Commission's experts group on FAIR Data has been released[28]: on this occasion, the adoption of the FAIR Data Principles as guidelines for the management of digital objects within the European Open Science Cloud (EOSC) was officially established[29]. As stated by the then Commissioner for Research, Science and Innovation, the aim was to present an action plan for the management of digital objects in the EOSC "[…] to make research data findable, accessible, interoperable and reusable (FAIR): attributes which are essential to extract the full scientific value from data resources and to unleash the potential for large-scale, machine-driven analysis"[30].

The FAIR Data Principles have become a real milestone in the development of the EOSC: one of the Working Groups is specifically dedicated to FAIR Data[31].

At the same time, alongside the developments by European institutions in the definition of the EOSC, several investigations have been conducted by scholars in relation to the implementation of the FAIR Data Principles, in different domains. For example, in the field

[27] European Commission, Directorate-General for Research & Innovation, *Guidelines on FAIR Data Management in Horizon 2020*, version 3.0 (2016), doi: 10.25607/OBP-774.
[28] SANDRA COLLINS, *et al.*, *"Turning FAIR into reality", op. cit.*
[29] On EOSC, in particular, see: Section 2.2.3 and 2.2.4.
[30] SANDRA COLLINS, *et al.*, *"Turning FAIR into reality", op. cit.*, p. 7.
[31] "The EOSC FAIR working group will provide recommendations on the implementation of Open and FAIR practices within the EOSC. It addresses cross-disciplinary interoperability, gather requirements relevant to the EOSC services, and advise the EOSC Executive and Governance Boards on FAIR-related matters.", in: https://www.eoscsecretariat.eu/working-groups/fair-working-group. The EOSC FAIR Working Group chair is Sarah Jones, digital curator at the University of Glasgow, UK, and Open Science advocate. Recently, the FAIR Data Working Group has published: EOSC Executive Board Working Groups FAIR and Architecture, "EOSC interoperability framework." *Publications Office of the European Union* (2021), doi: 10.2777/620649.

of genomics[32], a critical analysis on the concept of applied FAIR Data, led to the assumption that for this specific field "[T]he FAIR guiding principles constitute necessary, even if not sufficient, principles for responsible research data stewardship"[33].

Figure 4.1: The FAIR Guiding Principles in Wilkinson, et al. (2016)

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

---

[32] Note that the field of genomics is one of the most advanced in terms of data management and sharing. In fact, the scientific community in this field has been carrying out data sharing practices for a long time, at global level. For instance, the Global Alliance for Genomics and Health: https://www.ga4gh.org. This tradition is linked to the tradition of sharing in the field of genetics: the Nagoya Protocol on access and benefit-sharing on biodiversity, i.e., on access to genetic resources and the fair and equitable sharing of the benefits arising out of their utilization, was signed in 1993, in Rio de Janeiro, Brazil, see: https://www.cbd.int/abs/. On the analysis of the Nagoya Protocol, see: Chiara Ragni, *Scienza, diritto e giustizia internazionale*. (Milano: Giuffrè, 2020), pp. 50-53.

[33] Martin Boeckhout, Gerhard A. Zielhuis, Annelien L. Bredenoord, "The FAIR guiding principles for data stewardship: fair enough?." *European journal of human genetics* 26.7 (2018): 935, doi: 10.1038/s41431-018-0160-0.

Further studies, domain-based, have been conducted in recent years[34]. Such developments led the scientific community to come up with the idea of an Internet of FAIR Data and Services[35], i.e., a space in which "[…] the degree to which any piece of data is available, or even advertised as being available (via its metadata) is entirely at the discretion of the data owner"[36].

However, with the emergence of the FAIR Data Principles and the intention of the Open Science paradigm to make research data as open as possible, some challenges have been identified. Before proceeding with the investigation of these challenges, it is necessary to preliminarily shed light on the nature of these challenges.

4.1.3 A Preliminary Issue: Between Law and Ethics

In order to understand the challenges related to the openness and FAIRness of research data, it is important to briefly resume the premises from which the investigation about the European scenario of scientific research started, in Chapter 2. In other words, the intention is to apply these premises to the specific context of the research data, in order to shed light on the nature of the research data challenges.

The first premise, which has been called methodological, concerns the transformation that science has undergone due to the impact of digital ICTs. Both the way of doing science and the way of wondering about science have changed[37]. In Section 4.1.1, while defining the

---

[34] On material science see: CLAUDIA DRAXL, MATTHIAS SCHEFFLER, "NOMAD: The FAIR concept for big data-driven materials science." *Mrs Bulletin* 43.9 (2018): 676-682, doi: 10.1557/mrs.2018.208; on biopharmaceutical filed: JOHN WISE, *et al.,* "Implementation and relevance of FAIR data principles in biopharmaceutical R&D." *Drug discovery today* 24.4 (2019): 933-938, doi: 10.1016/j.drudis.2019.01.008; on the identification of *ad hoc* applications for FAIR Data (as an example of implementation) see: DAVID WILCOX, "Supporting FAIR data principles with Fedora." *LIBER Quarterly* 28.1 (2018), doi: 10.18352/lq.10247. Here Fedora is presented as "[…] a flexible, extensible, open source repository platform for managing, preserving, and providing access to digital content".

[35] BAREND MONS, *et al.*, *"Cloudy, increasingly FAIR", op. cit.*, p. 51; but also: PAOLO BUDRONI, JEAN CLAUDE-BURGELMAN, MICHEL SCHOUPPE, "Architectures of knowledge: the European open science cloud." *ABI Technik* 39.2 (2019): 130-141, doi: 10.1515/abitech-2019-2006.

[36] BAREND MONS, *et al.*, *"Cloudy, increasingly FAIR", op. cit.,* p. 51.

[37] On the methodological premise, see: Section 2.1.1.

concept of research data, the methodological premise has been declined precisely in relation to research data: in the input phase of the Open Scientific Research Process, for instance, the impact of Big Data triggered a profound transformation.

The second assumption is represented by the institutional premise[38]. It concerns the intention of institutions to strongly support the emergence of the new paradigm of science, namely the Open Science paradigm. In general terms, this intention emerges from the entire *corpus* of documents and strategies put in place by the European institutions from 2015 onwards[39]. Applying this institutional premise to the specific context of the research data, we attain a proper "right to access and reuse digital research data"[40]: it could therefore be defined as a right to *open* research data, subject to compliance with any restrictions set out in the terms of conditions.

In other words, regarding the research data, the first methodological premise is represented by the transformation of science determined by the massive use of Big Data and the necessary techniques to analyse it; and the institutional premise is represented by what the European institutions themselves define as a right of access to research data.

Consider that the Article 38 of the recent EU Regulation 2021/695[41], on the establishment of the Horizon Europe programme, set out that: "Open access to research data shall be the general rule under the terms

[38] On the institutional premise, see: Section 2.1.2.
[39] On the analysis of the entire European *corpus* of documents and strategies about Open Science, see: Section 2.2.
[40] European Commission, Directorate-General for Research & Innovation, *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*, version 3.2 (2017), p. 4, https://webgate.ec.europa.eu/funding-tenders-opportunities/display/OM/Online+Manual. The document refers to the "right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement". Consider that the mentioned "Grant Agreement" is part of the funding granted to research within the Horizon 2020 project, now replaced by Horizon Europe, discussed in Section 2.2.6.
[41] Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021, *establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013*, OJ L 170, 12.5.2021, p. 1–68, ELI: http://data.europa.eu/eli/reg/2021/695/oj.

and conditions laid down in the grant agreement, ensuring the possibility of exceptions following the principle 'as open as possible, as closed as necessary'"[42].

In light of these two premises, a set of issues arise related to the openness of research data and the adoption of the FAIR Data Principles.

Starting from the assumption that Open Data is not equivalent to FAIR Data, it is therefore crucial to clarify the nature of the challenges: ethical and legal issues should not be confused. As regards to open research data, in fact, there are two different types of challenges. The first type of challenges is legal: they are related to the European framework of Open Data and the specificities of the scientific research field.

The second type of challenges go beyond the legal domain, pertaining to ethics[43], and are mainly related to data quality. The equivalence between good research data and FAIR data is frequently proposed. However, this assumption is not entirely supportable. The FAIR Data Principles are useful guidelines in determining the structural properties of the data from a technical point of view since they do not indicate anything about the substantial accuracy or correctness of the data.

Why is it crucial to keep the two types of challenges separated when proceeding with the investigation? The *ratio* is related to the role of the institutions: introducing legislations that support Open Data, good data management and the adoption of the FAIR Data Principles do not concern the intrinsic quality of the data conveyed. In order to intervene (or to evaluate whether or not to intervene) in this sphere it is necessary to identify different strategies or types of actions.

[42] The Article 38, Regulation EU 2021/695. On the accessibility by default see, also, the following Section 4.2.3.2.
[43] This level is further investigated in Section 4.3 below.

The need to emphasise this distinction stems from a widespread lack of clarity. As an illustration, consider the management of FAIR Data Principles within the EOSC ecosystem. Whilst an eager work is conducted in order to implement the FAIRness within the EOSC[44], there is a lack of accuracy regarding the distinction between ethical and legal issues. Although some specialists are aware that "[…] many of the challenges are more on the social and policy side than on the technical side"[45], ambiguity persists. The implementation of FAIR Data is essentially delegated to two working groups: the technical issues related to FAIR Data are handled by the "Architecture Working Group"; the remaining problems are handled by the "FAIR Data Working Group". The "FAIR Data Working Group" should deal with all non-technical issues, which are summarily referred to as "cultural issues"[46]. However, under the label of "cultural issues" fall very different problematic matters: first of all, some of them belong to the purely legal sphere and others to the ethical sphere.

Keeping in mind this preliminary and essential distinction, next section explores the first type of issues: the legal challenges related to research data within the Open Science scenario.

4.2 Legal Issues of the Open Research Data

A wide range of legal issues emerge in relation to research data, their management and especially their reuse. This section starts by investigating the limits of the vague notion of legal interoperability,

---

[44] In the annual symposium on the development of the EOSC (held in 2021 from 15 to 19 June, online, available here: https://www.eoscsecretariat.eu/eosc-symposium-2021) a major effort in this direction clearly emerged. See the final report: Veronica Bertacchini, *et al.*, *EOSC Symposium Report*, 2021, Zenodo, doi:10.5281/zenodo.5176089.

[45] As stated, at EOSC Symposium 2021, by Sarah Jones, the new EOSC Engagement Manager at GÉANT, see: Veronica Bertacchini, *et al.*, *EOSC Symposium Report, op. cit.,* p. 13.

[46] "The FAIR & Architecture WGs operate in close alignment. The former addresses cultural aspects such as semantic and legal interoperability, certification and community data standards, while the latter focusses on the related technical specifications that address FAIR requirements." in https://www.eoscsecretariat.eu/working-groups/fair-working-group.

which seems to be the focus of the few studies on the Open Science legal issues conducted by European institutions (Section 4.2.1). Starting from the gaps provided by the concept of legal interoperability, attention is then drawn to the main legal framework on data sharing and reuse in Europe, i.e., the European Open Data framework (Section 4.2.2). After a general overview of this framework, the analysis proceeds from the overall picture to the specifics, shedding light on the Article 10 of the Open Data Directive (ODD), specifically addressed to research data (Section 4.2.3). The critical interpretation of this provision leads to the analysis of the hurdles of the openness of research data, in the EU Open Data legal framework (Section 4.2.4), which are: (i) the difficulties in the transposition of the ODD in the Member States, with the consequent risk of fragmentation of the discipline (Section 4.2.4.1); (ii) the limits of the enforceability of the openness by default (Section 4.2.4.2); as well as (iii) the still unresolved knots in terms of licences through which the effective sharing and reuse of research data can be achieved (Section 4.2.4.3).

4.2.1 Redefining Legal Interoperability: A Concept of Coordination and Harmonisation

The documentation related to the implementation of the EOSC has, so far, always referred to legal issues in very general and vague terms[47]. Chiefly, from the outset of the project, the soft law policies and strategies have often referred to the concept of legal interoperability. The notion of legal interoperability is far from being well-established, nor can it be found in any legal system. Yet, this notion is accompanying the development of the major European Open Science projects, first and foremost the EOSC: for this reason, further consideration is needed.

[47] Except for a very recent report, which provides interesting recommendations on the legal challenges of the EOSC, see: Caterina Sganga, *et al.*, "Recommendations for legal and policy harmonization of open and FAIR science in the EU", *EOSC-Pillar Project* (2022): 1-8, doi: 10.5281/zenodo.6327691.

The concept of legal interoperability *de facto* transposes a typically technical concept, that of interoperability, from the technical domain to the law domain.

From a technical point of view, the definition of interoperability is provided by the International Organization for Standardization, ISO, as "[…] the ability of two or more systems or components to exchange information and to use the information that has been exchanged"[48]. Similarly, the National Institute of Standards and Technologies, NIST, in its publication on Big Data, defines interoperability by adding an essential factor, namely the conditions that make exchange possible: "[…] the ability for tools to work together in execution, communication, and data exchange under specific conditions"[49].

The concept of *legal* interoperability, however, seems to be different. This notion, applied to data sharing and reuse projects, has been introduced already since 2016, by the organisation CODATA-RDA[50], defined as follows[51]:

> Legal interoperability occurs among two or more datasets when: the legal use conditions are clearly and readily determinable for each of the datasets, typically through automated means; the legal use conditions imposed on each dataset allow creation and use of

[48] ISO/TS 27790:2009 in PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica.* (Trento: Collana della facoltà di Giurisprudenza dell'Università di Trento, 2021), pp. 233-234. The Author underlines how this technical definition of interoperability has also been embraced by the European lawmaker in two directives: Directive 2009/24/EC on the legal protection of computer programs, i.e. Computer Programs Directive, in Recital 12; Directive 2007/2/EC, establishing an Infrastructure for Spatial Information in the European Community, so-called INSPIRE, in the Article 3(7).

[49] WO L. CHANG, NANCY GRADY, "NIST big data interoperability framework: Volume 1, big data definitions." *National Institute of Standards and Technology*, (2015): 7, doi:10.6028/NIST.SP.1500-1r2.

[50] RDA stands for Research Data Alliance, an organisation whose role has already been outlined in Section 2.4.3.3, referring to Open Science actors. CODATA is the Committee on Data of the International Science Council (ISC), see: https://codata.org. The joint effort of RDA and CODATA resulted in an RDA-CODATA interest group, which develops studies on certain aspects related to the RDA community, i.e., the broader research community. See: PAUL UHLIR, GAIL CLEMENT, "Legal Interoperability of Research Data: Principles and Implementation Guideline." *RDA-CODATA Legal Interoperability Interest Group* (2016).

[51] PAUL UHLIR, GAIL CLEMENT, "*Legal Interoperability of Research Data", op. cit.*, p. 36. The document explicitly mentions this study: CATHERINE DOLDIRINA, *et al.*, "Legal approaches for open access to research data." *LawArXiv*, (2018).

combined or derivative products; and users may legally access and use each dataset without seeking authorization from data rights holders on a case-by-case basis, assuming that the accumulated conditions of use for each and all of the datasets are met.

Later, this concept has been adopted in the document "Turning FAIR into reality"[52], as well as in the 2019 EOSC Strategic Implementation plan document[53], and finally it also appears in the SRIA, the living document following the implementation of the EOSC, version 1.0 of which was released in 2021[54]. This recurring mention to the "legal interoperability" would seem to suggest that this concept is part of the approach adopted by the European Open Science policies. However, the notion is never defined, except through a reference to the aforementioned RDA-CODATA document.

It should be noted that, in December 2020, a study entitled "Legal interoperability and FAIR Data Principles" was released[55]. This study, commissioned by the FAIR Data Working Group (i.e., one of the aforementioned EOSC Working Groups) and carried out by the legal consultancy firm X-Officio[56], provided several recommendations about "[…] the key issues in legal interoperability in connection with the implementation of the FAIR Principles within the context of the EOSC"[57]. In this report, legal interoperability is defined as "[…] the ability to combine datasets from multiple sources without conflicts among the restrictions that each dataset carries (i.e., support of one

[52] Sandra Collins, *et al.*, *"Turning FAIR into reality", op. cit.*, p. 23, which specifically refers to the RDA-CODATA study.
[53] Sarah Jones, Jean-François Abramatic, "European Open Science Cloud (EOSC) Strategic Implementation Plan." *Publications Office of the European Union* (2019): 17, doi: 10.2777/202370.
[54] SRIA stands for "Strategic Research and Innovation Agenda", EOSC Executive Board, February 2021, https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf, p. 13, 171; on SRIA, see: Section 2.1.2 and Section 2.2.3.
[55] Ohad Graber-Soudry, *et al.,* "Legal Interoperability and the FAIR Data Principles" (1.0). *Zenodo*, (2021), doi:10.5281/zenodo.4471312.
[56] https://www.xofficio.eu/about.
[57] Ohad Graber-Soudry, *et al.*, *"Legal Interoperability and the FAIR Data Principles", op. cit.*, p. 5.

restriction inherently negates support of another)"[58], again following the direction pointed by the RDA-CODATA document.

By broadening the analysis, going beyond the Open Science domain, the notion of legal interoperability is also found in the field of the Internet governance. For instance, a 2013 study proposes legal interoperability as a means to overcome regulatory fragmentation, resulting from multiple different jurisdictions as well as multiple legislative disciplines[59]. In this context, however, a different definition is proposed, identifying interoperability as a "normative tool": "Legal interoperability addresses the process of making legal rules cooperate across jurisdictions, on different subsidiary levels within a single state or between two or more states"[60]. This perspective seems to evoke the more traditional concept of harmonisation, typical of the international and European law.

Another field, where the use of the expression legal interoperability emerges, is the management of Open Government Data. In this strand of research, legal interoperability is defined as a prerequisite for technical interoperability: "Legal interoperability could be defined as the possibility of legally mixing data coming from different sources (including governmental data, data generated by online communities and data held by private parties). Legal interoperability is similar to technical interoperability, since it is a prerequisite for mixing data and create new knowledge or services"[61].

The references to these different interpretations of the notion of legal interoperability, proposed in diverse fields, clarifies the vagueness of

[58] OHAD GRABER-SOUDRY, *et al.*, *"Legal Interoperability and the FAIR Data Principles", op. cit.*, p. 19. In turn, it draws on the study: CATHERINE DOLDIRINA, *et al.*, *"Legal approaches for open access to research data", op. cit.*, mentioned earlier and linked with RDA-CODATA.
[59] ROLF H. WEBER, "Legal interoperability as a tool for combatting fragmentation." *Global Commission on Internet Governance Paper Series n. 4,* (2014): p. 5.
[60] ROLF H. WEBER, *"Legal interoperability as a tool", op. cit.*, p. 5.
[61] FEDERICO MORANDO, "Legal Interoperability: Making Open (Government) Data Compatible with Businesses and Communities.", *Italian Journal of Library and Information Science*, 4.1, (2013): 452, doi:10.4403/jlis.it-5461.

the concept. The major limitation of its current formulation in the context of Open Science, in fact, is that the lack of a precise connotation leads to minimising the legal problems that characterise the reuse of research data. In other words, the legal challenges related to sharing and re-using research data are often reduced to the formula "legal interoperability": it evokes a number of issues, mixing technical and legal nature, without providing the overall insight.

If the intention of the European institutions is indeed to use the term legal interoperability, it should be considered, at least, as the result of a dual process: on the one hand, a coordination process; on the other, a harmonisation process.

The co-ordination process aims to enable the legal reuse of data, taking into account the multiple disciplines involved (e.g., copyright, database regulation, data protection, etc.). According to Paolo Guarda, legal interoperability, in terms of coordination, should be "[…] the ability to combine datasets from multiple sources without conflicting restrictions that each of these may have"[62].

The harmonisation process, on the other hand, relates to the barriers resulting from the possible multiple jurisdictions involved in the sharing and reuse of research data. This aspect is more akin to the interpretation of the concept of legal interoperability adopted in Internet governance field of study, mentioned above. Moreover, the EOSC aims to present itself as an Internet of FAIR Data and Services[63]; as a result, it is to be expected that, from a legal point of view, the EOSC governance should face the same issues as the Internet governance as a whole.

The several mentions of the concept of legal interoperability in the different documents concerning the European Open Science projects suggested an increased relevance of this formula[64].

---

[62] PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica, op. cit.,* p. 234. [Translation from the Italian original text].
[63] See: Section 4.1.2.
[64] This is even more evident considering the consequences of legal interoperability. In fact, the legal

The interpretation of legal interoperability as a process of coordination and harmonisation, however, does not lead to the same conclusions as the study conducted by the X-Officio, with regard to one specific aspect: this study interprets the role of some European legal disciplines as barriers to Open Science, and other legal disciplines as enablers[65]. Specifically, the study indicates as barriers the discipline of data protection and that of copyright, whereas the *corpus* of provisions in the field of Open Data as enablers. It has already been argued above why the view of the European discipline of data protection as a barrier to scientific research is not supported[66]. Beyond this specific aspect, in the context of Open Science, it seems fair to reject the view of some legislative frameworks as barriers and others as facilitators, for two main reasons.

First, the vision of the law as a barrier to scientific research should be rejected because it implies that certain legal texts are opposed to the promotion of scientific research. The right to science, as seen above[67], is a human right that is protected on several levels (i.e., international, European, and constitutional). Therefore, even when a discipline aims to protect another fundamental right, the European lawmaker is always required to strike a balance between the rights, so that the protection of one does not completely override the other. From a legal point of view, this issue is a matter of balance between human and fundamental rights, which can – and must – coexist, not with mutually opposing requirements.

The second reason, then, is more concrete. Interpreting the Open Data discipline in Europe as an enabler provides only a partial view of the context: although the goal of the Open Data policies in Europe is to

---

interoperability seems to represent the condition allowing the creation of derivative works, based on the sharing and reuse of datasets. On this aspect, see: PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica, op. cit.,* p. 234; and OHAD GRABER-SOUDRY, *et al.*, *"Legal Interoperability and the FAIR Data Principles", op. cit.*, p. 19.

[65] OHAD GRABER-SOUDRY, *et al.*, *"Legal Interoperability and the FAIR Data Principles", op. cit.*, p. 8.

[66] On this aspect, see: Section 2.3.2. This topic is also further explored in Chapter 5.

[67] On the fundamental and human rights framework of Open Science, see: Section 3.3.

foster data sharing and reuse, in accordance with the Open Scientific Research Process, there are still several challenges posed by the Open Data disciplines that scientific research must face. Similarly, some scholars argued that "[…] the EU regulatory landscape is highly complex when it comes to data reuse. Claiming that as a general proposition the EU law should be labelled as a barrier or as an enabler does not hold much water"[68].

Therefore, next section aims to discuss the legal framework of the European Open Data disciplines.

4.2.2 EU Open Data Framework

The European legal framework on Open Data is heterogeneous and not all concerns the field of scientific research. The Open Data domain, in fact, concerns the public sector data. It is therefore necessary to take into account the different legislative acts involved and their different impact on the scientific research domain.

Chiefly, a terminological threshold should be clarified. Data produced, generated, processed, and stored by the public sector have traditionally been referred to as "public sector information", blurring the difference between the concepts of data and information[69]. In general, it should be considered that this expression has traditionally been used to refer to all materials, produced and managed by public entities, available for reuse and sharing: data that may be weather report, data deriving from statistics commissioned or carried out by public bodies, environmental or geographical data, etc. More generally, the reference is to the concept of public sector "documents", as defined in the Article 2(3) of Directive 2003/98/EC, the so-called PSI Directive: "document means: (a) any content whatever its medium (written on

[68] HELENA URSIC, BART CUSTERS, "Legal Barriers and Enablers to Big Data Reuse." *European Data Protection Law Review*, 2 (2016): p. 221.

[69] On this aspect see: UGO PAGALLO, ELEONORA BASSI, "Open Data Protection: Challenges, Perspectives, and Tools for the Reuse of PSI", in MIREILLE HILDEBRANDT, *et al.*, (eds), *Digital Enlightenment Yearbook 2013* (Amsterdam: IOS Press, 2013): p. 180, doi:10.3233/978-1-61499-295-0-179.

paper or stored in electronic form or as a sound, visual or audio-visual recording) (b) any part of such content"[70].

Consider that the European initiatives on Open Data and sharing and reuse of public authority material (including both hard and soft law), starting from the early 2000s and following the general trend of European policy development[71], referred to the concept of "information" or "document", rather than "data"[72].

In analysing the current European legal framework of Open Data, here the intention is to refer only to hard law interventions, excluding from the analysis many soft law interventions, represented by strategies, guidelines, and action plans.

There are essentially two disciplines to take into account: the INSPIRE Directive[73]; the Open Data Directive (ODD)[74].

---

[70] Directive 2003/98/EC, o*n the re-use of public sector information*, OJ L 175, 27.6.2013, p. 1–8, ELI: http://data.europa.eu/eli/dir/2013/37/oj. The Directive is no longer in force as of 16 July 2021.

[71] See: MASSIMO DURANTE, "Potere computazionale e questioni giuridiche. Dalle informazioni ai dati." in MASSIMO DURANTE, UGO PAGALLO (eds.), *Il governo delle nuove tecnologie tra diritto, economia e società*. (Milano-Udine: Mimesis, 2022), pp. 59-80, where it is analysed how the focus of attention and the European strategies in the early 2000s was the concept of "information"; while now the reference to information is, almost everywhere, replaced by the reference to "data".

[72] Consider that the proposal for the Directive on the sharing and reuse of public sector documents is from 2002: Proposal for a Directive of the European Parliament and of the Council, *on the re-use and commercial exploitation of public sector documents*, COM/2002/0207, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52002PC0207.

[73] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007, *establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*, OJ L 108, 25.4.2007, ELI: http://data.europa.eu/eli/dir/2007/2/oj.

[74] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019, *on open data and the re-use of public sector information*, OJ L 172, 26.6.2019, 56–83, ELI: http://data.europa.eu/eli/dir/2019/1024/oj. The scenario would also include the Environmental Information Directive: Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003, *on public access to environmental information and repealing Council Directive 90/313/EEC*, OJ L 108, 25.4.2007, p. 1–14, ELI: http://data.europa.eu/eli/dir/2003/4/oj. It covers environmental information, understood as: "'Environmental information' shall mean any information in written, visual, aural, electronic or any other material form", on several aspects, such as: "(a) the state of the elements of the environment, such as air and atmosphere, water, soil, land, landscape and natural sites […]; (b) factors, such as substances, energy, noise, radiation or waste, […]; (c) measures (including administrative measures), such as policies, legislation, plans, […]; (d) reports on the implementation of environmental legislation; (f) the state of human health and safety, including the contamination of the food chain", *ex* the Article 2(1). As claimed in the X-Officio study, already mentioned: "[…] applies to public institutions at all levels and to information held by other entities on

The objective of the INSPIRE Directive is the establishment of "the Infrastructure for Spatial Information in the European Community (hereinafter referred to as Inspire), for the purposes of Community environmental policies and policies or activities which may have an impact on the environment"[75]. In other words, the aim is to identify a regulatory framework for an infrastructure which aims to share "[…] metadata, spatial data sets and spatial data services; network services and technologies"[76].

Throughout this directive, there are no explicit references to scientific research. However, the implementation of INSPIRE has been presented by some scholars as the first step towards integrated research, capable of addressing the "[…] key scientific challenges of humanity in the 21st century"[77]. It can be argued that the INSPIRE Directive was an anticipation of the new European Data Spaces, the development of which has been established in the EU Data Strategy of 2020[78].

More relevant to this dissertation is, instead, the Open Data Directive (ODD), essentially for two reasons. The first reason is connected to the objective of the directive: the ODD intends to bring order to the scenario, repealing the previous directive on the reuse of Public Sector Information, the so-called PSI Directive[79], and, as a result, adapting the legal framework to the current context.

behalf of public institutions, as well as institutions performing functions on behalf of public institutions" but "there is no obligation for public authorities to make information publicly available without a request, e.g., in an open database. Access may be subject to reasonable fees.", in OHAD GRABER-SOUDRY, *et al.*, *"Legal Interoperability and the FAIR Data Principles"*, *op. cit.*, p. 67.

[75] The Article 1, Directive INSPIRE, 2007/2/EC.

[76] The Article 3(1), Directive INSPIRE, 2007/2/EC.

[77] LORENZINO VACCARI, *et al.* "Integrative research: the EuroGEOSS experience." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5.6 (2012): p. 1603, doi: 10.1109/JSTARS.2012.2190382.

[78] ALEXANDER KOTSEV, *et al.* "From spatial data infrastructures to data spaces—A technological perspective on the evolution of European SDIs." *ISPRS International Journal of Geo-Information,* 9.3 (2020): p. 176, doi: 10.3390/ijgi9030176.

[79] The PSI Directive, 2003/98/EC, has already undergone a major revision in 2013, by the Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive

The second reason is related to the space specifically dedicated to research data. Starting from the Recitals, scientific research (financed by public grants) is clearly included as a relevant sector within the scope of the reuse of public data[80]. This is the most important difference to the repealed PSI Directive. The Article 1(2)*e* of the PSI Directive expressly excluded the following from its scope of application "[…] documents held by educational and research establishments, such as schools, universities, archives, libraries and research facilities including, where relevant, organisations established for the transfer of research results"[81].

By contrast, in the ODD, data from publicly funded research projects are expressly included in the scope of application and defined in the Article 2(9) as: "documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results". This change of direction of the European policies on Open Data was the result of the public consultation on the revision of the legal framework of Open Data, preceding the ODD[82]. In addition, the ODD – with regard to research

2003/98/EC, o*n the re-use of public sector information*, OJ L 175, 27.6.2013, p. 1–8, ELI: http://data.europa.eu/eli/dir/2013/37/oj.

[80] In Section 2.3.1, it has already been underlined that: "Recitals 27 and 28 of the ODD underline the importance of exploiting the huge amount of data generated by scientific research by defining policies, at national level, so that "[...] certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidized by public funding or co-funded by public and private-sector entities".

[81] Nevertheless, on the – albeit – limited impact of the PSI Directive and its 2013 adaptation on the sharing and reuse of research data and university libraries, see: ANDREAS WIEBE, NILS DIETRICH, *Open Data Protection-Study on legal barriers to open data sharing-Data Protection and PSI* (Göttingen: Universitätsverlag Göttingen, 2017), pp. 211-258.

[82] European Commission, DG Communications Networks, Content and Technology, *Public consultation on the review of the directive on the re-use of Public Sector Information (PSI Directive)*, from 19 September 2017 to 12 December 2017, https://ec.europa.eu/info/consultations/public-consultation-review-directive-re-use-public-sector-information-psi-directive_en; Specifically, the consultation concerned Directive 2013/37/EC. In PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica, op. cit.,* pp. 187-188, three main reasons for excluding research data from the former and repealed PSI

data – is also aligned with the previous European Commission Recommendation "on access to and preservation of scientific information", issued in 2018[83]. This Recommendation, in fact, specifically mandated Member States to implement Open Access policies, not only with regard to research results (i.e., publications), but also with regard to data[84]. Some of the suggestions of this Recommendation became mandatory with the introduction of the ODD.

The most significant aspect is the Article 10 of the ODD, which is specifically entitled "Research data" and which establishes considerable obligations on Member States. For this reason, next section investigates the Article 10 of the ODD.

4.2.3 The Art. 10 of the ODD: A Critical Interpretation

The Article 10 of the ODD is entitled "Research data" and, as specified in Recital 27, it addresses "[…] statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images meta-data, specifications and other digital objects". The Article 10 of the ODD has two paragraphs, which are highly detailed.

The first paragraph states:

> Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible

---

are underlined: "[…] the high burden of clarifying the legal status of research data in order to make them reusable under the PSI Directive would have outweighed the benefits; the existence of a dynamic and well-established system for the dissemination and exploitation of research findings and results; the peculiar character of the debate on Open Access, which is in some ways conceptually separate from the debate on PSI." [Translation from the Italian original text].

[83] Commission Recommendation (EU) 2018/790, *On access to and preservation of scientific information*, of 25 April 2018, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018H0790. Some considerations on this Recommendation have already been expressed in: Section 2.2.4; Section 2.4.3; and also, Section 2.3.1.

[84] On the missed expectations in relation to the implementation of this Recommendation see: ROBERTO CASO, *La rivoluzione incompiuta: La scienza aperta tra diritto d'autore e proprietà intellettuale*. (Milano: Ledizioni, 2019), p. 39, 173.

with the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'. Those open access policies shall be addressed to research performing organisations and research funding organisations.

Some considerations may be developed from the wording of the legislative provision. In particular: (i) on the actors involved; (ii) on the concept of Open Access applied to data; (iii) on the reference to the FAIR Data Principles; as well as (iv) on the limits to the openness of data.

(i) Actors involved. The first aspect to take into account concerns the actors involved: i.e., Member States and bodies that fund and conduct public scientific research.

The scope assigned to the national level is crucial: Member States are mandated to adopt policies for sharing and reuse of research data. This means that the transposition of the directive by the Member States is fundamental, assuming that there will be indications on national initiatives to make open and FAIR research data.

The recipients of these policies, conversely, are "research performing organisations" and "research funding organisations". In other words, these policies are addressed to the bodies that conduct and fund public scientific research. Recital 28 then specifies that these recipients can be indifferently structured as "sector bodies or public undertakings"[85], subject to the condition that the provisions apply only "[…] in their capacity as research performing organisations and to their research data"[86]. This effectively extends the requirements for sharing and

[85] The definition of "public undertakings" is provided in the Article 2(3): "any undertaking active in the areas set out in point (b) of Article 1(1) over which the public sector bodies may exercise directly or indirectly a dominant influence by virtue of their ownership of it, their financial participation therein, or the rules which govern it". The areas mentioned, in accordance with Article 1(1)*b*, are water, energy, public transport – road, railway and maritime – and postal services.
[86] Recital 28, ODD.

reuse of research data to all cases of publicly funded research, regardless of whether it is conducted wholly or partially by public bodies[87].

(ii) Open Access applied to research data. Second, consider that the European lawmaker makes a general mention of "open access policies". As described in Section 3.1.2[88], Open Access generally refers to free access to scientific publications, understood as articles in scientific journals, rather than access to data or datasets. In the Article 10, however, the expression Open Access is related to free access to research data, since the ODD itself excludes scientific publications from its scope. The European lawmaker, in Recital 27, stresses that "[R]esearch data is different from scientific articles reporting and commenting on findings resulting from their scientific research". Moreover, the ODD interprets the Open Access "[…] as the practice of providing online access to research outputs free of charge for the end user and without restrictions on use and reuse beyond the possibility to require acknowledgement of authorship"[89]. However, the reference to Open Access policies, in this context, must only be related to research data and this is clarified by Recital 28, where it explicitly states that "[D]ocuments other than research data should continue to be exempt from the scope of this Directive". Yet, an explicit connection with publications is established in the same Recital 28, where it is stressed that the application of the ODD provisions may be – at the discretion of Member States – extended to research data made "publicly available" by means other than repository, such as "[…] open access publications, as an attached file to an article, a data paper or a paper in a data journal"[90].

[87] In this regard, Recital 28 ODD states that "[…] certain obligations stemming from this Directive should be extended to research data resulting from scientific research activities subsidised by public funding or co-funded by public and private-sector entities".
[88] The issue was also discussed in Section 3.2, where the dynamic interpretation of Open Science, in terms of the Open Scientific Research Process, was illustrated.
[89] Recital 27, ODD.
[90] Recital 28, ODD.

(iii) FAIR Data Principles. A third aspect that should be considered is that, for the first time, the FAIR Data Principles are mentioned within a hard law legislative act: the ODD marks the acknowledgement by the European lawmaker of these guidelines for data reuse, explored in Section 4.1.2. The European institutions go further, establishing openness "by default", i.e., as a standard option[91]. An opt-out approach, thus, seems to be adopted: data closure becomes an exception in a scenario where openness is the rule. Recital 27 then specifies that the FAIR Data Principles are an aspect related to data management, rather than to Open Access[92]. Here, the formal distinction between FAIR Data and Open Data becomes clear: the FAIR Data Principles aim to achieve good data management, making research data *potentially* findable, accessible, interoperable and reusable. While the openness of research data consists in making them *effectively* accessible online, without restrictions on possible reuse[93]. Going a step further what has been stated by the European lawmaker, as well as pointing out what will be said later in relation to the ethical challenges of open research data[94], it seems fair to assume that data management based on the FAIR Principles is a prerequisite and a condition for the openness of research data. In other words, although FAIRness and openness are two distinct concepts, having FAIR data is a necessary but not sufficient condition to have open research data. This aspect is further explored and clarified in Section 4.3.

(iv) Limits to openness. A final consideration regarding the first paragraph of the Article 10 of the ODD is related to the restrictions of such openness. As stressed on several occasions in the analysis of European soft law strategies on Open Science, investigated in Section

[91] This EU approach has been confirmed in the Regulation EU 2021/695, see: Section 4.1.3.
[92] "Beside open access, commendable efforts are being made to ensure that data management planning becomes a standard scientific practice and to support the dissemination of research data that are findable, accessible, interoperable and re-usable (the FAIR principle)", Recital 27, ODD.
[93] As clarified by Recital 27, ODD.
[94] See: Section 4.3.

2.2, the openness of research is never blind, but rather subject to the principle "as open as possible, as closed as necessary". This formula has been officially adopted by the European institutions, who also go a step further to identify what the reasons for closure might be. The interests to be balanced against the need for openness are listed in the Article 10 and specified in Recital 28: (1) intellectual property rights of third parties, involved in different ways in the research process; (2) privacy and personal data protection; (3) confidentiality; (4) national security; and (5) legitimate commercial interests, "such as trade secrets"[95]. In this manner, the European lawmaker is setting the boundaries within which national institutions can operate in providing for – exceptional – closure options. The list seems closed and fixed, as it does not identify residual clauses or clauses open to the discretion of either Member States or researchers. However, these interests for closing the research data are so broad that they can cover a very wide range of situations. In addition, it is not clear which should be meant by "confidentiality" from a legal point of view, considering that the concept of confidentiality is often related to other interests, already included in the list of exceptions (e.g., privacy or data protection). The definition of data confidentiality can be derived from the Communication on the protection of confidential information by national courts in proceedings for the private enforcement of the EU competition law[96]. Although this communication deals with "confidential information in civil proceedings before national courts"[97], it is suitable insofar as it incorporates a definition of confidential information, as determined by the jurisprudence of the European Courts[98]: "The EU courts qualify as

---

[95] Recital 28, ODD. These limits are also reaffirmed in the Article 38 of the Regulation EU 2021/695.
[96] European Commission, *on the protection of confidential information by national courts in proceedings for the private enforcement of EU competition law*, C/2020/4829, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020XC0722(01).
[97] C/2020/4829, p. 1.
[98] The European Commission's references are to the following cases: "Judgment of the General Court of 30 May 2006, *Bank Austria* v *Commission,* CaseT-198/03, EU:T:2006:136, paragraph71; judgment of the General Court of 8 November 2011, *Idromacchine* v *Commission*, Case T-88/09, EU:T:2011:641,

confidential information that meets the following cumulative conditions: (i) it is known only to a limited number of persons; and (ii) its disclosure is liable to cause serious harm to the person who provided it or to third parties; and (iii) the interests liable to be harmed by the disclosure of confidential information are, objectively, worthy of protection"[99]. As a result, the notion of confidentiality referred to in the Article 10 of the ODD could legitimately be linked to that which emerges from the European case law, with reference to "confidential information".

The second paragraph of the Article 10 of the ODD, on the other hand, states:

> Without prejudice to point (c) of Article 1(2), research data shall be re-usable for commercial or non-commercial purposes in accordance with Chapters III and IV, insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository. In that context, legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights shall be taken into account.

The second paragraph deals with data, from public scientific research, already made publicly available through the archiving in institutional repositories (i.e., repositories managed – from an infrastructural point of view – directly by the university or the research center involved) or repositories relevant to the scientific domain (e.g., GSA – Genom Sequence Archieve for genomics; or EARTHCHEM, for environmental sciences, etc.), by researchers, or by organisations that conducted or funded the research.

---

paragraph 45; judgment of the General Court of 28 January 2015, *Akzo Nobel and Others* v *Commission*, Case T-345/12, EU:T:2015:50, paragraph 65; and judgment of the Court of Justice of 14 March 2017, *Evonik Degussa* v *Commission*, Case C-162/15 P, EU:C:2017:205, paragraph 107", in C/2020/4829, section n. 20.

[99] C/2020/4829, section n. 20.

The *ratio* at the basis of this provision on data already made publicly available is given in Recital 65, which states that "[L]ibraries, including university libraries, museums and archives hold a significant amount of valuable public sector information resources, in particular since digitisation projects have multiplied the amount of digital public domain material. Those cultural heritage collections and related metadata are a potential base for digital content products and services and have a huge potential for innovative reuse in sectors such as learning and tourism"[100].

Furthermore, it is underlined that such open data can be reused for both commercial and non-commercial purposes, depending on the fulfilment of a triple condition: the openness must not harm third parties' intellectual property rights; the reuse of such data must respect the conditions for reuse, as regulated by Chapter III of the ODD; and it must be non-discriminatory, in accordance with Chapter IV of the ODD.

The purpose of this second paragraph of the Article 10 is spelled out in Recital 28, which explains that, as regards data that already exist, the requirements set out should only be applicable to research data already made publicly available, "[I]n order to avoid any administrative burden".

The complexity of the scenario is, however, expressed by the final provision, which requires to take into account opposing interests such as commercial or pre-existing IP rights, and those of knowledge transfer.

The ODD, on the one hand, strives to emphasise that the openness and sharing are never blind, considering also the commercial interests

[100] On this aspect, it is worth to mention that the EU lawmaker has adopted a very different approach in another piece of legislation, identifying a clear distinction between library and university library, i.e., in the Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019, *on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*, OJ L 130, 17.5.2019, 92–125, ELI: http://data.europa.eu/eli/dir/2019/790/oj. Here, the Article 2 consider university library as "research organisation" (Article 2(1) of Dir. EU 2019/790) and by contrast library as "cultural heritage institution" (Article 2(3) of Dir. EU 2019/790).

and IP rights of any third parties involved. On the other hand, however, it insists on knowledge sharing, which is defined in terms of "right to knowledge", in Recital 43: "Making public all generally available documents held by the public sector — concerning not only the political process but also the legal and administrative process — is a fundamental instrument for extending the right to knowledge, which is a basic principle of democracy. That objective is applicable to institutions at every level, be it local, national or international". In doing so, an implicit reference is drawn to the so-called fifth European freedom, the circulation of knowledge[101], which finds one of its multiple dimensions precisely in sharing and reuse of research data.

The examination of the Article 10 of the ODD has anticipated some open issues regarding the application of the Open Data framework in the context of scientific research. Next section explores these issues in more detail, showing that considering the EU Open Data legal framework as "enabler" is a shortcoming: it only underlines one sided perspective, i.e., the intentions of the European institutions. It is therefore necessary, instead, to shift attention to the downside, i.e., the still unresolved knots.

4.2.4 Hurdles to the Openness of Research Data

The European Open Data legal framework is not devoid of challenges, nor can the legal issues related to data sharing and reuse be reduced to the concept of legal interoperability. For these reasons, it is now worth examining the most problematic legal aspects representing a barrier to the openness of research data.

First of all, attention should be drawn to the national transposition of the ODD and to the limits related to it, *in primis* the fragmentation of the discipline (Section 4.2.4.1). Secondly, some remarks are illustrated about the enforceability of the provision of openness "by default" of research data, as established by the Article 10 of the ODD

[101] See: Section 2.2.1.

(Section 4.2.4.2). Finally, a last observation is raised on the issue of licenses to release research data, instrumental to allow their sharing and reuse (Section 4.2.4.3).

4.2.4.1 National Transpositions

The ODD embodies a twofold achievement: on the one hand, it has brought order to the system of sources concerning Open Data in Europe; and on the other hand, it has widened the scope of application, by including research data[102]. However, one of the constraints identified is the legislative act chosen to embody this European data-sharing strategy, namely the directive. This legal act requires the transposition into national law by the Member States, leaving "[…] to national authorities the choice of form and methods", according to the Article 288 of the TFEU. In particular, the Article 17 of the ODD states that "Member States shall bring into force the laws, regulations and administrative provisions necessary to comply with this Directive by 17 July 2021".

On several occasions, the academic community has questioned the choice of the European institutions to opt for the directive as the suitable legal instrument. This occurred, for instance, in the field of Data Protection, with Directive 95/46/EC, before its repeal by the GDPR[103], or, similarly, in the field of consumer law[104]. One of the weaknesses traditionally raised is the risk of fragmentation resulting

[102] The specificities that had traditionally led to the exclusion of research data from the scope of the disciplines of reuse of public data have been already mentioned, referring to PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica, op. cit.,* pp. 187-188. Although these reasons are totally endorsed, the ODD introduced, for the first time, an aspect of the Open Science vision – i.e., open access to data – within a hard law discipline. Although there are obvious limitations to this operation (which will be investigated in these sections) it was nonetheless an attempt to unhinge a years-old approach to data.

[103] Among others, see: PAUL DE HERT, VAGELIS PAPAKONSTANTINOU, "The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals." *Computer law & security review* 28.2 (2012): 130-142, doi: 10.1016/j.clsr.2012.01.011.

[104] Among others, see: COLETTE CUIJPERS, BERT-JAAP KOOPS, "How fragmentation in European law undermines consumer protection: the case of location-based services." *European Law Review* 33 (2008): 880-897.

from the implementation and reduced political impact of the European strategies conveyed through this piece of legislation.

By contrast, there are circumstances in which the directive is the most suitable act for the purpose in hand: it is a means by which to achieve harmonisation between national disciplines, without necessarily unifying the law, in the name of the "union in diversity" which became the motto of the European Union in the aftermath of the Second World War. In fact, the directive is characterised by flexibility: it enables national specificities and the particular needs of the different Member States to be guaranteed.

The key question, therefore, is whether the directive is the most effective means for meeting the needs and pursuing the objectives of scientific research.

On the one hand, knowledge is, by its very nature, without borders of any kind. This assumption underpins the framework of fundamental and human rights underlying the Open Science, both at international level, with the right to science, and at European level, with the identification of the fifth EU freedom of movement[105]. Providing for a national discipline of sharing and reuse of scientific research may impact the effectiveness of the initiative itself.

On the other hand, the reason why the European institutions adopted the directive, instead of other legal instruments, e.g., the regulation, is readily identifiable. The competence in the field of scientific research is essentially in charge of the Member States[106]: the European Union has a coordinating role in this matter, deriving from the Article 4(3) of the TFEU. Nevertheless, it should not be forgotten that the Article 167 of the TFEU stresses, among other aspects, that "[A]ction by the Union shall be aimed at encouraging cooperation between Member States and, if necessary, supporting and supplementing their action in the following areas: improvement of the

[105] See: Section 3.3.
[106] Based on the Articles 4(4) and 165 TFEU.

knowledge and dissemination of the culture and history of the European peoples".

Without prejudice to the framework of competences laid down in the EU Treaties, the sphere within which the EU can act appears to be quite flexible. So far, in the field of scientific research and knowledge sharing, a rather restrictive view of the Union's competence has always prevailed, in favour of a wide discretionary power of the Member States[107]. An emblematic example of this trend is the already mentioned European regulation on personal data. Despite the fact that it is ruled by a regulation that has unified the matter at European level, scientific research has nevertheless remained an area in which the Member States exercise considerable discretion[108].

The decision to devolve competence for scientific research mainly to the Member States, also in light of the leeway left by the EU treaties, seems, rather, to be the expression of a specific market-oriented approach: "[…] the underlying idea is that each Member State should bear the costs for its own nationals or, at most, for the European citizens who reside permanently on its territory […], since it is only that Member State which will benefit in the future from the higher training acquired, enjoying a 'gain' on the investment made"[109]. This vision, however, no longer corresponds to the current needs. First, because knowledge, scientific research, and the outcomes of each of its phases necessarily circulate within the European territory, and certainly beyond it, even easier thanks to new technologies. Second, an economic approach on a national basis to the European policies on

---

[107] In Section 2.1.1, it has been analysed how, however, this traditional view seems to be in slight reverse trend.

[108] Many academics have investigated the consequences of this choice of the European lawmaker. See, *ex multiis*: PAOLA AURUCCI, "Legal Issues in Regulating Observational Studies: The Impact of the GDPR on Italian Biomedical Research." *European Data Protection Law Review* 5(2) (2019): 197-208, doi: 10.21552/edpl/2019/2/9; and MARIA LUISA MANIS, "The processing of personal data in the context of scientific research. The new regime under the EU-GDPR." *BioLaw Journal-Rivista di BioDiritto* 11.3 (2017): 325-354. On this aspect, also see: Section 5.2.2.

[109] BARBARA GAGLIARDI, *La tutela amministrativa della libertà accademica.* (Milano: CEDAM, 2018), p. 121. [Translation from the Italian original text].

scientific research does not work very well insofar as the main funding entity for research is the European Union: the last framework programme of the European research funding, i.e., the Horizon Europe[110] allocated EUR 95.5 billion to scientific research, an upward trend of 30% compared to the first framework programme in 1994[111].

Earlier, the discretionary power that the Article 10(1) of the ODD endows to Member States in identifying Open Access policies to support the availability of research data has been discussed. As a result, the possibility of fragmentation related to the regulation of sharing and reuse of research data appears to be real. The possibility of having several national disciplines that do not necessarily agree with each other, may generate the risk of transforming the European Open Data disciplines from enablers into factual bottlenecks: "Most developed countries spend large amounts of public resources on research and related scientific facilities and instruments that generate massive amounts of data. Yet precious little of that investment is devoted to promoting the value of the resulting data by preserving and making them broadly available. The largely *ad hoc* approach to managing such data, however, is now beginning to be understood as inadequate to meet the exigencies of the national and international research enterprise"[112].

Consider the *ratio* of the ODD: settling the matter of Open Data, in order to foster, as much as possible, the reuse of data held by the public administrations and by the public sector, in order to extract new value from them. As a result, the choice of the directive as a legislative instrument is particularly suitable. In this way, the Member States have the opportunity to adapt their specific national needs, connected

---

[110] See: Section 2.1.1.2; Section 2.2.6; Section 3.2.1, in particular "(O_2) Actors"; and also, Section 5.1.2, about "Entities Funding Research".

[111] European Commission, DG Research and Innovation, *Horizon Europe: Budget*, 2021, p. 2 doi: 10.2777/714209. On this aspect, see: Section 2.2.6. Whereas, on the division of competences between the EU and the Member States, see: Section 5.2.2.

[112] PAUL F. UHLIR, PETER SCHRÖDER, "Open data for global science." *Data Science Journal* 6 (2007): OD36-OD53.

to the work of their public sector[113], to the general framework established by the European institutions. The peculiarities of public administrative systems strongly require national adaptations, given: the specificity of internal mechanisms, which have been rooted in national needs for decades (if not centuries); the fact that are strongly anchored in the domestic languages; and, also, the partially different organisations of the public bodies in each Member State.

However, the same does not apply to scientific research, which is based on sharing of knowledge and cooperation among researchers. The research domain is characterised by mechanisms that may differ from one area of research to another, but not from one Member State to another.

These significant differences lead to the conclusion that there is a need for greater coordination in the field of scientific research and, specifically, in the sharing and reuse of research data.

In light of this legislative scenario, there appear to be two ways forward. First, it is necessary to wait for the national transpositions of the ODD, in order to understand if and to what extent the Open Access policies for research data diverge in the different Member States[114]. Second, it is worth envisaging another type of coordination, beyond the national dimension: the action at the local level, of universities and

---

[113] Assuming that territorial boundaries apply in this sector but taking into account that "[I]n the information age, politics is beginning to be understood as the efficient and effective management and control of the information life cycle, which almost always exceeds the spatial constraints of nation state territories" in MASSIMO DURANTE, "An Informational Approach to Politics." in MASSIMO DURANTE (eds.) *Ethics, Law and the Politics of Information. The International Library of Ethics, Law and Technology*, vol. 18 (Dordrecht, Springer: 2017), doi: 10.1007/978-94-024-1150-8_9.

[114] It should be noted that the transposition of the previous PSI Directive had very different results in the different Member States. On this point, see: CRISTINA DOS SANTOS, ELEONORA BASSI, CECILE DE TERWANGNE, MANUEL SALMERON, POLONA TEPINA, "Policy Recommendation on Privacy and Personal Data Protection as Regards Re-Use of Public Sector Information (PSI)", *Masaryk University Journal of Law and Technology MUJLT*, 6(3), 2012, p. 8. Here, to monitor the current situation about the transposition and the implementation of the ODD in the different Member States: https://eur-lex.europa.eu/legal-content/EN/NIM/?uri=uriserv%3AOJ.L_.2019.172.01.0056.01.ENG. However, as previously argued, it is not sufficient a mere transposition. What need to be seen are the national open access action plans and strategies of each Member State, in light of the mandate of the ODD.

research centres. While this latter aspect is investigated later[115], now it is time to explore a further legal barrier to the openness of research data, related to the concept of "openness by default".

4.2.4.2 "Open by Default" Enforceability

The Article 10 of the ODD, paragraph 1, stipulates that publicly funded scientific research data should be "open by default": as analysed in Section 4.2.3, this means that openness is identified as the preset option. As a consequence, if the intention is to close the research data, a specific and exceptional choice in this sense will have to be made. In other words, the institutional or domain-based repository in which the data are stored will provide the human user the open option as predetermined, by default. It will be the human user, i.e., the researcher, who will have to specifically choose the option to close the data, if there are conflicting interests against openness: i.e., those reasons justifying closure, listed in the Article 10 of the ODD.

One of the reasons behind this legal requirement is the aforementioned "reproducibility crisis". As has already been examined[116], there are many reasons for the reproducibility crisis: one of them is the lack of availability of the data on the basis of which the research projects were developed[117]. A vicious circle is triggered: publications contain the results of research projects, but it is common that the data used in carrying out the research itself and needed to replicate the experiments are not available (note: not open, just available). In a recent study, Miyakawa states: "As an Editor-in-Chief of Molecular Brain, I have handled 180 manuscripts since early 2017

---

[115] See: Chapter 7, Conclusions.

[116] The issue of reproducibility of the results and processes of science, in the Open Science paradigm, has already been analysed in this Chapter 4, Section 4.1.1 (the reproducibility has been considered one of the features characterising the operational definition of research data, adopted in this dissertation); in Chapter 3, Section 3.2: specifically investigating the Observable O_3 Methods ($V_{3.2}$, Verifiability).

[117] On this aspect, see also: BERNARDO A. HUBERMAN, "Sociology of science: Big data deserve a big audience." *Nature*, 482 (2012): p. 308; and ALAWI A. ALSHEIKH-ALI, *et al.*, "Public availability of published research data in high-impact journals." *PloS one* 6.9 (2011): e24357.

and have made 41 editorial decisions categorized as 'Revise before review,' requesting that the authors provide raw data. Surprisingly, among those 41 manuscripts, 21 were withdrawn without providing raw data, indicating that requiring raw data drove away more than half of the manuscripts"[118]. A further relevant study, published in Nature in 2016, showed that, out of a sample of 1576 researchers surveyed, "[M]ore than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments"[119].

Accordingly, the "open by default" clause, set out in the Article 10 of the ODD, would seem to have a very significant impact on the scientific research scenario. As mentioned before, in general terms it does, considering that for the first time some of the key concepts of the Open Science paradigm have been conveyed in a piece of legislation.

In addition, this requirement of the Article 10 of the ODD is not only intended to be a way of addressing the issue of the lack of availability of research data, and thus indirectly help to overcome the reproducibility crisis, but it also represents something else. Considering the general framework of rights, in fact, the European lawmaker, in the Article 10, strikes a balance: on the one hand, there is the right to science, which in relation to research data is declined in

[118] TSUYOSHI MIYAKAWA, "No raw data, no science: another possible source of the reproducibility crisis." *Molecular Brain* 13, 24 (2020): 13-24, doi:10.1186/s13041-020-0552-2.

[119] MONYA BAKER, "1,500 scientists lift the lid on reproducibility." *Nature* 533, (2016): 452–454, doi: 10.1038/533452a. These concerns led to "A Manifesto for Reproducible Science", a contribution by various scientists, published by the journal Nature, which aims to draw attention to the necessary improvement practices that should be put in place to promote the integrity of science as much as possible: "Data from many fields suggests reproducibility is lower than is desirable; one analysis estimates that 85% of biomedical research efforts are wasted, while 90% of respondents to a recent survey in Nature agreed that there is a 'reproducibility crisis'. Whether 'crisis' is the appropriate term to describe the current state or trajectory of science is debatable, but accumulated evidence indicates that there is substantial room for improvement with regard to research practices to maximize the efficiency of the research community's use of the public's financial investment in research.", in MARCUS R. MUNAFÒ, BRIAN A. NOSEK, DOROTHY V. M. BISHOP, *et al*. "A manifesto for reproducible science." *Nature Human Behaviour* 1, 0021 (2017): 1-9, doi: 10.1038/s41562-016-0021.

terms of the right to access research data[120]; on the other hand, there is the autonomy of researchers which derives directly from the academic freedom, enshrined in the Article 13 of the EU Charter of Fundamental Rights[121], as well as in the constitutional charters of the various Member States.

By stipulating that the default option is the openness of research data, and the exception is the closure of such data, the European lawmaker is striking a balance between these fundamental rights, making the right to science prevail over the autonomy of the individual researcher deriving from academic freedom.

Bear in mind the relevance of the academic freedom: this freedom makes the field of public scientific research a peculiar branch of the public sector, with its own and different rules, mechanisms and institutions.

This aspect clearly emerges in relation to the disciplinary actions against members of the scientific community. Unlike civil servants, members of the scientific community, when subject to disciplinary proceedings, are submitted to the judgement of peers, thus bringing out "[…] the persistence of a significant peculiarity of the university status, which is related to the exclusion from the area of the contractual public employment, where for the latter the 'system' of disciplinary responsibility results from the provisions of collective agreements and codes of conduct"[122].

In addition, the typified reasons triggering the disciplinary actions are often related to failures in teaching and educational functions: any

---

[120] See: Section 4.1.3, in which the general right to science, underlying the Open Science paradigm, has been represented in terms of "right to access and reuse digital research data", as a real right to *open* research data.

[121] Section 3.3.2, where the Case C-66/18, *European Commission v Hungary* (2020) was specifically investigated.

[122] BARBARA GAGLIARDI, *La tutela amministrativa della libertà accademica.* (Milano: CEDAM, 2018), p. 66. [Translation from the Italian original text].

mentions to scientific community's misconduct in carrying out the research activity are missing[123].

On the contrary, the obligations and duties of the researchers regarding the scientific research activity are generally set out in the ethical and deontological codes of conduct, produced by the different universities or research centres, which go beyond the formal discipline, falling within the sphere of the soft law; or they are embodied in the research evaluation system. This system, in fact, bases career advancement on the "production" – almost more quantitative than qualitative – of publications, being almost more effective than a legislative provision.

Moreover, the traditional autonomy of scientific research, beyond having a strong background in terms of fundamental rights, has also traditionally been supported and guaranteed by case law at all levels: international[124], European[125], and national[126].

---

[123] BARBARA GAGLIARDI, *La tutela amministrativa della libertà accademica.* (Milano: CEDAM, 2018), p. 68.

[124] Consider, for instance, *Mustafa Erdoğan and Others v. Turkey*, released in 2014, by the European Court of Human Rights (ECHR). This judgment focuses on the aspect of academic freedom to express (and eventually the freedom not to express) of the researchers ("[…] academic freedom in research and in training should guarantee freedom of expression and of action, freedom to disseminate information and freedom to conduct research and distribute knowledge and truth without restriction […]"). Although this aspect is not related to the dimension of academic freedom that is relevant here, the ECHR also emphasises an aspect that is profoundly relevant to the discussion: the autonomy characterising research activity, which derives from this freedom. In particular, the joint concurring opinion by Judges András Sajó, Nebojša Vučinić and Egidijus Kūris, specifically pointed out that "[T]raditionally, academic freedom referred to a crucial element of university autonomy: non-interference by external powers in university teaching. This core academic freedom has increasingly been accepted as including personal freedom of expression, often in the sense of scholars' autonomy. It is in this sense that the maxim of the independence of university teachers and researchers was recognised as a constitutional principle by the French Constitutional Council […]. A similar approach can be found in the constitutional case-law of many other European countries". See: Case 346/04 and 39779/04, *Mustafa Erdoğan and Others v. Turkey* (2014), ECHR, ECLI:CE:ECHR:2014:0527JUD000034604; URL: http://hudoc.echr.coe.int/eng?i=001-144129.

[125] In Section 3.3.2, a recent ruling of the ECJ on this aspect has been analysed (see: Case C-66/18 *European Commission v Hungary* (2020) ECJ, ecli: EU:C:2020:792.

[126] To give just a few examples, it should be noted that in Italy, a considerable strand of court case law enshrining the autonomy of universities and research organisations is linked to the issue of evaluation and access to the research field, through public or national selections. See, *ex multiis*: Tribunale

Precisely in light of this strong academic freedom, the European lawmaker, in striking a balance in the Article 10 of the ODD, always guarantees, however, a decision against the openness of research data: where one of the opposing interests, investigated in Section 4.3.2, is prevailing in the specific case, the researcher may (or must) not make her data open.

Yet, in relation to the current "openness by default" clause set out in the Article 10 of the ODD, two different aspects of distress are noticed straight away.

First, the adoption of the openness by default approach is however left to the implementation of the Member States, which are in charge of putting in place the "national policies and relevant actions"[127] for research data. This bottleneck immediately points to the considerations already made in relation to the hurdle of national transposition, i.e., the potential weakening of the European strategy, which is subject to the decisions of the Member States. However, to express a more informed view on this aspect, it is necessary to wait and observe how the requirement will be concretely implemented in the various Member States[128].

Amministrativo Regionale per il Lazio (Sezione Terza), n. 201900108, 22/07/2020, ECLI:IT:TARLAZ:2020:8579SENT; Tribunale Amministrativo Regionale per la Lombardia (Sezione Terza), n. 201900007, 20/01/2020, ECLI:IT:TARMI:2020:115SENT. In France, this topic i salso highly debated, see: CLÉMENTINE GOZLAN, "L'autonomie de la recherche scientifique en débats: évaluer l''impact' social de la science." *Sociologie du travail* 57,2 (2015): 151–174, doi: 10.1016/j.soctra.2015.02.001; and also, LAMY ERWAN, SHINN TERRY, "L'autonomie scientifique face à la mercantilisation. Formes d'engagement entrepreneurial des chercheurs en France." *Actes de la recherche en sciences sociales*, 164,4 (2006): 23-50, doi: 10.3917/arss.164.0023. As regards Germany: "The Federal Constitutional Court of Germany has repeatedly declared the constitutional guarantee of the freedom of science, albeit not imposing a particular organisational model of the university" in GIOVANNI CORDINI, "Università istituzioni e imprese, aspetti di diritto comparato." *Politico* 60.3.174 (1995): p. 468. [Translation from the Italian original text].

[127] The Article 10(1), ODD.

[128] Not only considering the internal law of transposition of the entire Directive; but especially the national policies and plans through which national strategies for sharing and re-using research data will be effectively identified. Consider that in December 2021 the European Commission opened infringement procedures against 19 Member States "for failing to provide complete information on how the Open Data Directive is being transposed into national law", see: https://data.europa.eu/en/news/commission-urges-19-member-states-comply-open-data-directive.

Second, beyond the implementation of the openness by default clause by Member States, some considerations on the actual enforceability of the provision are required. In fact, a provision imposing the openness of publicly funded research data to researchers seems to be hampered by the very nature of the system. According to the wording of the ODD, the recipients of national policies are "research performing organisations" and "research funding organisations"[129]. Although the organisations funding and conducting research are crucial, as already been stressed above, the major role in the publicly funded scientific research is played by the researcher and her research team, who actually conduct the specific research project. These actors, as seen above, by their nature, benefit from wide discretionary powers, stemming from the Article 13 of the EU Charter of Fundamental Rights. Decisions concerning the organisation of the research project are generally the exclusive responsibility of the researcher and her research team – within the limits of what is granted by the eventual grant agreement, which actually defines the boundaries of this autonomy. Consequently, in the final instance, the decision to share (or, in this case, to close) the research data (as well as any other output of a scientific research project) is the exclusive responsibility of the researcher and her research team[130].

As just argued, the European lawmaker, directly in the Article 10 of the ODD, establishes a balance of rights that would seem to let the right to Open Science prevail over the autonomy of the individual researchers, in not sharing their research data.

However, what needs to be underlined here is that the practical implementation of this provision may risk going in the opposite direction. The autonomy of the researcher could allow her to opt for

[129] This aspect has been investigated in Section 4.2.3, (i) Actors.

[130] Disclaimer: please note, however, that it will be essential to study the national policies on this matter, in order to understand whether and how the "research performing organisations" and "research funding organisations" will be required to put in place controls on these choices, realised exclusively by individual researchers or individual research teams.

closure of the research data, to some extent even without having to account for that choice. In other words, the autonomy of the researcher, which stems from the academic freedom, traditionally protected by the courts – *de facto* – weakens, if not completely void, the legal provision of the Article 10 of the ODD that requires the openness of research data by default. In this context, therefore, the cultural approach, which traditionally makes closure (or competition) prevail over sharing, risks thwarting the aim set by the European lawmaker. As a result, it seems one of those cases that led Martin Weller to wonder "How openness won and why it doesn't feel like victory"[131].

In order to prevent the legal provision under investigation becomes an empty shell, it would be recommended to adopt a different approach. Consider that it is not argued that the normative intervention is irrelevant. However, the effort to support the openness of research data cannot be limited exclusively to legal requirements to the Member States: once again, it is a matter of legal coordination between different systems, levels and actors[132]. In other words, the regulatory provision needs to be complemented by a precise design of the processes, in order to support the effective adoption of the "openness by default" option, by the researcher. The aim should be to give a concrete implementation to what has been defined in France as "*services d'accompagnement adaptés*": the openness of research data, in France, was already included in a law introduced in 2016, that also envisaged such mentioned supporting mechanisms for the researchers[133].

---

[131] MARTIN WELLER, *The battle for openness*. (London: Ubiquity Press, 2014), pp. 1-27, 62.
[132] See: Section 2.4.3.
[133] See: Ministre de l'enseignment supérieur de la recherche et de l'innovation, *Deuxième Plan national pour la science ouverte*, 2021, p. 13, https://www.enseignementsup-recherche.gouv.fr/cid159131/plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-des-pratiques-de-science-ouverte-en-france.html: "L'obligation d'ouverture des données de la recherche publique, posée par la loi pour une République numérique de 2016, doit désormais se traduire dans les pratiques scientifiques grâce à des infrastructures et des services d'accompagnement adaptés. Elle est limitée par les exceptions légitimes encadrées par la loi, par exemple en ce qui concerne le secret professionnel, les secrets industriels et commerciaux, les données personnelles ou les contenus protégés par le droit d'auteur. Dans ces cas, les pratiques de partage des données devront être favorisées à travers la définition de

Similarly, the design of the process should be aimed at assisting the scientific community to the effective adoption of the openness of data by default. It is envisaged that the process should operate along two main strands: the introduction of incentives and the effective involvement of *ad hoc* actors.

First, the process of unlocking publicly funded research data should incentivise researchers in this direction. Opening the data (and making it FAIR as a precondition to openness in the first place[134]) takes time and energy in the daily operation of the researchers[135]. The rewarding perspective can be realised by evaluating these efforts within the evaluation process of the research activity and therefore of the researcher. It means, in other words, using the same mechanism that has traditionally made closure (and competition) prevail over collaboration, but reversing it.

Second, a fundamental step is the effective involvement of the *ad hoc* actors identified within the data unlocking process, i.e., the so-called data stewards. Data stewardship is defined as: "[T]he process and attitude that makes one deal responsibly with one's own and other people's data throughout and after the initial scientific creation and discovery cycle"[136]. Data stewards are, indeed, experts in a specific scientific domain, dealing with the management and curation of data, based on the assumption that "[I]f science has become indeed data driven and data is the oil of the 21st century, we better put data centre stage and publish data as first-class research objects, obviously with

---

protocoles maîtrisés".

[134] On this aspect, see: Section 4.1.2, on the notion of FAIR Data Principles; and also, see: Section 4.3 on the role of FAIRness as precondition of the openness, in an infraethic perspective.

[135] See: "From incidental surveys conducted under young PhD students, the gloomy picture emerges that they have to spend roughly 50% to 70% of their time on a process that is now called data munging or data wrangling" in BAREND MONS, *Data stewardship for open science: Implementing FAIR principles*. (Boca Raton, Florida: CRC Press, 2018), p. 11; the expression "data munging or data wrangling" refers to those complex operations of data manipulation and processing, which then allows the subsequent analysis. In other words, these are preliminary and preparatory tasks to the actual analysis, i.e., the actual implementation of the research project.

[136] BAREND MONS, *Data stewardship for open science, op. cit.*, p. 36.

supplementary narrative where needed, steward them throughout their life cycle, and make them available in easily reusable format"[137].

Traditionally, the role of data stewardship has been linked to the implementation of data FAIRness, but it is fundamentally much more than that: "[A]ny domain specialist in open, data-driven science should pay due respect to, and work closely with, data experts. This is far from trivial and a frequent reason for failure of projects, or even entire e-infrastructures"[138].

Going a bit further and taking a bigger picture look, it is suggested to adopt a vision of the role of data stewards as real enablers of openness. In other words, data stewards could be considered as the human interface between the systems of institutional repositories adopting openness as a default option and the users, i.e., researchers, often reluctant to share. Data stewardship could thus not only be a crucial aspect of good data management, but also a key element of the openness of publicly funded research data[139].

To sum up, this second legal hurdle to the openness of research data is therefore the applicability of the legal provision providing for openness by default, in terms of enforceability. In order to address this challenge, it has been suggested to adopt a different approach, starting from the European legal framework (the directive) and the national legal frameworks (the transpositions), to operate on the design of the

---

[137] BAREND MONS, *Data stewardship for open science, op. cit.*, p. 13. The need to foresee the presence of data stewards in every research area is not very different from the need expressed at the very beginning of the establishment of data science, by those who believed that "[…] data analysis needs to be part of the blood stream of each department and all should be aware of the workings of subject matter investigations and derive *stimulus* from.", in WILLIAM S. CLEVELAND, "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics." *International statistical review* 69.1 (2001): p. 22, doi: 10.2307/1403527. The difference is that for the Author, the presence of the data scientist was seen as a *stimulus* for research, a means to derive new and unprecedented benefits from data. In our context, data stewardship is a real necessity, which makes the difference between good and bad data management.
[138] BAREND MONS, *Data stewardship for open science, op. cit.*, p. 8.
[139] On the competences of data stewards, see: VALENTINA PASQUALE, EMMA LAZZERI, ELENA GIGLIA, "Data steward per i dati FAIR." *Conferenza GARR 2021 "Sostenibile/Digitale"* (2021): 77-80, doi: 10.26314/GARR-Conf21-proceedings-16.

process of opening up research data, in two strands: incentives and data stewardship.

In light of these considerations, the study proceeds with the third legal challenge to the openness of the research data in publicly funded research projects, represented by the licences under which such data are released.

4.2.4.3 Licenses for Research Data Sharing and Reuse

The Article 8 of the ODD states that "Member States shall encourage the use of […] standard licences". The standard licence is defined in the Article 2(5) as "[…] set of predefined re-use conditions in a digital format, preferably compatible with standardised public licences available online". Licences allow data reuse to be restricted to some extent, by imposing certain specific conditions that must be "[…] objective, proportionate, non-discriminatory and justified on grounds of a public interest objective"[140].

Recital 44 of the ODD emphasises the relevance of standardised licences, stating that "[O]pen licences in the form of standardised public licences available online which allow data and content to be freely accessed, used, modified and shared by anyone for any purpose, and which rely on open data formats, should play an important role in this respect".

Chiefly, the licence is an atypical contract allowing the transfer of a temporary right of use, for a given length of time, by a licensee to a licensor[141]. The latter remains the owner of the right granted for use, shifting from a model of transfer of ownership to a regime of access[142].

---

[140] The Article 8(1), ODD.

[141] PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica, op. cit.,* p. 212. See, also: PAOLO GUARDA, "Creation of Software Within the Academic Context: Knowledge Transfer, Intellectual Property Rights and Licences." *IIC-International Review of Intellectual Property and Competition Law* 44.5 (2013): pp. 512-515, doi: 10.1007/s40319-013-0078-6.

[142] JACOPO CIANI, "Governing Data Trade in Intelligent Environments: Taxonomy of Possible Regulatory Regimes between Property and Access Rights." *Workshop Proceedings of the 14th International Conference on Intelligent Environments, Ambient Intelligence and Smart Environments Series*, 23, 2018: 285-297, doi: 10.3233/978-1-61499-874-7-285. On the theorisation of the transition to

By contrast, the licensor benefits from the use of the licensed right, complying with the conditions laid down in the licence itself[143].

The Article 10(2) of the ODD, regulating research data, refers specifically to the compliance with Chapter III of the ODD itself, which contains the Article 8, concerning the licences.

The licences applied to the reuse of research data requires some further considerations, compared to the reuse of public sector data. The reason is made very clear by Koščík and Myška, who argued that "[T]he realisation of open access to data is customarily realised by granting a license allowing the use of such protected subject matter or ideally by waiving the respective existing rights. Opening up research databases without the consent of the right-holder equals an infringement of the granted exclusive rights. In the context of open research data, the question of who actually is the right-holder has the utmost importance, as only this person is legally entitled to license or waive the rights to the database"[144].

Data, if understood as mere facts, cannot be the object of copyright. As described in Section 2.3.3, copyright does not cover the idea or the mere fact[145] – which always remain in the public domain[146] – but the creative piece of work. If data, *per se*, are not protected by copyright protection, databases are. The notion of database is "[…] a collection of independent works, data or other materials arranged in a systematic

---

the access regime, see, in particular: JEREMY RIFKIN, *The age of Access: The new culture of hypercapitalism, where all of life is a paid-for experience.* (New York, Putnam Publishing: 2000), pp. 26-36.

143 MARK ANDERSON, "How to draft a licence agreement that is fair, reasonable and non-discriminatory: a ten-point plan." *Journal of Intellectual Property Law & Practice*, 13.5 (2018): 377–392, doi: 10.1093/jiplp/jpx212.

144 MICHAL KOŠČÍK, MATĚJ MYŠKA, "Database authorship and ownership of sui generis database rights in data-driven research." *International Review of Law, Computers & Technology* 31.1 (2017): 43-67, doi: 10.1080/13600869.2017.1275119.

145 See, the Article 9(2) of the Trade-Related Aspects of Intellectual Property Rights Agreement (TRIPS Agreement): "Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such".

146 On the concept of public domain, see: JACOPO CIANI, *Il pubblico dominio nella società della conoscenza.* (Torino: Giappichelli Editore, 2021), pp. 275-440.

or methodical way and individually accessible by electronic or other means"[147].

Such data collections can be protected by the copyright or the *sui generis* right. In fact, on the one hand, if a database is the result of a creative effort, i.e., the outcome of an original work of creativity, then the author will be protected by copyright. On the other hand, however, a *unicum* of the European law is represented by the so-called *sui generis* right, regulated by Directive 96/9/EC[148], that is, a right held by the creators of databases, aimed at protecting the effort made in the creation of the database. This effort is expressed in terms of "investment", qualitative and quantitative, both by the wording of the law[149], and by the case law of the ECJ, which has specified the boundaries of the notion[150]. Moreover, these two levels of protection may also be presented at the same time since the protection of one does not exclude that of the other[151].

The situation is extremely complex mainly for four reasons. First, as regards the protection given by copyright, the identification of the criterion of originality of a research data collection must be determined on a case-by-case basis[152].

Second, collections of research data which are created and not derived from other sources are outside the scope of application of the Directive 96/9/EC and thus outside the protection of the *sui generis* right[153]. In addition, "[R]aw data and collections thereof made without

[147] The Article 1, Directive 96/9/EC.
[148] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996, *on the legal protection of databases*, OJ L 77, 27.3.1996, p. 20–28, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31996L0009.
[149] Recital 7, the Article 7 "Object of protection", the Article 10(3) "Term of protection" of the Directive 96/9/EC.
[150] See: LUCIE GUIBAULT, ANDREAS WIEBE, *Safe to be open, op. cit.*, pp. 24-26; see, also: Case C-304/07, *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg* (2018) ECJ, ECLI:EU:C:2008:552.
[151] PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica, op. cit.*, p. 89.
[152] LUCIE GUIBAULT, ANDREAS WIEBE, *Safe to be open, op. cit.*, p. 21.
[153] LUCIE GUIBAULT, ANDREAS WIEBE, *Safe to be open, op. cit.*, p. 26.

substantial investment into its obtaining, verification or presentation do not enjoy either copyright or *sui generis* protection"[154].

Third, the situation is further complexified by the ODD itself, which states in Recital 54 that "[P]ublic sector bodies should, however, exercise their copyright in a way that facilitates re-use", where "the term 'intellectual property rights' refers to copyright and related rights only, including *sui generis* forms of protection"[155].

Finally, a further crucial aspect concerns the increasing relevance of the private sector in scientific research. Today the intertwining of private and public actors within research projects is more evident than ever[156]. In a context where boundaries are blurred, it is extremely more complex for the actors directly involved, i.e., researchers whose projects are funded with public money, to determine ownership on a case-by-case basis.

The European framework regulating copyright and *sui generis* rights raises a considerable number of issues related to the scientific research domain, which have been extensively investigated[157], and which will not be delved into in this dissertation. The aim pursued here is to consider only the relationship of these issues with what is regulated in the ODD directive, with specific reference to the sharing and reuse of research data: this relationship is characterised by

---

[154] MICHAL KOŠČÍK, MATĚJ MYŠKA, *"Database authorship and ownership of sui generis database rights in data-driven research.", op. cit.*, p. 45.
[155] Recital 54, ODD.
[156] This aspect is further investigated in Section 6.3.
[157] On the constraints of copyright law in scientific research, see: ROBERTO CASO, "Open Access to legal scholarship and copyright rules: a law and technology perspective." *Proceedings law via the Internet: free Access, quality of information, effectiveness of rights*, (2009): 97-110; THOMAS MARGONI, "The harmonisation of EU copyright law: the originality standard." *Global governance of intellectual property in the 21st century* (Cham: Springer, 2016): 85-105, doi:10.1007/978-3-319-31177-7_6; LUCIE GUIBAULT, THOMAS MARGONI, "Legal aspects of open access to publicly funded research." *Enquiries into Intellectual Property's Economic Impact* (2015): 373-414. For more insights specifically on the database directive, see: ROSSANA DUCATO, "'Adiós sui generis': A Study of the legal Feasibility of the Sui Generis Right in the Context of Research Biobanks." *Revista de derecho y genoma humano*, 38, 2013, 125-146; SAMUEL E. TROSOW, "Sui generis database legislation: A critical analysis." *Yale Journal of Law and Technology* 534.7 (2004): 534-642.

substantial ambiguity[158]. Although "[…] authorship cannot be decided upon contractually"[159], uncertainty must necessarily be overcome in order to allow a practical understanding of "who" can do "what", with certain research data, for which sharing and reuse are promoted. In this scenario the contractual dimension acquires wide relevance: "[…] the contractual parties often try to contractually bind themselves in order to establish a regime such that the desired beneficiary of a project acquires (in the broadest possible meaning) all the IP"[160].

In particular, standardised licences are of crucial interest: "[T]he definition of standardised contractual terms of use reduces, in fact, transactional costs between right holders and users and facilitates the reuse of information, even in cases where the identification of rights may appear problematic. In addition, such standardised terms can solve the user's problem of identifying the rights holder and negotiating the terms of use. In order to promote the widest access and reuse of scientific publications and data, lawmakers globally are promoting policies to create the conditions for open access"[161].

In the research context, the most widely used licences are primarily the Creative Commons (CC) licences[162], created by Lawrence Lessig in 2001, which are composable and irrevocable[163]; then, the General Purpose Licenses (GPL)[164], closely related to the release of free

[158] MIREILLE VAN EECHOUD, "A Serpent Eating Its Tail: The Database Directive Meets the Open Data Directive." *IIC - International Review of Intellectual Property and Competition Law volume* 52 (2021): 376, doi:10.1007/s40319-021-01049-7.
[159] MICHAL KOŠČÍK, MATĚJ MYŠKA, *"Database authorship and ownership of sui generis database rights in data-driven research.", op. cit.*, p. 58.
[160] *Ibid.*
[161] PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica, op. cit.,* p. 214. [Translation from the Italian original text].
[162] https://creativecommons.org.
[163] LAWRENCE LESSIG, "The Creative Commons." *Montana Law Review* 65.1 (2004); and, see, also: ADRIENNE K. GOSS, "Codifying a commons: copyright, copyleft, and the Creative Commons project." *Chicago-Kent Law Review*, 82.2 (2007): 963-996.
[164] https://www.gnu.org/licenses/gpl-3.0.html.

software; finally, the Open Data Commons (OCD) licenses[165], more recent, developed since 2008 by the Open Knowledge Foundation[166].

The scenario is clearly complex. On the one hand, if the goal of the Open Data European legal framework is to foster as much as possible the circulation of research data, then the conditions under which such sharing and reuse is made possible are fundamental. On the other hand, these conditions under which sharing and reuse are made legally possible relate to the rules on data ownership and the adoption of licences.

In light of this problematic scenario, therefore, two different levels of intervention are envisaged to address the issue, i.e., *de lege ferenda* and *de lege data*.

As regards the first level, i.e., *de lege ferenda*, it is fair to admit that the intervention of the European lawmaker is necessary to clarify and harmonise the discipline of the ownership of data and databases held by the public sector: "[…] the EU itself has no harmonised rules on copyright and neighboring rights for information held by the public sector. If in the EU we do not even have a common understanding of if and when certain information should be excluded from copyright, or be treated differently because it is publicly funded, or produced for the purposes of public tasks, what then is the yardstick against which the regimes of other countries are to be measured?"[167]. This reasoning is even more crucial if applied to the scientific domain and research data.

---

[165] https://opendatacommons.org.

[166] SIMONE ALIPRANDI, *Il fenomeno Open Data: indicazioni e norme per un mondo di dati aperti*. (Milano: Ledizioni, 2014), pp. 71-75. In addition, a phenomenon of so-called "proliferation of licenses" is emerging. It is mainly due to a substantial overlap between the concept of "share alike" and "strong copyleft". "Share alike" is a requirement according to which the material further released, the so-called derivative works, must be released under the same conditions as the original licence. Similarly, the "strong copyleft" feature has an impact on derivative works, requiring them to adopt the copyleft licence.

[167] MIREILLE VAN EECHOUD, *"A Serpent Eating Its Tail: The Database Directive", op. cit.*, p. 377. This issue is connected to the more complex need for revision of the current IP discipline, in general, based on the fact that "[…] the existing law is becoming outdated because of its inability to be enforced", and also on the fact that the digital world has profoundly changed the perception of "piracy" – even among legal professionals – as argued in a recent study: MALGORZATA CIESIELSKA, DARIUSZ JEMIELNIAK,

The second level of intervention proposed, *de lege data*, addresses the practical situation in which a researcher finds herself, faced with a partially contradictory European legislation: on one side, the legal discipline pushes for her research to be as open as possible; on the other, it does not provide support mechanisms to generate such openness. Once again, the ODD risks going from enablers to obstacles: "[I]f you do not issue a clear license that is legally binding on yourself and the other party, these other scientists may be afraid to touch your data for fear of being sued"[168].

Part of the solution to this problem lies in the standardised nature of the licences whose use it is promoted. Such standardised licences, expressly suggested in the Article 8 of the ODD, "[...] are made part of the metadata and can be read by a computer without a law degree. Now a would-be data user can instruct her computer to go and download all datasets with a particular license (or with an even broader mandate), and she can start dealing with the scientific issue that has her interest"[169].

Yet, the adoption of standardised licences is not sufficient *per se*. The focus here is again on the design of the processes towards openness, sharing and reuse of research data is realised: the scientific community should be guided in the choice of the most suitable licences for any given research project, as well as discharged from dealing with legal issues that may – legitimately – fall outside the scope of the individual researcher's expertise.

"Fairness in digital sharing legal professional attitudes toward digital piracy and digital commons." *Journal of the Association for Information Science and Technology* (2021): 1-14, doi: 10.1002/asi.24592. The interesting aspect that emerges from this study is the factual emergence of a "culture of sharing", which has been empirically established and concretised, and which goes in the opposite direction to what the legislative discipline wants to establish.

[168] Marcel P. Dijkers, "A beginner's guide to data stewardship and data sharing." *Spinal Cord* 57, (2019): 179, doi: 10.1038/s41393-018-0232-6.

[169] *Ibid.*

Once again, similarly to the second hurdle[170], the suggestion is to strengthen the profile of data stewardship, within the design of the processes of the openness of research data, at an institutional level. Supporting the training and participation of data stewards in research centres and universities should not be seen solely from a technical point of view. Data stewards should not only be highly expert in the field of research in which they operate, or enablers of the FAIR Data Principles. As seen before, these legal issues have acquired, and are acquiring, increasing relevance in an Open Science scenario, where the impact of new digital technologies, huge amounts of data, and immense computational power have deeply changed the traditional mechanisms of science. For this reason, the data steward should represent an interface between the systems of institutional repositories adopting openness as a default option and the researchers, not only as regards technical issues, but also legal ones.

In light of these considerations on ownership and licensing of research data, it is understood that there are many more complex issues to be faced concerning sharing and reuse of research data than sharing and reuse of public sector data in general.

Having thus clarified the scenario regarding the purely legal challenges facing the openness of research data in relation to the European legal framework of Open Data, now it is time to shed light on the challenges that go beyond legal boundaries, to fall into the realm of ethics.

4.3 Ethical Issues of the Open Research Data

Section 4.2 illustrates that the European legal framework of research data sharing and reuse provides for several moves: openness of research data as a "by default" option; balancing openness and closure based on the evaluation of certain interests (e.g., privacy, data

[170] See: Section 4.2.4.2.

protection, security, legitimate commercial interests, etc.); FAIRness of Open Data; adoption of licences, primarily standardised.

Beyond the legal domain, however, there is a sphere pertaining to other aspects, such as values, preferences, or choices, which no longer belong to the legal field, but rather to the ethical one, since "[E]thical judgements and choices are embedded in every aspect of data management, including choices that at first sight appear to be purely technical and therefore socially neutral"[171]. There is, precisely, an ethical dimension of the scientific knowledge, which, as argued by Vayena and Tasioulas, is its "*raison d'etre*"[172]. The framework of human and fundamental rights, investigated in Chapter 3, which are moral rights, "[…] possessed by all human beings, simply by virtue of their humanity"[173], underlies the Open Science paradigm.

Section 4.3.1 deals with the most problematic ethical issue of research data, namely the quality of research data. Then, in Section 4.3.2, the attention is drawn to the interplay between openness, quality and FAIRness of research data, proposing an infraethical view of the latter.

4.3.1 Research Data Quality

The quality of research data is a frequently mentioned but rarely defined feature in many contexts in which data are the key element: data-driven innovation, data-driven economy, data-driven decision making, data-driven science, etc.

The notion of data quality has multiple meanings. This concept has been connected to the so-called veracity, one of the "Vs" that traditionally defined the Big Data[174]. Recently, the OECD, in its study

---

171 SABINA LEONELLI, *La ricerca scientifica nell'era dei big data, op. cit.*, p. 83. [Translation from the Italian original text].

172 EFFY VAYENA, JOHN TASIOULAS, "The dynamics of big data and human rights: the case of scientific research." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): p. 3, doi: 10.1098/rsta.2016.0129.

173 EFFY VAYENA, JOHN TASIOULAS, *"The dynamics of big data and human rights", op. cit.,* p. 5.

174 STUART G. NICHOLLS, SINEAD M. LANGAN, ERIC I. BENCHIMOL, "Reporting and transparency in big

on "Enhanced Access to Publicly Funded Data for Science, Technology and Innovation", mentioned the data quality as a principle that "[…] comprises quality control through peer review, documenting the origin of sources, linking to original research materials and datasets, and data citation practices"[175]. This OECD study, therefore, reports that "[…] more needed to be done on overall quality assurance, by defining explicit and verifiable quality standards that could be captured quantitatively where possible"[176].

On the basis of this declared gap, the interpretation of data quality needs to be slightly expanded. According to Leonelli, good research data are (i) not outdated; (ii) reliable; (iii) a sample of reality; (iv) accurate[177].

(i) Updating. Out-dated data are those that have been collected and not subsequently revised. Such data are to some extent affected by the so-called "scaffolds"[178], i.e., "[…] the conceptual, social and material assumptions necessary for the construction of theories, technologies or infrastructures"[179]. In other words, these data are affected by the criteria on the basis of which they have been collected, structured, clustered, stored. Although these scaffolds do not linger for, to be

---

data: the nexus of ethics and methodology." in BRENT DANIEL MITTELSTADT, LUCIANO FLORIDI (eds.) *The ethics of biomedical big data* (Cham: Springer, 2016): p. 340, doi: 10.1007/978-3-319-33525-4_15. Big Data have traditionally been defined by multiple characteristics, including veracity. There is a broad strand of analysis on this aspect. See, among many: VIKTOR MAYER-SCHÖNBERGER, KENNETH CUKIER. *Big data: A revolution that will transform how we live, work, and think.* (Boston, Massachusetts: Houghton Mifflin Harcourt, 2013); ROB KITCHIN, GAVIN MCARDLE, "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets", *Big Data & Society*, 3.1 (2016): 1-10, doi:10.1177/2053951716631130; MONICA PALMIRANI, "Big Data e Conoscenza." *Rivista di filosofia del diritto*, 1 (2020): 73-91, doi: 10.4477/97021.

175 OECD, *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation* (Paris: OECD Publishing, 2020), p. 75, doi: 10.1787/947717bc-en.

176 *Ibid.*

177 Leonelli also identifies a fifth category, that of potential "social harm" arising from the processing of personal data or special categories of data (the category of data called "sensitive data", before the GDPR, now regulated in the Article 9 of the GDPR), which is not considered here. On the challenges of Open Science in relation to the European data protection legal framework, see: Chapter 5.

178 LINNDA R. CAPORAEL, *et al.*, *Developing Scaffolds in Evolution, Culture and Cognition.* (Cambridge: MIT Press 2014).

179 SABINA LEONELLI, *La ricerca scientifica nell'era dei big data, op. cit.*, p. 47.

replaced by the data themselves, a deep mark of them still persists, since they affect the result obtained. The presence of outdated data in scientific databases can become a real pitfall for research projects that have relied on this poor-quality raw material. By some academics, this problem has been exposed in terms of a lack of maintenance: "[T]his lack of maintenance undermines data quality and consequently hampers the sustainability of the resource. Other measures, such as curation and quality assessment are needed to counter the tendency towards degradation of the shared repositories"[180]. It is therefore essential that the data shared and reused are constantly up to date.

(ii) Reliability. Reliable data, then, are data whose quality is – to some extent – certified or certifiable through submission to specific control mechanisms. The most problematic aspect of this feature is that the reliability is variable, not universal, and, as such, changes in relation to many factors: scope of research, purpose of the project, end user, etc[181].

However, it is deemed that the issue of data reliability can only be addressed by focusing on the informational asymmetry that underlies it. In fact, at the basis of the issue of data reliability, there is an informational asymmetry between those who collect data and those who reuse such data[182]. Essentially, this informational asymmetry is precisely what should generate a benefit, in the wake of the famous statement "the best thing to do with your data will be thought up by someone else"[183]. Conversely, it is problematic to the extent that such

---

[180] POLYXENI VASSILAKOPOULOU, ESPEN SKORVE, MARGUNN AANESTAD, "Premises for clinical genetics data governance: Grappling with diverse value logics." in BRENT DANIEL MITTELSTADT, LUCIANO FLORIDI (eds.) *The ethics of biomedical big data* (Cham: Springer, 2016): p. 250, doi: 10.1007/978-3-319-33525-4_11.

[181] The impossibility of adopting universal standards is in fact stressed in SABINA LEONELLI, *La ricerca scientifica nell'era dei big data, op. cit.*, p. 53.

[182] The idea of an asymmetric distribution of data and rights has been described in see: MASSIMO DURANTE, *Computational Power: The Impact of ICT on Law, Society and Knowledge.* (New York: Routledge, 2021), pp. 127-146.

[183] ELENA GIGLIA, "Open Access e Open Science: per una scienza più efficace." *Journal of Biomedical Practitioners* 1.1 (2017): p. 16.

an informational asymmetry can generate a lack of reliability of the data itself, which, as a consequence, may compromise its reuse.

(iii) Sample of reality. Data are incomplete insofar as they are not an expression of the entire and complex reality. They are, instead, a part of it: in other words, a sample of the real word. It has been stressed by many scholars that "[…] the informative value of data is highly dependent on the context in which they are placed"[184]. The selective nature of the data is not in itself a problem. The problematic aspect "[…] is instead posed by the tendency of Big Data users to forget that what they are processing is not a comprehensive nor particularly well-balanced sample of reality, but rather a selection made partly for practical limitations and partly for conceptual reasons"[185]. The lack of awareness of this incompleteness is a central problem, which risks having major implications for the entire cycle of scientific research: "The idea that Big Data embodies a complete representation of reality is an illusion that is destroying the critical thinking with which researchers approach the analysis and interpretation of empirical data"[186].

(iv) Accuracy. The accuracy of data has to do with their cycle of evolution: many of the data used for scientific research purposes were initially collected, produced, stored and processed for other reasons. These reasons may be commercial: in this case, the aforementioned problem of public and private interplay, which is increasingly affecting scientific research, emerges in relation to their primary use. This problem, expressed in terms of trustworthiness, is actually based on a lack of transparency. The shortfall in transparency occurs when the purpose of scientific research disguises commercial aims: "In situations

---

[184] PETER MILLS, "Ethical reuse of data from health care: Data, persons and interests." in BRENT DANIEL MITTELSTADT, LUCIANO FLORIDI (eds.) *The ethics of biomedical big data* (Cham: Springer, 2016): 429-444, doi: 10.1007/978-3-319-33525-4_18. p. 430.

[185] SABINA LEONELLI, *La ricerca scientifica nell'era dei big data, op. cit.*, p. 58. [Translation from the Italian original text].

[186] SABINA LEONELLI, *La ricerca scientifica nell'era dei big data, op. cit.*, p. 64. [Translation from the Italian original text].

where the commercial value attributed to data far outweighs the interest in their scientific value, it is perfectly possible to completely dismiss the search for data that are accurate, correct and whose processing provides a reliable representation of reality. Thus, data production procedures proliferate for the sole purpose of providing credibility to pre-established positions and hypotheses that are politically, commercially or socially convenient. In these cases, the production of data cannot have the outcome of modifying what is already believed, because the only data that matter are those that can be used to support and strengthen opinions that already exist, or to improve products that have already been planned regardless of their scientific and social value"[187].

Other times, simply, data that will be secondarily used for research purposes, have been primarily collected with little accuracy: "For instance, electronic health records typically consist of data written by clinicians for clinical work without the interests of researchers, standardisation and interoperability in mind, while aggregation of observational data for purposes of identifying causal links is prone to selection, confounding and measurement biases […]. If data come to be processed automatically without 'human checks' […] or by algorithms beyond the capabilities of human understanding, the variable quality of the data undermines justification of the actions taken on their behalf"[188].

In light of these various features which help to define the notion of data quality, it is now necessary to wonder about who exercises control over this quality. Three levels of control over data quality can be identified: chiefly, the scientific community, which by assessing the outcome of the research project, indirectly also assesses the quality of

---

[187] SABINA LEONELLI, *La ricerca scientifica nell'era dei big data, op. cit.*, pp. 72-73. [Translation from the Italian original text].

[188] BRENT DANIEL MITTELSTADT, LUCIANO FLORIDI, "The ethics of big data: current and foreseeable issues in biomedical contexts." *Science and Engineering Ethics* (2016): 320-321. doi: 10.1007/s11948-015-9652-2.

the data on the basis of which the results were obtained[189]; then, the researcher who processes that data, in person; eventually, the database manager, who is also involved in assessing the quality of the data stored in its repositories[190].

Accordingly, starting from the analysis of the concept of research data quality and the three levels of actors involved in quality assessment (i.e., scientific community; researchers; database managers), two remarks are due.

First, it is essential to underline an aspect that is too often misunderstood, involving the concept of FAIRness. Although in the context of Open Science it is strongly recommended to adopt the FAIR Data Principles, FAIRness of data does not mean having good quality data.

The nature of FAIR research data certainly participates in the definition of data quality[191]. Consider, for instance, the reliability of the data: this is *also* built by a detailed description of the provenance, which is a fundamental aspect of the reusability of the FAIR guidelines, as previously explored. However, the definition of the provenance of the data is only part of the reliability issue: in fact, it is related to the process of decontextualisation of the data[192], required for sharing and reuse. Subsequently, nonetheless, it is always necessary a process of recontextualisation of data coming from others, carried out by the researcher herself, within her own research project[193]. In other words, it is not considered sustainable to narrow the issue of data quality only to a technical dimension. As an illustration, consider the implementation of the EOSC, organised in Advisory Groups (AGs) and Tasks Forces (TFs), which act on certain priority areas. According to

---

[189] Assuming that these data are submitted together with the final publication.

[190] Think, for instance, about the issue of obsolete and outdated data: it is mainly from this point of view that database administrators play a relevant function.

[191] This assumption is further clarified in the next Section 4.3.2.

[192] SABINA LEONELLI, *La ricerca scientifica nell'era dei big data, op. cit.*, p. 55.

[193] *Ibid.*

the EOSC Symposium 2021, "The AG Metadata and Data Quality consists of Semantic Interoperability and FAIR Metrics and Data Quality"[194]. It would thus seem that the issue of data quality is primarily considered from a technical point of view, in parallel with the implementation of FAIR metrics. In doing so, the dimension of the data quality issue that goes beyond the technical one, is lost. However, the "EOSC Task Force FAIR Metrics and Data Quality Charter" clearly indicates the very early stage of the implementation of this aspect, bringing out the possibility of a change of direction[195].

The second consideration to note in light of the quality criteria of the research data (i.e., updating; reliability; awareness of incompleteness; accuracy) is about the relevance of the openness. It can be admitted that although the data FAIRness is not equivalent to the good data quality, the openness is a fundamental guarantee of the latter. Only data that are "as open as possible, as closed as necessary" allow the scientific community the widest and deepest possible scrutiny of the criteria involved in achieving good data quality. If data are open and available, there is a way of reproducing experiments and overcoming the already discussed reproducibility crisis[196]. This consideration shifts the focus of the data quality issue to two other aspects. First, the major problem – at this level – becomes, from the legal perspective, the enforceability of the "openness by default" option, discussed in Section 4.2.3.2[197]. Moreover, it becomes essential to strengthen the control

[194] VERONICA BERTACCHINI, *et al.*, *EOSC Symposium Report, op. cit.,* p. 12.

[195] "As far as data quality is concerned, this topic has been less explored in EOSC", in The Task Force (TF) FAIR Metrics and Data Quality, EOSC Task Force FAIR Metrics and Data Quality Charter, 1st draft, 2021, p. 2, available at: https://www.eosc.eu/sites/default/files/tfcharters/eosca_tffairmetricsanddataquality_draftcharter_20210614.pdf. See, also, Section 4.1.3.

[196] See: Section 3.1.1. and, Section 3.2, in particular the "(O_3) Methods".

[197] As a further illustration of the strong need for more openness, also through bottom-up instances, see a recent proposal suggesting that journals accepting publication with results should request the submission of the data from which the researchers started: "[…] journals, in principle, should try to have their authors publicize raw data in a public database or journal site upon the publication of the paper to increase reproducibility of the published results and to increase public trust in science" in TSUYOSHI MIYAKAWA, "No raw data, no science: another possible source of the reproducibility crisis."

mechanisms of the scientific community, *ex post*. In this regard, in fact, "[…] lack of governance mechanisms that can assure proper quality control and adherence to current standards regarding methods, nomenclatures and documentation, render the open databases a potentially dangerous source of information for clinical purposes"[198].

Having clarified the fact that FAIRness, quality, and openness of research data are three distinct concepts, which go beyond the domain of law into that of ethics: what should be the relationship between these three factors? This interplay is investigated in next section.

4.3.2 Infraethics of FAIR Principles

By investigating the concept of data quality, it has been claimed that the good quality research data are (i) not outdated (and thus able to provide a real picture corresponding to a certain context), (ii) reliable, (iii) accurate and (iv) with the understanding that they are not an expression of the whole complex reality, but rather an incomplete representation of it.

The sharing and reuse of research data is expressly suggested by legal acts of the European Union, most recently by the ODD; but the requirement to conduct high-quality scientific research, and therefore to produce, collect and process good quality research data, is a provision embodied in codes of conduct or charters of deontological practice produced *by* the scientific community, *for* the scientific community.

New challenges and opportunities generated by the impact of new technologies and digital ICTs on science – which is represented in terms of Open Science – require to have conceptually a clear outline of the different concepts involved: it becomes crucial to dwell further on the relationship between openness, FAIRness and quality of data.

Molecular Brain (2020): pp. 13-24, doi: 10.1186/s13041-020-0552-2.

[198] POLYXENI VASSILAKOPOULOU, *et al. "Premises for clinical genetics data governance: Grappling with diverse value logics.", op. cit.*, p. 248. See also BAREND MONS, *Data stewardship for open science, op. cit.*, p. 9, which states: "A rigorous quality check on the supporting data for any conclusion in narrative is badly needed, and in many cases very cult or even impossible, if only due to the sheer size of the data".

As mentioned above, the FAIR Data Principles are guidelines that operate at a technical level to ensure that data are kept securely, are accessible, are interoperable, and that their formal validity has been checked. The FAIR Data Principles are not a means of ensuring good data quality: they provide for good data management. At the first sight, the two concepts of quality and FAIRness may appear to be overlapping. Certainly, they are related, since having good data management – hence, in other words, formal quality – is the first step to achieve good data quality. But differences do persist, and their interplay is more complex.

To understand the relationship between FAIR Data Principles and good quality data, the concept of infraethics, developed by Luciano Floridi, is adopted. According to Floridi, infraethics is "[…] the not-yet-ethical framework that can facilitate or hinder evaluations, decisions, actions, or situations, which are then moral or immoral"[199]. A close bond between infraethics and ethics certainly exists: "[…] the moral behaviour of a society of agents is also a matter of 'ethical infrastructure' or simply infraethics"[200]. Ethics is thus profoundly dependent on infraethics. Precisely for this reason, infraethics should strive not to be paternalistic, in order to avoid exerting excessive restraint on personal autonomy of agents. Floridi explains precisely how a pro-ethical design choice is not paternalistic to the extent that it allows a certain space for the individual's decision-making autonomy: "[S]trategies based on pro-ethical design may not allow you to obtain a driving licence unless you have indicated whether you wish to be an organ donor: the unbiased choice is still all yours"[201].

This does not mean that infraethics is morally neutral. Bear in mind that infraethics "[…] is the not-yet-ethical framework of implicit

---

[199] LUCIANO FLORIDI, "Infraethics–on the Conditions of Possibility of Morality." *Philosophy & Technology* 30.4 (2017): p. 392, doi: 10.1007/s13347-017-0291-1.
[200] LUCIANO FLORIDI, *The fourth revolution: How the infosphere is reshaping human reality.* (Oxford: OUP, 2014), p. 362, EPub.
[201] LUCIANO FLORIDI, *The fourth revolution, op. cit.*, p. 361.

expectations, attitudes, and practices that can facilitate and promote moral decisions and actions"[202]. The verb "can" is the key to understand the dynamic. In fact, infraethics has a dual nature: depending on how it is conceived and designed, it can facilitate different kinds of behaviour, actions, and attitudes. In other words, "[…] an infraethics is the grease that lubricates the moral mechanism in the right way and successfully"[203].

In order to clarify the concept of infraethics, consider the following analogy: "[…] the best pipes (infraethics) may improve the flow but do not improve the quality of the water (ethics); and water of the highest quality is wasted if the pipes are rusty or leaky"[204].

Applying the concept of infraethics to the topic under investigation, the FAIR Data Principles are thought as pertaining to infraethics and the quality of research data as pertaining to ethics. The most careful and meticulous respect of the FAIR Data Principles (infraethics) may improve the dissemination and sharing of research data but does not improve – *per se* – the very quality of the research data shared (ethics). By contrast, research data created or collected and processed with the highest quality are wasted if they do not respect those guidelines, i.e., the FAIR Data Principles, that allow them to be found, accessed, interoperable, and eventually reused. Following Floridi's analogy, the FAIR Data Principles are the necessary pipelines through which updated, reliable, wittingly incomplete, and accurate research data should flow. Compliance with FAIR Data Principles does not in itself, necessarily, produce good data quality, but certainly non-compliance with these principles leads to bad data.

In other words, considering the FAIR Data Principles to be infraethical and not ethical is relevant insofar as it underlines a substantial difference: on one side, FAIR data are not necessarily of

---

[202] Luciano Floridi, *The fourth revolution, op. cit.*, p. 364.
[203] Luciano Floridi, *"Infraethics–on the Conditions", op. cit.*, p. 392.
[204] Luciano Floridi, *The fourth revolution, op. cit.*, p. 367.

high quality; on the other hand, it is not possible to assess the quality of data solely on the basis of compliance with the FAIR Data Principles: "So, it is easy to mistake the infraethical for the ethical because whatever helps goodness to flourish or evil to take roots, it partakes of their nature"[205]. This assumption is not intended to undermine the concept of FAIRness, a pillar of the Open Science paradigm. But, on the contrary, the aim is to define the dimension of FAIRness, within the context of data sharing and reuse, which is extremely complex from a technical, legal, ethical and, more generally, from a governance point of view.

Since the full adoption of the FAIR Data Principles and their adaptation in the different fields of scientific research is one of the priorities of the current EOSC implementation phase[206], it is therefore essential to understand their scope, especially in relation to the concepts of openness and data quality: "[…] creating the right sort of infraethics and maintaining it is one of the crucial challenges of our time, because an infraethics is not morally good in itself, but it is what is most likely to yield moral goodness if properly designed and combined with the right moral values"[207].

This perspective, therefore, considers FAIRness as occupying the infraethical level; data quality as the ethical one; while the openness of research data – expressed in terms of sharing and reuse – is the goal to be aimed at, as an important asset provided by new technologies to scientific research. This view is crucial to understand the distinction between these apparently similar notions: having FAIR data does not mean having high quality data, although it is an essential precondition.

The adoption of this perspective can have a twofold impact, on the researchers and on the institutions.

[205] LUCIANO FLORIDI, *"Infraethics–on the Conditions", op. cit.*, p. 392.
[206] See the example given in Section 4.3.1, regarding the Task Force on FAIR Data as described in the EOSC Symposium 2021.
[207] LUCIANO FLORIDI, *The fourth revolution, op. cit.*, p. 367.

First, the proposed interpretation has an impact on the practical activities implemented in a research project (e.g., sharing FAIR data without quality assessment of the data). Some scholars have expressed this by using the term "fitness for use": "While the FAIR Data Principles are valuable for enabling trustworthy and interoperable access, use and reuse, data literacy is a separate issue, and potential reusers may lack the competencies when interacting with data, regardless of its FAIR-ness. Since no data are perfect, some data are good enough for reuse and reusers may take what data are available, that appear to meet basic criteria, and also come from a trusted organization […]. An informed understanding of fitness for use of science data would outline these considerations and inform the functionality and design of data, metadata, and, critically, the tools and guidance (e.g., documentation) to access both"[208].

Second, the interpretation of the FAIRness as the infraethical dimension has an impact how institutions, at any level, can foster certain practices (e.g., data curation; widespread presence of data stewards; promotion of coordinated multi-level strategies on the definition of codes of conduct; reform of the science evaluation model, etc.) or can hinder others (e.g., "publish or perish" mechanisms; purely quantitative evaluations of science; lack of coordination of strategic actions, etc.).

In addition, a clear conceptual framework can become a forward-looking model of data sharing, which can go beyond the boundaries of scientific research to be applied to different contexts[209]. Without a clear conceptual framework, it will never be possible to overcome the

---

[208] BRADLEY WADE BISHOP, *et al.*, "Scientists' data discovery and reuse behavior: (Meta) data fitness for use and the FAIR data principles." *Proceedings of the Association for Information Science and Technology* 56.1 (2019): 21-31, doi: 10.1002/pra2.00004.

[209] Consider, as an illustration, according to the European Data Strategy, the EU institutions represented the EOSC as a model for the definition of European Data Spaces. On this aspect see: LUDOVICA PASERI, "COVID-19 Pandemic and GDPR: When Scientific Research becomes a Component of Public Deliberation", in DARA HALLINAN, RONALD LEENES, PAUL DE HERT, *Data Protection and Privacy: Enforcing Rights in a Changing World* (London: Hart Publishing, 2021): pp. 165-167.

frequently mentioned "[…] lack of an adequate governance framework"[210].

Hence, having framed the legal (i.e., the EU Open Data framework), the ethical (i.e., data quality) and the infra-ethical (i.e., the FAIR Data Principles) dimensions of research data, the analysis moves forward to one of the recent European proposals on digital governance: the Data Governance Act (DGA)[211]. What implications will the DGA raise in relation to research data?

4.4 Looking Forward: Data Governance Act

The Data Governance Act (DGA) is a proposal for a regulation that could be defined as complementary to the ODD by integrating the European framework on data sharing and reuse. The complementarity is made evident by the scope of application. The Article 3 of the DGA states that the DGA provides for the reuse of certain categories of data, such as data held by the public sector that are protected on the basis of commercial confidentiality, statistical confidentiality, protection of third parties' intellectual property rights and protection of personal

[210] MARIAROSARIA TADDEO, "Data philanthropy and the design of the infraethics for information societies." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): p. 1, doi: 1.1098/rsta.2016.0113.

[211] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on European data governance (Data Governance Act)*, COM/2020/767, 25.11.2020, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767. It was released together with other proposals: European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, COM/2020/825 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN; European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on contestable and fair markets in the digital sector (Digital Markets Act)*, COM/2020/842 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN; and European Commission, Proposal for a Regulation of the European Parliament and of the Council, *laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM/2021/206, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. For a general overview, see Section 2.3.4, specifically devoted to introduce the proposed new regulations.

data. Therefore, the DGA concerns, in fact, the reuse of those public sector data excluded from the scope of the ODD[212].

The aspect that seems to have the greatest impact on scientific research is the so-called "data altruism". Section 4.4.1 explores what the DGA means by data altruism, who are the actors involved, and how such a reuse mechanism should actually operate. Later in Section 4.4.2, some general considerations on the impact of this mechanism in the context of sharing and reuse of research data is illustrated.

4.4.1 Data Altruism

Among the novelties introduced by the DGA, the most relevant as regards the research field, is the so-called "data altruism". Data altruism is defined by the DGA, in the Article 2(10):

> 'data altruism' means the consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking a reward, for purposes of general interest, such as scientific research purposes or improving public services.

Starting from the definition of the data altruism, it is worthwhile to explore (i) the actors involved in the mechanism; (ii) the condition of the voluntary release of data; and, (iii) the phases of the process of data altruism.

4.4.1.1 Data Altruism Actors

The mechanism of data altruism identifies the activities of several agents, which needs to be framed: (i) the data subject; (ii) the data holder; (iii) the data altruism organisations; (iv) the data users; and (v) the competent authority for registration. First, the data altruism, understood in terms of granting consent, may be carried out by two

[212] "The proposed Data Governance Act seeks to extend the principles of the Open Data Directive to a wider range of data, which is held by public authorities but subject to third-party intellectual property rights, to commercial or statistical confidentiality, or data protection restraints." in MIREILLE VAN EECHOUD, *"A Serpent Eating Its Tail: The Database Directive", op. cit.*, p. 376.

entities: the data subjects of the personal data processed; or the data holders of the non-personal data.

(i) The data subject. The notion of data subject is indirectly derived from the definition of personal data provided in the Article 4(1) of the GDPR: the data subject is the natural person, identified or identifiable, to whom the personal data pertain.

(ii) The data holder. The notion of data holder is defined by the DGA, in the Article 2(5), as "[…] a legal person or data subject who, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal or non-personal data under its control"[213].

There are then three further actors that make the mechanism actually work, described below.

(iii) The data altruism organisations, i.e. legal entities, specifically registered as such, operating not for profit, independent of any profit-driven data processing activity.

(iv) The data users, i.e. any "[…] natural or legal person who has lawful access to certain personal or non-personal data and is authorised to use that data for commercial or non-commercial purposes"[214].

(v) Finally, the so-called competent registration authorities, designated in every Member States, regulated in the Article 20 of the DGA. They have the task of keeping, at national level, a register of entities recognised as data altruism organisations. Furthermore, according to the Article 21 of the DGA, they are also required to monitor and supervise the work of data altruism organisations.

4.4.1.2 Conditions

The data altruism is based on two conditions, expressed in the Article 2(10) of the DGA. The first condition is that reuse by data subjects and

[213] The Article 2(5), DGA.
[214] The Article 2(6), DGA.

data holders must be granted to the data altruism organisation, free of charge, not in return for a reward. This condition is clearly intended to avoid the establishment of a buying and selling of personal data[215].

The second condition is that the reuse must pursue public interest purposes[216]. These purposes of general interest are specified in Recital 35, which states: "Such purposes would include healthcare, combating climate change, improving mobility, facilitating the establishment of official statistics or improving the provision of public services. Support to scientific research, including for example technological development and demonstration, fundamental research, applied research and privately funded research, should be considered as well purposes of general interest"[217].

4.4.1.3 Phases

The process of data altruism, based on data "[…] voluntarily made available by individuals or companies for the common good"[218] consists of several stages, exposed below.

(1) The registration. Entities interested in being recognised as data altruism organisations have to complete a registration process, with the competent authority (or authorities) at national level, as regulated in the Article 17 of the DGA. It is noted that the entity requesting registration, among the various information required, should also indicate "[…] the purposes of general interest it intends to promote

---

[215] However, there is a strand of research on this proprietary dimension of personal data. See, among others, KONSTANTINA SAMARA, "Selling Personal Data: The Legal Framework and Nature of Personal Data Selling Transactions Under GDPR." *Personal Data Protection and Legal Developments in the European Union*, IGI Global (2020): 34-59, doi: 10.4018/978-1-5225-9489-5.ch003; and VANESA-MADALINA VARGAS, "The new economic good: Your own personal data. An integrative analysis of the Dark Web." *Proceedings of the International Conference on Business Excellence*. 13.1 (2019): 1216-1226, doi: 10.2478/picbe-2019-0107.

[216] On the concept of "public interest" see: Section 5.3.1.

[217] Recital 35, DGA.

[218] "Budgetary implication" of the Explanatory Memorandum, that was issued with the proposed regulation, giving some information about the decisions taken within the DGA, p. 8.

when collecting data"[219]. The general interest is primarily identified by scientific research[220], and as claimed by Pagallo and Bassi years ago, it is often difficult to be precise about the aims pursued in relation to research[221]. For this reason, Recital 36 specifies that "[D]ata subjects in this respect would consent to specific purposes of data processing, but could also consent to data processing in certain areas of research or parts of research projects as it is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Legal persons could give permission to the processing of their non-personal data for a range of purposes not defined at the moment of giving the permission"[222]. In this manner, a fair amount of leeway is granted.

After that, if the requesting entity meets all the requirements laid down by the DGA, it will be included in the national register of data altruism organisations, by the competent national authority or authorities, within 12 weeks from the date of application, pursuant to the Article 17(5) of the DGA.

(2) The voluntary exchange of data. The voluntary release of personal data by data subjects, or non-personal data by data holders, to data altruism organisations is done through the provision of consent. This consent is given in compliance with the two conditions just described above, i.e., no reward and public interest purposes. Furthermore, Recital 38 clearly connects the provision of consent for data altruism to the consent regulated in the GDPR and its conditions of validity, as set out in the Article 7 of the GDPR[223].

---

[219] The Article 17(4)h, DGA.

[220] The specific connection between "public interest purposes" and "scientific research" is made explicit by the DGA itself, in the definition of the concept of data altruism, in the Article 2(10): "[…] for purposes of general interest, such as scientific research purposes or improving public services".

[221] UGO PAGALLO, ELEONORA BASSI, *"Open Data Protection: Challenges, Perspectives, and Tools for the Reuse of PSI"*, *op. cit.*, p. 183.

[222] Recital 36, DGA.

[223] "Typically, data altruism would rely on consent of data subjects in the sense of Article 6(1)(a) and

In addition, the Article 22 of the DGA envisages the possibility for the European Commission to adopt a specific form for the provision of consent by data altruism, specifically aimed to promote it. However, the goal pursued at this stage is to create large data repositories, as specifically stated in Recital 35. The collection and storage of such data, by registered data altruism organisations, becomes a way to create rich data repositories: "[T]his Regulation aims at contributing to the emergence of pools of data made available on the basis of data altruism that have a sufficient size in order to enable data analytics and machine learning, including across borders in the Union"[224].

(3) The processing by data users. The registered data altruism organisations may give to several natural and legal persons the possibility to process the data they hold, for purposes of general interest, eventually on the basis of a fee. Each data altruism organisation is required to keep accurate records[225] – very similar to the processing register set out in the Article 30 of the GDPR – concerning a set of accurate information about the specific data processing activities, based on the data altruism consent. The main difference with the requirements of the GDPR register of processing activities is that in the data altruism register it is demanded to record the occurrence of any fees paid by the entities processing the data.

Moreover, every year, the data altruism organisations are required to submit an annual activity report to the competent national authority, summarising the activities carried out on those data voluntarily and freely given by data subjects and holders.

(4) Duty of communication to data holders. Each data altruism organisation, pursuant to the Article 19 of the DGA, has several reporting obligations towards data holders.

9(2)(a) and in compliance with requirements for lawful consent in accordance with Article 7 of Regulation (EU) 2016/679.", in Recital 38 of the DGA.

[224] Recital 35, DGA.

[225] The Article 18, DGA.

In particular, entities are required to communicate the purposes for which further processing of data is permitted to third parties, and "[…] any processing outside the Union"[226]. Extremely significant (and also problematic, in certain respects) is the fact that, any organisation of data altruism also has a function of control over the entire lifecycle of the data that is given to third parties to process. The Article 19 of the DGA, states, in fact, that "[T]he entity shall also ensure that the data is not be used for other purposes than those of general interest for which it permits the processing"[227].

The problematic aspect emerging in this phase is the one already identified by Bassi as follow: "[T]he purpose of the notion and regulation of re-use of public sector data is intrinsically generic and open to any possible use of the data, and appears from the very definition of 're-use'"[228]. The assumption was referred to the discipline established by the PSI Directive, but the reasoning also applies to the DGA. What is underlined here, in fact, is the lack of control over a wide and broad range of possible actions, which fall under the concept of reuse. These possible actions, by specific nature, tend to drift easily out of the control of the given organisation[229].

After having shed some light on the fundamental aspects of the data altruism mechanism (i.e., actors, conditions, and phases of functioning), it is now necessary to slightly broaden the analysis by formulating some observations on the link between the DGA and the current European system of sharing and reuse of research data.

[226] The Article 19(1), DGA.
[227] The Article 19(2), DGA.
[228] ELEONORA BASSI, "PSI, protezione dei dati personali, anonimizzazione." *Informatica e diritto* 37.1-2 (2011): p. 67. [Translation from the Italian original text].
[229] This phase of the data altruism process is closely related to the remark number (iii) "on the duty to control in charge of the data altruism organisations", which will be discussed in the following Section 4.4.2.

4.4.2 Research Data Sharing and Reuse between ODD and DGA

The DGA, and in particular the data altruism mechanism, appear to have a considerable impact on sharing and reuse of research data. The DGA was put in place with the specific intention of complementing the scenario on the reuse of public sector data. In addition, scientific research is relatedly mentioned in both the DGA and the Explanatory Memorandum accompanying its publication.

The concept of data altruism is not new. On the contrary, there is a strand of research on data philanthropy, which has been slightly investigated over the years[230]. Starting from the studies carried out so far in this field, and from what is envisaged in the new regulatory proposal of the European Commission, some considerations has been developed. In particular, the remarks further discussed below, are about: (i) the practical implementation of data altruism; (ii) the security issues of the mechanism; (iii) the duty of control that is in charge of data altruism organisations; (iv) the data ownership; and, finally, (v) the consent.

(i) The practical realisation. Starting from the assumption that the data philanthropy or altruism is not a moral principle, as argued by the philosopher Mariarosaria Taddeo[231], it is necessary to wonder about the practical feasibility of this mechanism. In light of a well-established trend[232], it is not difficult to imagine that there might be a fair amount of success on the side of data subjects and data holders who release their personal and non-personal data for the pursuit of general interest purposes. Several experiences show a general inclination to

---

[230] See, primarly: MARIAROSARIA TADDEO, *"Data philanthropy and the design of the infraethics for information societies.", op. cit.*; but, also: MARIAROSARIA TADDEO, "Data philanthropy and individual rights." *Minds and Machines* 27.1 (2017): 1-5, doi: 10.1007/s11023-017-9429-2; YAFIT LEV-ARETZ, "Data philanthropy." *Hastings Law Journal* 70 (2018): 1491-1546. ALEXANDRA GIANNOPOULOU, "Access and Reuse of Machine-Generated Data for Scientific Research." *Erasmus Law Review* 2 (2019): p. 159, doi; 10.5553/ELR.000136.

[231] MARIAROSARIA TADDEO, *"Data philanthropy and the design of the infraethics for information societies.", op. cit.* 4.

[232] MARIAROSARIA TADDEO, *"Data philanthropy and individual rights.", op. cit.*, p. 1.

release more easily personal data for scientific research purposes[233]. Similarly, it is not difficult to imagine data users, which may include universities, research centres, but also private companies, foundations, etc.

It is more difficult to identify entities undergoing the registration process to become data altruism organisations. Related to the acquisition of this qualification, in fact, there are several tasks and obligations, such as the control of possible users, comparable to the obligations related to the role of the data controller, as identified in the GDPR. Conversely, the benefits seem to be lower for this entity, i.e., the data altruism organisation, than for the data controller. In fact, it should be remembered that it must be a non-profit organisation, "[…] independent from any entity that operates on a for-profit basis"[234], as well as it must "[…] perform the activities related to data altruism take place through a legally independent structure, separate from other activities it has undertaken"[235]. In addition, this organisation, beyond the set of control and reporting obligations, must also ensure a solid infrastructure system. The goal is to create pools of data and this data must be stored, transferred, and managed, which makes the infrastructure aspect absolutely central.

(ii) The security issues. Closely related to the last aspect of infrastructure, a remark about the security risks emerges. Recital 36 states that "[L]egal entities that seek to support purposes of general interest by making available relevant data based on data altruism at

[233] For instance, as illustrated in UGO PAGALLO, *Il dovere alla salute. Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario*, (Milano-Udine, Mimesis: 2022), Chapter IV, Section 2.2, "The Eurobarometer report released in March 2020 investigated "attitudes about the impact of digitisation on our daily lives". [...] In twenty EU Member States, the majority of respondents declared they would like to share their data mainly to improve research and medical treatment". [Translation from the Italian original text].

[234] The Article 16(b), DGA.

[235] The Article 16(c), DGA. On the necessary incentives to pool data, from an economic point of view, see: BRUNO CARBALLA-SMICHOWSKI, NÉSTOR DUCH-BROWN, BERTIN MARTENS, "To pool or to pull back? An economic analysis of health data pooling." *JRC Digital Economy Working Paper* (2021): 1-75.

scale and meet certain requirements, should be able to register as 'Data Altruism Organisations recognised in the Union'. This could lead to the establishment of data repositories"[236]. Thus, to some extent, it seems that a centralised approach to the management of data is envisaged. The centralisation of data always brings with it a number of challenges from a security point of view, making those holding the data both very powerful, and at the same time very vulnerable. Very powerful, because it generates "[…] the emergence of pools of data made available on the basis of data altruism that have a sufficient size in order to enable data analytics and machine learning, including across borders in the Union"[237]. Highly weak because they are more easily targeted by cyber-attacks and data breaches. This is precisely why the EOSC project is a system of a federated and non-centralised nature.

(iii) The duty to supervise. As mentioned above, the Article 19 of the DGA establishes a duty to control in charge of any data altruism organisations, over the third parties that are allowed to process the data: "The entity shall also ensure that the data is not be used for other purposes than those of general interest for which it permits the processing"[238]. Although this requirement is understandable on principle, it does not seem easy achievable in practice. This mechanism of mutual controls seems to strongly refer to the mechanisms of accountability of the GDPR which establish[239], for the data controller, a set of duties, also with regard to the data processors, those who actually process the data[240], in the name and on behalf of the controller.

---

[236] Recital 36, DGA.
[237] Recital 35, DGA.
[238] The Article 19(2), DGA.
[239] The meta-principle of accountability is established in the Article 5(2) of the GDPR, which states that "The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability')". It can be considered as a meta-principle because it is the core principle of the whole GDPR and guides the entire European approach to data protection. On this aspect, see, Section 5.2.4 and Section 6.4.3.
[240] According to the Article 28 of the GDPR, the processor process personal data on the behalf of the controller.

There is, however, a substantial difference between the mechanism provided for by the GDPR and that provided for by the DGA. In the GDPR, the data processor processes personal data on behalf of the controller, but it is always the latter who determines the means and purposes of the processing. In some ways, it is fair to admit that the data processor is a mere executor of tasks.

In the case of the DGA, by contrast, it would appear that entities registered as data altruism organisations enable third parties to process the data held by them on a much wider basis. Admittedly, data altruism organisations must only ensure that these users conduct processing for purposes of general interest. However, as described above, the same DGA considers such purposes in a very broad and general way, referring to scientific research or the improvement of public services. These categories are very broad, and without the identification of further boundaries, much can fall within the all-encompassing terms set by the European proposal.

As a consequence, the implementation of the duty of monitoring provided for in the Article 19(2) of the DGA does not seem at all effortless. It is not an easy task for the data altruism organisation to exercise control – similar to that of the data controller, under the GDPR – over a third party that processes data with an extremely greater freedom than that accorded by the GDPR to the processor.

(iv) The ownership. Related to the duty to supervise upon the data altruism organisations, there is another relevant consideration on any third parties that are given the possibility to process data from data altruism organisations. Such organisations, in order to obtain the registration, must demonstrate that they meet a number of criteria, including that they "[…] operate on a not-for-profit basis and be independent from any entity that operates on a for-profit basis"[241]. However, this condition of independence and absence of commercial

[241] The Article 16(1)*b*, DGA.

profit is not a requirement for third parties who may be allowed to process such data. The only condition is the broad limitation of processing for purposes of general interest. But apparently there is no reason why such a third party (i.e., data user) should not be allowed to process the data for activities which, while respecting the purpose of general interest, are at the same time directed towards the pursuit of a commercial purpose. This different purpose may constrain further processing of the data. Such a scenario would not in itself infringe the provisions of the DGA, but it would create some problems in terms of data ownership.

(v) The consent-based altruism. A further consideration concerns the identification of consent as the legal basis underlying the mechanism of data altruism. Consider that one of the goals of the introduction of the GDPR, compared to the previous discipline of the Directive 95/46/EC, was precisely to overcome the model of personal data processing primarily based on consent. The Article 6 of the GDPR provides for a set of mandatory legal bases for the implementation of the processing of personal data: consent thus becomes one of the possible bases. The *ratio* for this choice made by the European lawmaker in 2016 was precisely to replace a consent-based approach that had proved to be ineffective[242]. Hence, it seems peculiar to relate the whole mechanism of data altruism, again, to the consent.

These remarks suggest that several aspects of the data altruism mechanism still need to be further clarified: "Data philanthropy is morally ambiguous [...] as it can either foster social development, knowledge, and the flourishing of information societies or can help

---

[242] On the ineffectiveness of the consent-based: BART W. SCHERMER, BART CUSTERS, SIMONE VAN DER HOF, "The crisis of consent: How stronger legal protection may lead to weaker consent in data protection." *Ethics and Information Technology* 16.2 (2014): 171-182, doi: 10.1007/s10676-014-9343-8; MARTINO TREVISAN, *et al.* "4 Years of EU Cookie Law: Results and Lessons Learned." *Proceedings on Privacy Enhancing Technologies* 2019.2 (2019): 126-145, doi: 10.2478/popets-2019-0023; EOIN CAROLAN, "The continuing problems with online consent under the EU's emerging data protection principles." *Computer Law & Security Review* 32.3 (2016): 462-473, doi: 10.1016/j.clsr.2016.02.004.

steering the design of current and future societies in the opposite direction"[243]. In order to understand which direction the European institutions intend to promote, it will be necessary to explore further expected legislative developments.

4.5 Conclusive Remarks

Chapter 4 discussed the openness of research data, in terms of sharing and reuse, as a fundamental part of the Open Scientific Research Process.

The concept of research data and FAIR Data Principles have been explored. Specifically, research data have been defined as follow:

> Research data are factual records, (i) used for research purposes; (ii) as part of a process of analysis traditionally called scientific method; (iii) available to enable the reproducibility of experiments, in order to ensure the review of the research output by the scientific community.

This preliminary analysis has enabled a distinction between openness and FAIRness of research data, as well as shedding light on the distinction between legal and ethical issues involved in sharing and reuse research data.

Accordingly, then, the investigation focused on the legal challenges of open research data.

First, the analysis started by redefining the concept of legal interoperability. This concept is often mentioned in the implementation process of the EOSC and in other major EU Open Science projects to refer to the whole range of legal-cultural issues related to the openness of research data.

However, it is not frequently defined, and considerable uncertainty arises from it. For the sake of clarity, legal interoperability has been redefined emphasising two dimensions, i.e., coordination and harmonisation: the legal interoperability as a co-ordination process

[243] MARIAROSARIA TADDEO, *"Data philanthropy and individual rights."*, *op. cit.*, p. 2.

aims to enable the legal reuse of data, taking into account the multiple disciplines involved (e.g., copyright, database regulation, data protection, etc.); and as an harmonisation process relates to the barriers resulting from the possible multiple jurisdictions involved in the sharing and reuse of research data[244].

After that, an overview of the EU Open Data legal framework has been provided and the Article 10 of the ODD has been critically examined.

In light of these analyses, the legal challenges to the sharing and reuse of research data have been investigated. These challenges were represented by (i) the national transposition; (ii) the applicability of the openness-by-default clause; and (iii) the choice of licences through which to share data.

(i) With regard to the national transposition of the ODD, it was investigated how suitable the legal instrument of the directive was to regulate the matter. While such legal instrument might be effective for sharing and reuse of public sector data, it is not effective for research data. The specificity of the field could risk fragmentation of the discipline. While waiting for national transpositions to be investigated, attention was therefore drawn to a further level of coordination, which goes beyond the national dimension: the local level of research centres and university networks[245].

(ii) Regarding the "openness by default" option, two considerations must be stressed. First, this condition is left to the discretion of the Member States, with the risk of weakening the strategy set out by the European directive. Second, there are still some concerns about the actual enforceability of the clause: even if it were envisaged by the Member States, the openness of research data would still be left to the discretion of researchers. Although this leeway of autonomy for

[244] A specific aspect of the harmonisation dimension of the concept of legal interoperability is further discussed in Section 6.3.2 about the impact of the US CLOUD Act on the EU scientific research.
[245] The relevance of the local level in the governance of the scientific research is further discussed in Section 6.4.3, related to the case study of this dissertation.

researchers is vital, related to the academic freedom, the risk is that the provision of the Article 10 would be devoid of any effectiveness.

For this reason, a different approach was recommended, based on legal coordination[246]: the regulatory provision needs to be complemented by a precise design of the processes, in order to support the effective adoption of the "openness by default" option, by the researcher. In other words, the design of processes should (a) include an incentive system for the researcher who spends time and effort in generating FAIR and open data; (b) reinforce the role data stewards, interpreting them as real interfaces between the systems of institutional repositories adopting openness as a default option and the users, i.e. researchers, often reluctant to share.

(iii) Concerning licences, a complex scenario determined by the relationship between data ownership and data circulation emerged. Here, two possible levels of intervention were envisaged: *de lege ferenda*, the need for a legislative harmonisation at European level on data ownership; *de lege data*, the focus was again placed on the design of the processes, imagining a procedure in which researchers are guided in the choice of the most appropriate licence. Once again, therefore, the relevance of the interface function played by data stewards was stressed, not only as technical, but also as legal experts.

Then, considering the ethical challenges, the concept of quality of research data has been represented identifying four characteristic features (i.e., data update; reliability; awareness of bias; accuracy), and the three actors involved (i.e., scientific community; researchers; database managers).

This analysis highlighted the substantial difference between the concepts of FAIRness, quality and openness of research data. The link between these three concepts has been described adopting Floridi's notion of infraethics: the FAIRness concerns the infraethical level, the

[246] As described in Section 2.4.3.

data quality concerns the ethical level, while the openness of research data – expressed in terms of sharing and reuse – is the goal to achieve. This interpretation has shown an impact both on researchers, in implementing their research projects, and on institutions, in encouraging or hindering certain practices.

Afterwards, looking forward, the impact of the Data Governance Act (DGA) on the current EU framework for sharing and reuse of research data has been explored. The focus was on the data altruism mechanism, taking into account actors involved; conditions underpinning the mechanism; the various phases in which the process is articulated.

Finally, some remarks on the impact of the data altruism mechanism on the sharing and reuse of research data has been illustrated. In particular, these considerations concerned: (i) the practical implementation of data altruism; (ii) the security issues; (iii) the duty of control in charge of any data altruism organisations; (iv) the data ownership; (v) and the consent at the basis of the mechanism. However, much will depend on how this proposal for a regulation will take shape at European level, in the next phases.

Chapter 4 dealt with research data, providing an answer to RQ4:

> What is the relationship between Open Science policies and the European Open Data legal framework?
> How do the two interact?

Instead, next chapter focuses on another dimension of the Open Scientific Research Process, exploring data protection issues in the context of Open Science.

Chapter 5

Data Protection Issues in Open Science

As Giovanni Buttarelli, the then European Data Protection Supervisor (EDPS), stressed when the General Data Protection Regulation (GDPR)[1] was enacted, "[…] data protection is no longer an optional extra"[2]. The data protection discipline is the set of rules designed to safeguard the fundamental right to personal data protection, as enshrined in the Article 8 of the EU Charter of Fundamental Rights: it is now a real cornerstone of the European Union.

For this reason, a study on the European approach to Open Science, focusing on its legal challenges, needs necessarily to shed light on the often mentioned but under-researched topic of personal data protection in the Open Science scenario. This chapter investigates the data protection issues related to the processing of personal data for research purposes.

Section 5.1 starts identifying which kind of processing activities are carried out in the Open Science context[3].

---

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, *on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ L 119, 4.5.2016, p. 1–88, ELI: http://data.europa.eu/eli/reg/2016/679/oj.

[2] Giovanni Buttarelli, "The EU GDPR as a clarion call for a new global digital gold standard." *International Data Privacy Law*, 6.2 (2016): p. 77, doi: 10.1093/idpl/ipw006. In addition, the GDPR has a real "expansive nature" as claimed in Marco Orfino, "Minori e diritto alla protezione dei dati personali." in Marco Orfino, Franco Pizzetti (eds.), *Privacy, Minori e Cyberbullismo*. (Torino: Giappichelli Editore, 2018), pp. 1-30. Here, the Author argued an "expansive nature of the EU data protection regulation" ("*capacità espansiva*", p. 4), in comparison with the right to privacy. Going beyond, it seems fair to admit an expansive nature of the protection of the GDPR in the context of the digital revolution, broadly understood.

[3] Consider that the "processing activity" is the concept on which to base a data protection analysis. The Article 4(2) of the GDPR defines it as follow: "'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval,

Then, Section 5.2 untangles the discipline of data protection for research purposes, taking into account the European level, considering the requirements of the GDPR, and the national level, referring to the wide discretional power of Member States in the research domain.

Afterwards, Section 5.3 is dedicated to the main data protection issues in the field of scientific research, represented by: (i) the identification of the lawful legal basis of processing activities for research purposes; (ii) the compliance with the principle of data minimisation and the subsequent problems related to the anonymisation techniques; and (iii) the challenges posed by the transfer of personal data to third countries and international organisations. This analysis shows that the major data protection issues in the context of research are not generated by the emergence of the Open Science paradigm. Rather, in the Open Science scenario, the previously existing weaknesses of the system tend to be more evident. The purpose here is to provide some suggestions to tackle these weaknesses, guided by the tenets of the Open Science.

In contrast to the claim of an unbreakable divergence between data protection and Open Science, Section 5.4 investigates the basis for a potential convergence. This analysis is developed addressing the major claims against the convergence of data protection and Open Science represented by: (i) the development of inequalities supposed to be generated by the Open Science paradigm; (ii) the risk of science privatisation, stemming from the prevailing market interests over those of science; (iii) the use of untrustworthy technologies; (iv) and the clash of cultures, represented by the openness of science against the supposed closeness of the GDPR. This reasoning aims to tackle these major concerns, in order to argue the convergence between data protection and Open Science.

---

consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction". The identification of processing activities should always be the first step in an analysis of compliance with the set of rules laid down in the GDPR.

However, it is recognised that some steps still need to be taken to achieve an effective data protection in the field of scientific research. The intention is to bridge this gap, in Section 5.5, by identifying the persistent barriers to an effective protection of personal data processed for research purposes. The factors identified as persistent barriers are the following: (i) the national fragmentation; (ii) a lack of all-encompassing data governance; and (iii) a closed science. Overcoming these obstacles would allow a wide convergence between Open Science and data protection, ensuring a great level of protection for data subjects involved in research, without restricting the free movement of knowledge.

Finally, Section 5.6 draws some conclusive considerations.

5.1 Processing Activities in the Context of Open Science

The protection of personal data is more about individuals, than about data *per se*: this is illustrated by the fact that the protection of personal data is a fundamental right, enshrined in the Article 8 of the Charter of Fundamental Rights of the European Union, and that the related right to privacy is considered a right of personality under civil law[4]. The European discipline on data protection, i.e., the GDPR, starting from Recital 2, emphasises that the subject of protection is the individual as such: "The principles of, and rules on the protection of natural persons with regard to the processing of their personal data should, whatever their nationality or residence, respect their

[4] On this aspect, see: MARCO OROFINO, "Diritto alla protezione dei dati personali e sicurezza: osservazioni critiche su una presunta contrapposizione", *MediaLaws-Rivista di diritto dei media* 2 (2018): p. 93; BART VAN DER SLOOT, "Privacy as Personality Right: Why the ECtHR's Focus on Ulterior Interests Might Prove Indispensable in the Age of Big Data." *Utrecht Journal of International and European Law*, 25 (2015): 25-50, doi: 10.5334/u ijel.cp; GIORGIO RESTA, "The new frontiers of personality rights and the problem of commodification: European and comparative perspectives." *Tulane European and Civil Law Forum*, 26 (2011): 33-66.

fundamental rights and freedoms, in particular their right to the protection of personal data"[5].

As a consequence, the role that a given individual plays for the purposes of the protection granted by the European lawmaker doesn't constitute a relevant factor. However, the different roles played by the various actors in a specific scenario are necessary to determine the allocation of rights, obligations and responsibilities related to the European data protection rules.

As previously investigated[6], the individuals involved in science are many and even more in the Open Scientific Research Process. They are the scientific community; the entities funding research projects[7]; the scientific publishers; the private sector services providers; eventually,

---

[5] Recital 2, GDPR; in addition, consider, also, that the Article 3 of the GDPR, entitled "Territorial scope", states: "This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not". In this way, the scope of the GDPR extends not only to any individual, but also beyond the geographical borders of the European Union. On the scope of application of the GDPR, see: CHRISTOPHER KUNER, "Territorial Scope and Data Transfer Rules in the GDPR: Realising the EU's Ambition of Borderless Data Protection." *University of Cambridge Faculty of Law Research Paper,* 20 (2021), 1-35; and from an economic perspective, see: GREGORY W. VOSS, "Cross-Border Data Flows, the GDPR, and Data Governance." *Washington International Law Journal*, 29.3 (2020): 485-532.

[6] See: Section 2.4.3, in particular the analysis of the third aspect, "Actors".

[7] Scientific research necessarily needs an input, represented by economic resources able to cover the expenses, of various kinds, related to the practical implementation of research projects. Looking at the Open Science phenomenon and referring to our LoA "Open Scientific Research Process", the first Observable O_1, "Resources", has in fact been further specified in its two components, represented by the variables $V_{1.1}$, "Research Data" and $V_{1.2}$ "Economic Resources". The entities funding research are connected to this variable $V_{1.2}$, since they represent the fundamental trigger of scientific research, without which research cannot be materially conducted. Entities that fund scientific research can be either public or private. There are several dimensions to public research funding: local, national, European, or international grants. As shown by the Royal Society, according to estimates made by the League of European Research Universities (LERU), "[…]15% of publicly funded research conducted in EU Member States comes from, or is coordinated by, the EU or by intergovernmental organisations" in Royal Society, *UK research and the European Union. The role of the EU in funding UK research*. (London: The Royal Society, 2015): p. 4. In addition, as already described in Section 2.2.6 and in Section 4.2.4.1, the last framework programme of European research funding, i.e., Horizon Europe programme, allocated EUR 95.5 billion to scientific research, for the four-year period 2022-2027. This represents an upward trend of 30% compared to the first framework programme launched in 1994. On this aspect, see: QUIRIN SCHIERMEIER, "How Europe's€ 100-billion science fund will shape 7 years of research." *Nature* 591.7848 (2021): 20-21, doi: 10.1038/d41586-021-00496-z.

also the institutions commissioning research projects or studies[8]; and, also, citizens and society.

As regards the personal data processing activities within the scientific research process, there are mainly three actors involved: scientific community; individuals and society; and private sector, both as service provider and also as scientific publisher.

The scientific community is identified as a group entity, consisting of researchers and students, both of whom are engaged in the scientific development and in the definition of the *connaissance scientifique*[9]. This group entity, i.e., the scientific community, is entrusted with a dual function: on the one hand, it generates new scientific knowledge on the basis of the continuous and uninterrupted dialogue between researchers, students and society; on the other hand, it is entitled to assess, judge and examine the scientific knowledge it generates[10]. In the Open Science scenario, triggered by the profound impact generated by the new digital technologies and ICTs, the scientific community maintains its central role, even widening its scope. ICTs, by enormously facilitating communication and human and professional relations, have driven the science community closer together and have greatly facilitated the dialogue of scientific knowledge. Moreover, new challenges posed by digital technology, which are common to the community of science (or to scientific sub-communities related to

---

[8] On the role of an effective deliberation, see: GIOVANNI BONIOLO, *The art of deliberating: democracy, deliberation and the life sciences between history and theory*. (Berlin, Heidelberg: Springer Science & Business Media, 2012): pp. 19-25, doi: 10.1007/978-3-642-31954-9. Starting from the main finding of this analysis on the art of deliberating, the issue of scientific research as a basis for public deliberation, has been investigated in: LUDOVICA PASERI, "COVID-19 Pandemic and GDPR: When Scientific Research becomes a Component of Public Deliberation", in DARA HALLINAN, RONALD LEENES, PAUL DE HERT, *Data Protection and Privacy: Enforcing Rights in a Changing World* (London: Hart Publishing, 2022): 165-167.

[9] GASTON BACHELARD, *Le rationalisme appliqué*. (Paris: Presses universitaires de France, 1949), pp. 124-142. On this matter, see also SABINO CASSESE, *Intellettuali*. (Bologna, Il Mulino: 2021), pp. 70-80, in which the Author problematises the new role of the intellectual as a member of the academic and scientific community in a society profoundly changed by the digital revolution.

[10] The role of peer review has already been discussed above, specifically in Section 4.2.3.2 on "Open by Default" Enforceability.

different research fields), have strengthened its centrality also from the governance point of view[11].

In light of this complex scenario, two kinds of processing activities are identified[12], described below: (i) the scientific community processes personal data of citizens and individuals involved in research projects; (ii) personal data of members of the scientific community are processed by the private sector involved in the research activities, through the use of ICTs and digital technologies.

5.1.1 Individuals as Data Subjects

Individuals and society may be involved in the Open Scientific Research Process from multiple points of view[13]. As regards data protection, personal data of individuals may be processed for what the GDPR defines as "archiving purposes in the public interest, scientific or historical research purposes or statistical purposes"[14]: in this case, the processing of individuals' personal data is part of the research projects. The university (or the research center) plays the role of

---

[11] The role and relevance of the scientific community in the Open Science scenario also – and above all – from the point of view of research governance, has already been investigated earlier, in particular in Section 3.4.1. In that instance, the aim was to underline the relevance of the scientific community in the governance of Open Science, in understanding the transition between the human right to science towards the practical implementation of Open Science.

[12] Here, the two major types of personal data processing activities have been identified, which are closely related to the practice of research. It should be noted, however, that further and multiple personal data processing activities might be carried out by universities and research centres, e.g., related to students and teaching, or in relation to employees of the university or research centre, in the employee-employer dynamic. For an analysis of the first case, related to students' personal data, see: MAURO ALOVISIO, ELEONORA BASSI, "Protezione dei dati personali e riutilizzo dell'informazione del settore pubblico", in MARCO RICOLFI, CRISTINA SAPPA, *Quaderni del Dipartimento di Giurisprudenza dell'Università di Torino*, (Napoli: Edizioni Scientifiche Italiane, 2013): 193-246, especially, see Section 6 of this paper, about the case study on the reuse of personal data of students, conducted by the university of Turin, pp. 241-246.

[13] Consider here the phenomenon of so-called "Citizen Science" and the multiple roles that can be played by society and citizens in research. The phenomenon of Citizen Science is described in Section 3.1.3 "Evolution: Shaping the Open Science". In a Citizen Science context, citizens are envisaged as part of the scientific research process from many and different points of view. For example, in data collection; in data analysis; in the dissemination phase; etc.

[14] The Article 89, GDPR.

controller, as the definer of the means and purposes of processing[15]; whereas the individuals whose data are processed within the various research projects are defined as data subjects, pursuant to the Article 4(1) of the GDPR[16]. This typology of processing activities deserves to be taken into account in an Open Science scenario, aiming to open up every stage of the scientific research cycle.

5.1.2 Members of the Scientific Community as Data Subjects

The role played by the private sector in the scientific research today seems crucial. The participation of the private sector in the scientific research process is not a novelty of the Open Science paradigm. Traditionally, the private sector has been involved in science in various ways: since the mid-20th century, with the emergence of the business model of scientific publishers[17]; or in the relationship between the benefits deriving from the scientific development and market repercussions[18]; as well as in the role of a research trigger, i.e., private actors as project funders.

In the context of Open Science, however, the role of private sector in scientific research is changing. As already mentioned[19], Open Science is characterised by the pervasive use of ICTs and new technologies at

[15] According to the Article 4(7), GDPR: "'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data".

[16] The definition of "data subject" can be indirectly derived from the Article 4(1) of the GDPR, where the definition of personal data is provided: "'personal data' means any information relating to an identified or identifiable natural person ('data subject')".

[17] AILEEN FYFE, "The production, circulation, consumption and ownership of scientific knowledge: historical perspectives", *CREATe Working Paper*, 4 (2020): pp. 17-21.

[18] On the relationship between science and market, see *ex multis*: RICHARD R. NELSON, "The market economy, and the scientific commons." *Research policy* 33.3 (2004): 455-471, doi: 10.1016/j.respol.2003.09.008. The debate on this relationship is longstanding, considering that already in 1969, questions were raised about the link between the innovation process and scientific research, stating that: "It is also essential that we use this understanding to ensure that society obtains maximum benefits from scientific research and that the community benefits from these growing relationship.", in WILLIAM J. PRICE, LAWRENCE W. BASS, "Scientific Research and the Innovative Process: The dialogue between science and technology plays an important, but usually nonlinear, role in innovation." *Science* 164.3881 (1969): p. 806, doi: 10.1126/science.164.3881.802.

[19] See: Section 4.3.1 and, also, Section 4.2.2.

every stage of the scientific process: this has generated a deep intermingling of public and private sectors in science[20]. This mixture often makes it difficult to identify clear and well-defined boundaries of areas of relevance. The EDPS recently described this relationship as "[…] the blurring of the boundaries between public interest, academic freedom and private gain"[21].

Concerning data protection, the role of the private actors has experienced a significant transformation: the traditional scientific publishers changed their core activity; and, in addition, new private actors emerged in science, represented by those who provide IT services.

Several scientific publishers are changing their main activity from disseminators of research outputs to analysts of user data. A recent study provided by the *Deutsche Forschungsgemeinschaft* (DFG, the German Research Foundation, namely the leading national research funding entity in Germany) illustrated that "[…] the major academic publishers have been fundamentally changing their business model with significant implications for research: aggregation and the reuse or resale of user traces have become relevant aspects of their business"[22].

---

[20] This trend is even more evident in the health research domain. "The research environment emerging under the pressure of technological change and the growing employment of algorithmic processing techniques in health research is very complex, characterized by a deep interconnection between traditional research centres and large high-tech corporations.": see, GIULIA SCHNEIDER, "Disentangling health data networks: a critical analysis of Articles 9(2) and 89 GDPR." *International Data Privacy Law*, 9.4 (2019: 270, doi: 10.1093/idpl/ipz015. This aspect will be further discussed in Section 6.3.

[21] EDPS, *A Preliminary Opinion on data protection and scientific research*, January 2020, p. 27 available at: https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en.

[22] Committee on Scientific Library Services and Information Systems of the Deutsche Forschungsgemeinschaft (DFG) – German Research Foundation, *Data tracking in research: aggregation and use or sale of usage data by academic publishers*, 20 May 2021, p. 3, available at: https://www.dfg.de/en/research_funding/programmes/infrastructure/lis/index.html. Similarly, see, CLAUDIO ASPESI, *et al.*, "Landscape Analysis: The Changing Academic Publishing Industry – Implications for Academic Institutions." *SPARC* (2019): p. 1-53. The Authors illustrate that: "Elsevier, Pearson and Cengage in particular are transforming themselves into data analytics companies built atop their content, effectively adding ways to monetize it. […] the management teams of these companies clearly view the future as driven by adding the provision of data and data analytics services to their respective customers, rather than by growing only the traditional core business.", p. 7.

This transformation is profound to the extent that "[S]ome publishers now explicitly regard themselves as information analysis specialists. Their business model is shifting from content provision to data analytics. This involves the tracking – i.e. recording and storage – of the usage data generated by researchers (i.e. personalised profiles, access and usage data, time spent using information sources, etc.) when they utilise information services such as when carrying out literature searches"[23]. In this context, private companies that carry out such data analysis activities are the controllers of such processing activities; conversely, here, members of the scientific community, who daily make use of such services in their work, should be considered as data subjects. New data protection challenges arise in this specific type of processing activities[24]. Moreover, this situation arises not only data protection challenges. Tim Berners-Lee, the founder of the Web, two decades ago, warned against the risk of search engine selection of information. Specifically, Berners-Lee advocated the need to define them as "impartial boxes", stating that "[I]f a company claims to give access to the world of information, and then presents a filtered view, the Web loses credibility"[25]. This risk is even greater when the access to be granted is to scientific knowledge[26].

---

[23] DFG, *"Data tracking in research", op. cit.*, p. 3.
[24] To some extent, this issue is related to the protection of group privacy, which is an aspect extremely sensitive, even more in this period characterised by the COVID-19 pandemic, as argued in: MARIAROSARIA TADDEO, "The Ethical Governance of the Digital During and After the COVID-19 Pandemic." *Minds and Machines* 30 (2020): 171-176, doi: 10.1007/s11023-020-09528-5. Here, the Author refers to the risk related to the adoption of digital technologies tracking the spread of the virus and the related risk of mass surveillance. However, the reasoning about the privacy group is relevant for the topic under investigation: "The protection of group privacy is crucial in the age of big data and artificial intelligence, where data collection is often finalised to identify categories, groups, of individuals rather than to single out a specific person. Consider commercial profiling, for example, it rests on the identification of groups (e.g. those who like Amarone or rock-folk music, those who live in the UK), and is independent from the identification of a specific data subject (e.g. Mariarosaria)", p. 173.
[25] TIM BERNERS-LEE, *Weaving the Web. The original design and ultimate destiny of the World Wide Web.* (New York: HarperBusiness, 2000), p. 132.
[26] On the risk related to the managing of access to information, see: MARIAROSARIA TADDEO, "The Civic Role of OSPs in Mature Information Societies.", SSRN (2020), doi: 10.2139/ssrn.3584187.

Although this second kind of processing activities, realised by private actors, certainly needs further analysis, in this dissertation the attention is on the first type of processing activities, namely the one carried out by the scientific community that processes personal data of individuals for scientific research purposes[27].

In order to identify the main data protection issues related to this kind of processing activity, next section casts light on the complex discipline provided for processing activities for research purposes, according to the GDPR.

5.2 Data Protection Discipline for Scientific Research

The intention is now to focus on the type of Open Science personal data processing activity in which personal data of individuals are processed for scientific research purposes, by researchers, both public and private, in carrying out research projects.

In order to investigate the challenges that this type of processing activity arises, it is essential to begin with the analysis of the discipline provided by the GDPR (Section 5.2.1). The kernel of the discipline is represented by the Article 89 of the GDPR which, therefore, represents our starting point. Then, attention is drawn to the national dimension of the protection of personal data processed for scientific research

[27] The expression "research purposes" is adopted for the sake of clarity. It should be considered, however, that whenever reference is made to the purposes of scientific research, it is always meant to refer implicitly to the "processing for archiving purposes in the public interest, scientific or historical research purposes", according to the GDPR. Basically, the aim is to focus on research in a broad sense, excluding only the processing activities for statistical purposes, due to the partially different nature. On the distinction between "archiving", "research" and "statistical proposes", see: CHRISTOPHER KUNER, CHRISTOPHER DOCKSEY, LEE BYGRAVE, *The EU General Data Protection Regulation: A Commentary* (Oxford: Oxford University Press, 2020), pp. ff. 1243; for a further analysis on this distinction, see: ROSSANA DUCATO, "Data protection, scientific research, and the role of information." *Computer Law & Security Review* 37 (2020): pp. 2-4, doi: 10.1016/j.clsr.2020.105412.

purposes, given the wide leeway enjoyed by the Member States (Section 5.2.2).

5.2.1 GDPR Discipline for Scientific Research

The Article 89 of the GDPR represents the core of the discipline of data protection in the field of scientific research and is entitled "Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes". Hence, the Article 89 takes into account different purposes: (i) archiving; (ii) research, both scientific and historical; and, finally, (iii) statistics.

In order to comprehend what is meant by "scientific research" according to the GDPR, it is necessary to refer to Recital 159:

> For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research. In addition, it should take into account the Union's objective under Article 179(1) TFEU of achieving a European Research Area. Scientific research purposes should also include studies conducted in the public interest in the area of public health.

From the wording of Recital 159, some considerations can be drawn. First, a very broad and widely encompassing interpretation of the notion of scientific research clearly emerges in the context of European data protection, considering that there is an explicit reference to privately funded research. This is particularly relevant given the fact that the GDPR entails a derogatory and more flexible regime for the processing of personal data for research purposes: "[…] it truly is an advantage, legally speaking, to have certain personal data processing formally labelled as scientific research"[28].

[28] CECILIA MAGNUSSON SJÖBERG, "Scientific research and academic e-learning in light of the EU's legal framework for data protection." in MARCELO CORRALE, MARK FENWICK, NIKOLAUS FORGÒ (eds.), *New*

Second, another significant aspect is the mention to the Article 179 TFEU[29] and to the European Research Area (ERA). This reference represents the *ratio* behind the derogation accorded to the processing of personal data for scientific research purposes, compared to the other purposes of the processing. The European Union intends to encourage as much as possible the circulation of knowledge and researchers on the territory of the European Union. This objective is enshrined in the Treaties, the source of law underpinning the European Union, and has been implemented over the years. In Chapter 2, investigating the development of the European projects in the field of scientific research, the focus was on the establishment of the fifth European freedom, the so-called "free movement of knowledge"[30], until the reform of the ERA in 2020[31].

The Article 1 of the GDPR, in its second paragraph supports the objective of protecting "[…] fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data". But, then, the Article 1, paragraph 3, states: "The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data". Therefore, the Article 1 of the GDPR expresses the intention of the European lawmaker to abide by policies on the circulation of knowledge that have already been implemented for years, despite pursuing the main objective of regulating the processing of personal data[32].

Technology, Big Data and the Law (Singapore, Springer: 2017): 43-63, doi: 10.1007/978-981-10-5038-1_3.

[29] Consolidated version of the Treaty on the Functioning of the European Union (TFEU), OJ C 326, 26.10.2012, p. 47–390, http://data.europa.eu/eli/treaty/tfeu_2012/oj.

[30] See: Section 2.2.1.

[31] See: Section 2.2.5.

[32] Here the first elements supporting the convergence – even only in principle – between Open Science and data protection, start to emerge, further discussed in Section 5.4. On the different perspectives behind the wording of the Article 1 of the GDPR, see: "Paragraphs 2 and 3 define the purposes of the Regulation. Paragraph 2 explicitly identifies, as an objective of the Regulation, the protection of the fundamental rights and freedoms of natural persons, with particular reference to the right to protection

Finally, a third consideration concerns the concept of "public interest". Public interest is one of the legal bases, pursuant to the Article 6(1)*e* of the GDPR. Both, Recital 159, and the Article 89 outline an explicit link between the processing of personal data for research purposes and the legal basis of public interest. Exploring the origins of Open Science in Modern Science, around the 17th century, the previous Section 3.1.1 underlined how the concept of public interest has gained centrality in this field already since that age[33]. However, the role of public interest in the context of research is still controversial to some extent. For this reason, it is further investigated later[34].

The GDPR provides for a derogatory regime for the processing of personal data for research purposes. Yet, the exceptions provided for in this context must be subject to a condition, represented by the fact that "[P]rocessing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards […] for the rights and freedoms of the data subject"[35]. Compliance with appropriate safeguards is identified through the adoption of technical and organisational measures, such as pseudonymisation, specifically mentioned in the Article 89(1) of the GDPR.

---

of personal data. This paragraph conveys a conception of law as a means to affirm and protect a human person in some fundamental respect, that is, with regard to what is situated at the foundation of the construction of personality. Paragraph 3 states that the free circulation of personal data cannot be limited or prohibited for reasons relating to the protection of natural persons with regard to the processing of their personal data. This paragraph conveys a different conception of the law: here the law is intended to build a framework of legal certainty within which to develop economic investments, innovations, and business, or in other words, the digital single market.", in MASSIMO DURANTE, *Computational Power: The Impact of ICT on Law, Society and Knowledge.* (New York: Routledge, 2021), p. 130. Considering the Article 1 of the GDPR, what the Author identifies as "dialectical tension between natural persons and the free market." (p. 130) here is a dialectic between the natural person and free movement of knowledge: it is fair to admit that the tension perceived between the two elements in the first interpretation does not have the same force in relation to the scenario under investigation in this dissertation.

[33] Section 3.1.1, in particular "(i) Publicity".

[34] Later on, the issue of the legal bases for processing will be further investigated, see: Section 5.3.1

[35] The Article 89(1), GDPR.

Without prejudice to these "appropriate safeguards", two groups of exceptions are laid down for the processing of personal data for research purposes: the exceptions set out directly in the GDPR; and the exceptions that the GDPR provides may be established by Member States.

Regarding the first group of exceptions, i.e., those specifically provided for in the GDPR, it is necessary to take into account a set of provisions, embodied in several articles, concerning: (i) purpose limitation; (ii) storage limitation; (iii) special categories of personal data; (iv) duty of communication to data subject; (v) right to erasure; (vi) right to object.

(i) Purpose limitation. The Article 5(1)*b* deals with the principle of purpose limitation. It provides that personal data shall be "[…] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". This article, however, provides that "further processing" for research purposes should not be considered incompatible with the original purpose. By establishing this exemption, the GDPR provides *de facto* for a so-called "secondary use" of personal data where it is carried out for scientific research purposes.

(ii) Storage limitation. Then, the Article 5(1)*e* lays down the principle of data retention limitation, which aims to ensure that personal data of data subjects are only kept as long as necessary for a specific processing activity. Regarding processing for research purposes, it provides that "[…] personal data may be stored for longer periods"[36]. However, the GDPR doesn't specify the expression "longer periods"[37].

[36] The Article 5(1)*e*, GDPR.

[37] In addition, also national authorities generally don't set a defined period for data retention. For example, the Information Commissioner's Office (ICO), i.e., the UK national authority established: "The UK GDPR does not dictate how long you should keep personal data. It is up to you to justify this, based on your purposes for processing. You are in the best position to judge how long you need it.", see: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-

(iii) Special categories of personal data. The Article 9 of the GDPR, in its first paragraph, provides for a general prohibition on the processing of special categories of personal data, namely those relating to "[…] revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation". These special categories of personal data were defined by the previous EU Data protection discipline[38] as "sensitive data". The second paragraph of the Article 9 of the GDPR, however, provides for a list of possible instances in which the processing of such special categories of data is permitted: letter *j* of this list specifically provides that the prohibition of paragraph 1 does not apply to processing for research purposes. The European lawmaker, however, stresses that this exempting processing in case of research purposes "[…] shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject"[39].

(iv) Duty of communication to data subject. The Article 14 of the GDPR provides for a list of information that must necessarily be communicated to the data subject, in case of processing of personal data concerning her, that the controller has received not from the data

regulation-gdpr/principles/storage-limitation/#retention_periods. The ICO refers to the UK GDPR, i.e., the UK Data Protection Act, adopted in 2018, https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted. However, the issue of data retention is addressed in the same way, so the ICO's considerations also apply to the European GDPR. The identification of the period of data retention should be an expression of the accountability of the controller. Regarding the research domain, it may be fruitful to identify some rules in an internal policy.

[38] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995, o*n the protection of individuals with regard to the processing of personal data and on the free movement of such data*, OJ L 281, 23.11.1995, p. 31–50, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046.

[39] The Article 9(2)*j*, GDPR.

subject herself, but from third parties. The fifth paragraph, however, excludes the application of this mandatory requirement in the case of processing of personal data for research purposes where "[…] the provision of such information proves impossible or would involve a disproportionate effort", or when "[…] the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing"[40].

(v) Right to erasure. The Article 17 of the GDPR endows the data subject with the right to have her personal data erased by the controller, in specific circumstances. However, the third paragraph, letter *d* provides that the right to erasure may not be guaranteed for processing operations that pursue research purposes "[…] in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing"[41].

(vi) Right to object. The Article 21 of the GDPR enshrines the right of the data subject to object to the processing of personal data concerning her. Yet, paragraph 6 states that: "[W]here personal data are processed for scientific or historical research purposes or statistical purposes pursuant to Article 89(1), the data subject, on grounds relating to his or her particular situation, shall have the right to object to processing of personal data concerning him or her, unless the processing is necessary for the performance of a task carried out for reasons of public interest"[42]. This condition introduced by the expression "unless" allows for the possibility to derogate from the right to object, at least in so far as the processing of personal data for research purposes is based on the legal ground of public interest[43].

---

[40] The Article 14(5)*b*, GDPR.
[41] The Article 17(3)*d*, GDPR.
[42] The Article 21(6), GDPR.
[43] This aspect is further clarified in Section 5.3.1.

Table 5.1: Exceptions for Research Purposes Established by the GDPR

Art.	Title	Contents
89	**Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes**	Member States may provide for certain derogations.
5(1)b	**Purpose limitation**	Authorisation for "further processing" in case of processing activities for research purposes
5(1)e	**Storage limitation**	Longer data retention for processing for research purposes
9(2)j	**Processing of special categories of personal data**	Option to process particular categories of personal data for research purposes
14(5)b	**Duty to provide information about processing**	Exemption from the duty of communication to data subjects, if data (not directly received) are processed for research purposes
17(3)d	**Right to erasure**	Option not to apply the right to erasure for data processed for research purposes
21(6)	**Right to object**	Option to reject the right to object for data processed on a lawful basis in the public interest (e.g. for scientific research purposes)

The Table 5.1 summarises the derogations provided directly by the GDPR for processing of personal data for scientific research purposes. The second type of exceptions foreseen for this specific area is left to Member States, directly by the Article 89 of the GDPR. Next section deals with this leeway granted to Member States, since, in the domain of research, raises some problematic considerations.

5.2.2 Member States' Leeway

In addition to the exemptions expressly provided for by the GDPR for the processing of personal data for research purposes, EU law and Member States may define further exceptional rules for this type of processing. Such provision is set out in the Article 89 of the GDPR, in paragraphs 2 and 3, which states[44]:

[44] The Article 89(2) and (3), GDPR.

> 2. Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.
>
> 3. Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.

Thus, Union or Member States law may provide for an exceptional regime derogating from certain provisions of the GDPR where compliance would irreparably undermine the achievement of the identified purposes: "More specifically, the GDPR establishes a 'three-step-test' for research derogations, centred on necessity and proportionality. To verify whether there are legitimate grounds for the introduction of exceptions to data subjects' rights the following elements must be present cumulatively. First, exercising the rights is likely to render impossible or seriously impair the achievement of scientific purposes. Second, the derogations must be necessary for the fulfilment of those purposes. Finally, appropriate safeguards for the rights and freedoms of the data subject must be adopted"[45].

Regarding research (scientific or historical) and statistical purposes, a potential derogation is permitted for some of the rights that the

[45] ROSSANA DUCATO, "Data protection, scientific research, and the role of information.", *op. cit.*, p. 6. The Author emphasies the relevance of the third step, referring to the following study: CIARA STAUNTON, SANTA SLOKENBERGA, DEBORAH MASCALZONI, "The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks." *European Journal of Human Genetics* 27 (2019): 1159-1167, doi: 10.1038/s41431-019-0386-5.

GDPR guarantees to the data subject: the Article 15 “Right of access by the data subject”; the Article 16 “Right to rectification”; the Article 18 “Right to restriction of processing”; the Article 21 “Right of object”[46].

Regarding archival purposes, the provisions for which a potential exemption is granted are the same as for research and statistical purposes, with the addition of two further provisions: the Article 19 “Notification obligation regarding rectification or erasure of personal data or restriction of processing”; and the Article 20 “Right to data portability”.

Table 5.2: Exceptions for Research Purposes

Art.	Title	Contents	Purposes
Exceptions directly imposed by the GDPR			
5(1)b	**Purpose limitation**	Authorisation for “further processing” in case of processing activities for research purposes	For scientific or historical research purposes; statistical purposes; or archiving purposes
5(1)e	**Storage limitation**	Longer data retention for processing for research purposes	
9(2)j	**Processing of special categories of personal data**	Option to process particular categories of personal data for research purposes	
14(5)b	**Duty to provide information about processing**	Exemption from the duty of communication to data subjects, if data (not directly received) are processed for research purposes	
17(3)d	**Right to erasure**	Option not to apply the right to erasure for data processed for research purposes	
21(6)	**Right to object**	Option to reject the right to object for data processed on a lawful basis in the public interest (e.g. for scientific research purposes)	
Exceptions imposable by Union or Member States			
15	Right of access by the data subject	Condition: if the observance of these rights is likely to render impossible or seriously impair the achievement of the specific purposes and and such derogations are necessary for the fulfilment of those purposes	For scientific or historical research purposes; statistical purposes; or archiving purposes
16	Right to rectification		
18	Right to restriction of processing		
21	Right of object		
19	Notification obligation regarding rectification or erasure of personal data or restriction of processing	Condition: if the observance of these rights is likely to render impossible or seriously impair the achievement of the specific purposes and and such derogations are necessary for the fulfilment of those purposes	For statistical purposes
20	Right to data portability		

[46] Although, regarding the “right of object”, the GDPR itself in paragraph 6 already provides for an exemption related to processing carried out on the lawful basis of public interest. See, Section 5.2.1 and Table 5.1.

With these provisions, the GDPR extends the derogatory legal framework for scientific research: adopting a teleological interpretation, the underlying *ratio* seems represented by the intention to provide a flexible legislative framework to adapt to the needs of research.

Therefore, the table summarising the exceptions envisaged for personal data processing activities conducted for research purposes (including scientific, historical, archival, and statistical research) is updated, as shown in Table 5.2.

To some extent, the GDPR intends to delegate to the national level the regime of data processing for research purposes. The reason should be found in the distribution of competences between the European Union and the Member States, in the various matters: the European Union has the competence to establish provisions relating to the protection of individuals with regard to the processing of personal data in the exercise of activities falling within the scope of the EU law, pursuant to the Article 16(2) of the TFEU, confirmed in the Article 2(2) of the GDPR itself, dedicated to the definition of the material scope of the Regulation. Concerning specifically the field of research, in fact, the Article 6 of the TFEU and the Title XIX of the same Treaty, entitled "Research and Technological Development and Space", underline the European Union's role of support, coordination and consolidation of the Member States' actions in this field. The Article 4 of the TFEU, which deals with matters of shared competence between the Union and Member States, establishes, in paragraph 3: "In the areas of research, technological development and space, the Union shall have competence to carry out activities, in particular to define and implement programmes; however, the exercise of that competence shall not result in Member States being prevented from exercising theirs"[47]. With the limitation of guaranteeing a competence to the Member States in this

[47] The Article 4(3), TFEU.

matter (which is, however, essential), the TFEU itself seems to recognise a fair leeway for the Union in this area[48].

In this dissertation, the issue of the EU competence in the field of scientific research has already been dealt with on several occasions, in connection with different topics. In Section 2.1.2, investigating one of the two premises underlying this dissertation, namely the institutional one, the growing interest and involvement of European institutions in supporting Open Science was described. Thus, in explaining the intention to adopt the European perspective in this dissertation, the coordinating role of the Union was emphasised. However, a particular ongoing trend was also stressed: the Member States, over the years, tended to cautiously overcome a traditional reluctance to grant competences in scientific research to the Union, precisely because of the inherent flexibility of the Union treaties.

Then, in Section 4.2.3.1, this flexibility was again mentioned when assessing the suitability of the choice of the legal instrument of the directive for the EU Open Data framework. There, it was argued that although this legal instrument appears to be suitable for fulfilling the needs of sharing and re-use of public sector data, it does not seem to be equally suitable for sharing and re-use of research data. It was pointed out that without compromising the distribution of competences between the Union and the Member States, the former may benefit of a fair leeway of flexibility.

Without dwelling too much on the general issue of the distribution of competences between the Union and the Member States, keep in mind that the European Union potentially has 27 different regimes for the protection of personal data for research purposes, with different

[48] The intention to move towards "a Community policy in this area" was already expressed at the 1972 Paris summit, which "[…] outlined the broadening of the Community's competences in scientific research, stating that all the possibilities offered by the [*then*, ed.] EEC Treaty should be used", see: FAUSTO POCAR, *Commentario Breve ai Trattati dell'Unione europea*, (Padova: Cedam, 2001), p. 163. [Translation from the Italian original text].

obligations and requirements, for a field – that of knowledge – which by nature transcends geographical borders.

The relevance of the constraints resulting from this fragmentary landscape needs further investigation and is further analysed later[49].

Now, attention is drawn to issues that have traditionally been considered problematic for personal data processing in the context of scientific research.

5.3 Data Protection Issues in the Research Field

The implementation of the GDPR "[…] has caused significant concern within the research community"[50]. Taking into account the legal framework and the relevant literature, three main challenges have been identified in the process of compliance for the research activities with the GDPR[51], which are further investigated below: (i) the identification of the lawful legal basis for processing activities for research purposes (Section 5.3.1); (ii) the compliance with the principle of data minimisation and the subsequent problems related to the anonymisation techniques (Section 5.3.2); (iii) and the challenges posed

[49] As argued in MARCELLO IENCA, *et al.*, "How the general data protection regulation changes the rules for scientific research." *European Parliamentary Research Service (EPRS), Scientific Foresight Unit (STOA)* (2019), p. 1, doi: 10.2861/17421, in the field of scientific research, there is a shared and deep perception of risk towards such fragmentation: "Research-based organisations have also expressed concerns regarding the potential risk of fragmentation deriving from the possibility of Member States' derogations. These derogations may establish uneven conditions for researchers and pose challenges for research collaboration between Member States, and globally." For this reason, see: Section 5.5.1. In addition, consider that during the drafting of the GDPR (before it come into force), a considerable debate among scholars was developed, on the discipline of data protection in the research domain. For instance, see: MARTINE CORRETTE PLOEM, MARIE-LOUISE ESSINK-BOT, KARIEN STRONKS, "Proposed EU data protection regulation is a threat to medical research." *Bmj Clinical Research* 346 (2013), doi: 10.1136/bmj.f3534.

[50] MARCELLO IENCA, *et al.*, *"How the general data protection regulation changes the rules for scientific research", op. cit.*, p. 1.

[51] During the complex process of compliance with the GDPR, multiple and different difficulties may arise for every research project, and for every entity conducting the research. Here, the aim is to underline the most tangled challenges from a legal point of view, summarised in the three that are investigated below. Yet, in Chapter 6, dealing with the case study represented by the ULHPC, more aspects on the compliance process are detailed.

by the transfer of personal data to third countries and international organisations (Section 5.3.3).

5.3.1 Legal Bases

The definition of the correct legal basis legitimising the processing of personal data for scientific research purposes is often considered problematic to identify.

As stated in the Article 6 of the GDPR, any processing of personal data shall be carried out in light of a legal basis, among those expressly listed: a) consent; b) performance of a contract; c) compliance with legal obligations; d) vital interests of the data subject or of another natural person; e) performance of a task carried out in the public interest; f) legitimate interests[52].

Concerning the identification of the legal basis for processing activities conducted for research purposes, a preliminary distinction is due. There are two uses of personal data processed for research purposes: the primary use of personal data for research, i.e., data collected and processed *ab origine* for research purposes; and the so-called secondary use, in which the research purpose is the secondary and additional purpose to data already collected and processed for other purposes[53].

The secondary use of personal data for research purposes, where research purposes are identified as additional and secondary, was already covered by Recital 29 of Directive 95/46/EC[54]. The GDPR establishes a real "presumption of compatibility"[55], by rendering incompatibility the exception, stating in the Article 5(1)*b* that: "further

---

[52] The Article 6, GDPR.

[53] The concept of secondary use was already mentioned in Section 4.4.1, defining "research data".

[54] Directive 95/46/EC, Recital 29 stated: "Whereas the further processing of personal data for historical, statistical or scientific purposes is not generally to be considered incompatible with the purposes for which the data have previously been collected provided that Member States furnish suitable safeguards; whereas these safeguards must in particular rule out the use of the data in support of measures or decisions regarding any particular individual".

[55] Giovanni Maria Riccio, Guido Scorza, Ernesto Belisario, *GDPR e normativa privacy* (Milano: Wolters Kluwer, 2018): 56-57.

processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes". In other words, the so-called secondary use of personal data for research purposes is only subject to a compatibility test[56].

In addition, the European Data Protection Board (EDPB), in the Opinion 3/2019, clarified that "[…] the controller could be able, under certain conditions, to further process the data without the need for a new legal basis"[57]: this means that it is not needed to *always* find a new legal basis concerning the secondary use. The problem is rather the identification of these "certain conditions" referred to by the EDPB. Similarly as for the compatibility test, the definition of the conditions that allow processing personal data as secondary use without changing the legal basis are part of a case-by-case evaluation to be carried out by the controller[58].

Regarding the processing of personal data conducted for research purposes, there are three legal bases that can be adopted: (i) consent; (ii) public interest; or (iii) legitimate interest.

---

[56] "When the secondary processing is not based neither on the data subject's consent or a Union or Member State law, the controller can still further process the personal data, but it needs to perform a purpose compatibility test. This test is a novel tool in the GDPR which helps to identify the crucial aspects of the processing to decide whether the new purpose is compatible with the original one. According to the test, when a controller willing to reuse the data will have to consider: – any link between the original and new purposes; – the context in which the personal data have been collected; the sensitivity of the personal data; – the possible consequences of the intended further processing for data subjects; – the existence of appropriate safeguards, which may include encryption or pseudonymisation." in JÁNOS MÉSZÁROS, CHIH-HSING HO, "Big data and scientific research: the secondary use of personal data under the research exemption in the GDPR." *Hungarian Journal of Legal Studies* 59.4 (2018): p. 406, doi: 10.1556/2052.2018.59.4.5.

[57] EDPB, *Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (art.70.1.b)*, 2019, p. 8, available at: https://edpb.europa.eu/our-work-tools/our-documents/opinion-art-70/opinion-32019-concerning-questions-and-answers_en.

[58] Once again, the accountability principle is at stake, and once again its role as a meta-principle emerges. On this aspect see Section 5.2.4.

(i) Consent. Processing may be carried out on the basis of the consent given by the data subjects, as provided for in the Article 6(1)*a* of the GDPR. The consent shall take the form of a "[…] freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her" pursuant to the Article 4 of the GDPR[59]. For this reason, it is not always the most feasible choice. In the research domain, frequently it is complex to identify precisely, *a priori*, at the time of collection, the specific purpose of the data processing. Therefore, as Recital 33 further specifies, consent-based processing of data for research purposes might be limited only to certain areas of research, in accordance with the relevant ethical provisions, stating that "[I]t is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose"[60]. This *modus operandi*, which seems to emerge from the GDPR itself, has been formalised by some scholars in terms of "broad consent"[61]. Without

---

[59] The Article 4(11), GDPR.

[60] Recital 33, GDPR.

[61] See, for example: DARA HALLINAN, "Broad consent under the GDPR: an optimistic perspective on a bright future." *Life Sciences, Society and Policy* 16.1 (2020): 1-18, doi: 10.1186/s40504-019-0096-3, where the Author claims that: "[...] broad consent allows each collected sample and associated data set to be used for multiple research purposes without the obligation to recontact research subjects to request new permissions for each new project – with the administrative and resource allocation this would require", p. 3. Similarly, EDPB, *Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research*, 2 February 2021, p. 8, available at: https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en, stated that: "[...] the GDPR cannot be interpreted to allow for a controller to navigate around the key principle of specifying purposes for which consent of the data subject is asked. Therefore, when research purposes cannot be fully specified, a controller must seek other ways to ensure the essence of the consent requirements are

bypassing the pillars of the discipline of consent, it is, however, necessary to adapt the discipline to the peculiarities of the research process, always bearing in mind the *ratio* of protecting the rights and freedoms of individuals.

(ii) Public interest. In the instance that the processing is carried out by public research bodies, the legal basis of the processing should generally be found in the Article 6(1)*e* of the GDPR, which identifies the lawfulness of processing if it is "[...] necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller"[62]. The limit of this legal basis is that "[...] it is only available where there is specific (European) Union or Member State law available [...]. This usually means that legislation must exist that identifies the controller in question as being able to carry out the type of processing in question"[63].

(iii) Legitimate interests. Finally, in order to process personal data for research purposes, the legal basis of legitimate interests can also be adopted. The Article 6(1)*f* of the GDPR establishes that: "[...] processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child"[64]. The concept of legitimate interest

---

served best, for example, to allow data subjects to consent for a research purpose in more general terms and for specific stages of a research project that are already known to take place at the outset". These assumptions of the EDPB are interpreted as an endorsement to the "broad consent" for scientific research. The broad consensus approach could also be complemented by the so-called "dynamic consensus", see: MEGAN PRICTOR, *et al*. "Consent for data processing under the General Data Protection Regulation: Could 'dynamic consent' be a useful tool for researchers?." *Journal of Data Protection & Privacy* 3.1 (2019): 93-112.

[62] The Article 6(1)*e*, GDPR. Significantly, something similar is provided for the EU lawmaker in the Data Governance Act (DGA) analysed in Section 4.4. In the Article 2(10) of the DGA the European lawmaker specifically identifies scientific research as an example of an activity of general interest.

[63] PAUL QUINN, "Research under the GDPR – a level playing field for public and private sector research?" *Life Sciences, Society and Policy* 17.4 (2021): p. 9, doi: 10.1186/s40504-021-00111-z.

[64] The Article 6(1)f, GDPR.

is quite discussed among scholars[65]. Recital 47 lays down the two conditions that must be met in order to identify the controller's legitimate interest as a legal basis for personal data processing: first, a balancing exercise must be carried out between the data subject's fundamental rights and freedoms and the interests of the controller; second, there must be "[...] the reasonable expectations of data subjects based on their relationship with the controller"[66]. It can be inferred that the reasonable expectation of the data subject is grounded in the relationship between data subject – controller, as revealed by the expression "based on".

As regards the legal basis of legitimate interests, it is relevant the Opinion of the then Article 29 Data Protection Working Party (WP29, now EDPB), issued in 2014, concerning the concept of legitimate interest of the controller under the then applicable Directive 95/46/EC[67]. In this opinion, the WP29 expressly included the processing of data for research purposes in "[…] a non-exhaustive list of some of the most common contexts in which the issue of legitimate interest in the meaning of Article 7(f) may arise"[68]. This assumption, even to some extent controversial, seems to admit the adoption of the legitimate interests as legal basis for processing activities conducted for research purposes.

---

[65] Among others, see: IRENE KAMARA, PAUL DE HERT, "Understanding the balancing act behind the legitimate interest of the controller ground: A pragmatic approach." in EVAN SELINGER, JULES POLONETSKY, OMER TENE (eds.), *The Cambridge Handbook of Consumer Privacy* (Cambridge: Cambridge University Press, 2018): 321-352, doi: 10.1017/9781316831960.019.

[66] Recital 47, GDPR.

[67] WP29, *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*, 9 April 2014.

[68] WP29, *Opinion 6/2014*, *op. cit.*, p. 24. Consider that the Article 7(f) of the Directive 95/46/EC establishes: "[…] processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed", substantially corresponding to the Article 6(1)f of the GDPR.

After almost four years since the GDPR came into force, some considerations can be developed on the identification of the lawful legal basis for processing of personal data for research purposes.

Chiefly, the European institutions have been able to further specify their positions. For instance, it has been stressed, on several circumstances[69], that in the case of processing of personal data for public scientific research purposes, the legal basis of the public interest represents the most suitable choice. This legal basis is also (and perhaps above all) justified by teleological reasons. The *ratio* of the GDPR's derogatory regulation for research is the promotion of scientific research, without harnessing it with excessive bureaucratic burdens. In other words, the European lawmaker strikes a balance *ab origine* between the right to science and academic freedom and the right to the protection of personal data, by simplifying the conditions under which personal data are processed for research purposes. For this reason, the legal basis of public interest appears to be the most coherent with the aims pursued by the GDPR.

Conversely, adopting consent as the legal basis, even sometimes might be necessary[70], is constraining. Compared to the previous

---

[69] EDPB, *Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research*, 2 February 2021, available at: https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en; EDPB, *Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak*, 21 April 2020, available at: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032020-processing-data-concerning-health-purpose_en. Consider that further EDPB guidelines on the subject are expected, as explicitly announced by the EDPB: "Guidelines on the processing of personal data for scientific research purposes (currently in preparation and due in 2021)", in EDPB, *Document on response, op. cit.*, 2021, p. 7.

[70] For instance, the Article 9(2)a of the GDPR permits to collect particular categories of data if "[…] the data subject has given explicit consent to the processing of those personal data for one or more specified purposes", and this might occur as regards medical, genomic, biomedical and medical research. To the question "How is consent for processing in scientific research obtained?", the European institutions replied: "Some flexibility in relation to the degree of specification and granularity of consent is allowed in the context of scientific research. When collecting personal data, researchers might not be able to fully identify the purposes for processing. In those cases they can ask individuals to give consent for certain areas of scientific research or parts of research projects.", available at: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-

European discipline on data protection, the intention of the GDPR is to exempt the field of research from the traditional "notice and consent" model, which proved to be ineffective[71]. Not to mention that in certain circumstances it is even impossible to obtain consent from data subjects involved in the scientific process[72].

Similarly, some problems occur also with regard to the legal basis of legitimate interests. Consider that "[…] 'legitimate interest' in the GDPR refers only to interest of private sector controllers"[73]: according to Recital 47, in fact, the legal basis of legitimate interest "[…] should not apply to the processing by public authorities in the performance of their tasks"[74]. These considerations seem to reduce the applicability of the legal basis of legitimate interest, at least for publicly funded research.

To sum up, the problem of identifying the legal basis of processing for research purposes should be addressed by taking into account several aspects, which are summarised below.

- It is necessary to make a case-by-case analysis, keeping in mind the various specifics of each research project, in order to identify the appropriate legal basis (among other things, taking into careful consideration the nature of the actors involved, whether public or private[75]).

grounds-processing-data/grounds-processing/how-consent-processing-scientific-research-obtained_en.

[71] See: Michiel Rhoen, "Beyond consent: improving data protection through consumer protection law." *Internet Policy Review*, 5.1 (2016): 1-15, doi: 10.14763/2016.1.404; and, also Eoin Carolan, "The continuing problems with online consent under the EU's emerging data protection principles." *Computer Law & Security Review* 32.3 (2016): 462-473, doi: 10.1016/j.clsr.2016.02.004.

[72] Think about the cases where the number of data subjects involved is very large, or the individuals involved are incapacitated or in an emergency situation. On this aspect, see, again EDPB, *Document on response, op. cit.*, 2021, p. 4: "[…] informed consent to participate in the medical research project is a necessary requirement, with some exceptions for situations where consent cannot be given (incapacitated individuals, emergency situations etc.)".

[73] Christopher Kuner, Christopher Docksey, Lee Bygrave, *The EU General Data Protection Regulation: A Commentary, op. cit.,* p. 337.

[74] Recital 47, GDPR.

[75] On this aspect, an interesting analysis has been proposed by: Janos Meszaros, Chih-hsing Ho, "AI

- Attention should be paid to distinguish the so-called secondary use of personal data for research, from the case of data collected from the beginning for research purposes: in the case of secondary use, although it is always crucial to ensure safeguards for data subjects, there are several simplifications (e.g., it is not necessary to always identify a separate legal basis, or it is not necessary to obtain the informed consent[76]).

- Consent is not the only possible legal basis in the field of scientific research; very often, especially in the field of public scientific research[77], the most appropriate legal basis is public interest.

- It is crucial to avoid the confusion between the consent as a legal basis for processing under the GDPR and the informed consent related to ethical obligations, of research projects, as recently very clearly emphasised by the EDPB[78].

---

research and data protection: Can the same rules apply for commercial and academic research under the GDPR?." *Computer Law & Security Review* 41, 105532 (2021): 1-10, doi: 10.1016/j.clsr.2021.105532. The Author argues that – with specific regard to AI research – the exceptions provided for public research do not (or should not be) extend to a context involving also private actors.

[76] The relevance of the consent within the domain of scientific research has a long story: "The WP29 took the position that the further processing of health data for historical, statistical and scientific research purposes should only be permitted after having obtained the explicit consent of the data subjects […], but this position has not been reflected in the GDPR.", CHRISTOPHER KUNER, CHRISTOPHER DOCKSEY, LEE BYGRAVE, *The EU General Data Protection Regulation: A Commentary, op. cit.,* p. 1245.

[77] GIOVANNI MARIA RICCIO, GUIDO SCORZA, ERNESTO BELISARIO, *GDPR e normativa privacy, op. cit,* p. 658.

[78] "Ethics standards cannot be interpreted in such a way that only explicit consent of data subjects can be used to legitimise the processing of health data for scientific research purposes. Article 6 and Article 9 GDPR contain other options for a legal basis and an exemption, that can be relied on for processing health data for scientific research purposes. The requirement of informed consent for participation in a scientific research project can and must be distinguished from explicit consent as a possibility to legitimise the processing of personal data for scientific research purposes. It can be argued that ethical statements and bio-ethics conventions primarily aim to protect individuals against being included in medical research projects against their will and/or without their knowledge. […] such consent can and should be distinguished from 'consent as a legal basis for processing of personal data' in Article 6(1)(a) of the GDPR. Taking into consideration that Article 6 (1) GDPR provides for legal bases other than consent and Article 9 (2) GDPR provides for exemptions other than explicit consent, it is foreseeable and not incompatible (with ethical standards) that the other legal grounds can be relied on for the processing health data for scientific research purposes. However, when relying on another legal basis in Article 6 other than consent and one of the other exemptions in Article 9 (2) GDPR, the 'ethical' requirement of informed consent for participation in the medical research project will still have to be met. In the GDPR-framework, this can be perceived as one of such additional safeguards as foreseen in

Clarified the conditions for lawful processing, it is now time to address the problematic aspects arising from compliance with the principle of data minimisation.

5.3.2 Minimisation and Anonymisation

The Article 89 of the GDPR identifies the principle of data minimisation as the pivotal principle of personal data processing for archiving purposes in public interest, scientific or historical research purposes or statistical purposes[79]. According to the principle of data minimisation, as stated by the Article 5(1)*c*, personal data collected shall be "[…] adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed". These criteria of adequacy, relevance and limitation are, however, considerably difficult to fulfil in a scenario such as that of science: research projects change over time, due to the previous stages of the scientific investigation, modifying aims, methodologies, objectives and consequently purposes of data processing. This dynamic is even more evident in an Open Science context, characterised by the deep pervasiveness of digital tools, AI, Big Data and ICTs. Therefore, the current approach to research often challenges the very nature of the principle of data minimisation. Although "[…] under the GDPR personal data must be 'limited to what is necessary' instead of being 'not excessive'"[80] as under the previous European Directive 95/46/EC, compliance with this principle is still problematic in the research field.

Article 89(1) GDPR that should be in place when processing personal data for scientific research purposes" in EDPB, *Document on response, op. cit.*, 2021, p. 4.

[79] "Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner", the Article 89(1), GDPR.

[80] Christopher Kuner, Christopher Docksey, Lee Bygrave, *The EU General Data Protection Regulation: A Commentary, op. cit.,* p. 317.

Traditionally, in order to address the principle of data minimisation, in the field of research, has been adopted a twofold strategy: first, "[T]o comply with the principle of data minimisation, scientific researchers should first consider whether it is possible to conduct data processing without using personal data"[81]; second, the employment of pseudonymisation or anonymisation techniques[82].

In this second case, the anonymisation techniques permit to process anonymised data, i.e., data considered not anymore personal according to the GDPR[83]; and the pseudonymisation techniques permit to process pseudonymised data, i.e., data still personal but guaranteed through technical and organisational measures[84].

However, sometimes, the adoption of such a strategy is not sufficient. Under certain circumstances the use of personal data cannot be avoided, as it is a fundamental part of the research project (e.g., think about the fields of health, biomedicine, or genomics). In addition, there are important strands of research, in many fields[85], that have

---

[81] MARCELLO IENCA, *et al.*, *"How the general data protection regulation changes the rules for scientific research", op. cit.,* p. 26.

[82] UGO PAGALLO, "The legal challenges of big data: putting secondary rules first in the field of EU data protection." *European Data Protection Law Review* 3 (2017): 41, doi: 10.21552/edpl/2017/1/7: "[…] the employment of pseudonymisation techniques so as 'to ensure respect for the principle of data minimisation' in the processing of personal data for statistical purposes, in accordance with the wording of Article 89(1)".

[83] Recital 26 of the GDPR explicitly states "The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes".

[84] The pseudonymisation is defined by the Article 4(5) of the GDPR as: "[…] the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".

[85] Consider, in the field of genomics: YANIV ERLICH, *et al.*, "Identity inference of genomic data using long-range familial searches." *Science* 362.6415 (2018): 690-694, doi: 10.1126/science.aau4832; or related to in environmental health (EH) studies, see: KATHERINE E. BORONOW, *et al.*, "Privacy risks of sharing data from environmental health studies." *Environmental health perspectives* 128.1.017008 (2020): 1-12, doi: 10.1289/EHP4817; on the residual risk of re-identification: MICHÈLE FINCK, FRANK PALLAS, "They who must not be identified—distinguishing personal from non-personal data under the

investigated the risks associated with anonymisation: "Even in the absence of basic personal identifiers, it is often possible to re-identify subjects using available contextual knowledge. […] Due to the massive amount of data that may contain hundreds or thousands of observations per individual, simple anonymization techniques such as making a few attributes more coarse-grained are not effective against de-anonymization anymore"[86].

These concerns led some scholars to claim that it is impossible to guarantee compliance with the principle of data minimisation, through anonymisation, especially in the Open Science scenario, based on circulation, sharing and reuse of research data. They observed that "[…] large-scale de-anonymization attacks on such shared data sets will eventually pose a real threat for researchers"[87]. Such de-anonymisation processes would be made possible by the contextual knowledge that may be exploited to achieve the re-identification of data subjects by more advanced technologies, starting from the assumption that "[…] as soon as data sets contain additional contextual or supplemental information (e.g., timestamps, indirect location information such as geo-locatable IP address fragments), things can change very fast and data sets can become highly sensitive to attacks. These attacks may occur at a later date and could be fueled by cross-linkage to other data sets not yet released and attack schemes not yet developed at the time of the publication"[88].

Although these concerns are justifiable given the empirical technical limitations of the anonymisation processes, the principle of data miminisation can be still satisfied in an Open Science context, mainly for three reasons, described below: (i) the distinction between

GDPR." *International Data Privacy Law* (2020): 11-36, doi: 10.1093/idpl/ipz026.

[86] Benjamin Erb, *et al.*, "Emerging Privacy Issues in Times of Open Science." *PsyArXiv* (2021): p. 2, doi: 10.31234/osf.io/u236e.

[87] Benjamin Erb, *et al.*, *"Emerging Privacy Issues, op. cit.*, p. 3.

[88] Benjamin Erb, *et al.*, *"Emerging Privacy Issues, op. cit.*, p. 4.

accessibility and openness in the Open Science paradigm; (ii) the principle of accountability; (iii) and a reasonable approach.

5.3.2.1 Accessibility v. Openness

First, in arguing that "[…] efforts towards more transparency and open science in the aftermath of the replication crisis increase tensions with privacy and data protection and anonymity of study participants"[89], one crucial aspect is ignored: the Open Science paradigm does not imply a blind and indiscriminate openness of every part of the research process, including datasets. Open Science, as argued in Chapter 3, represents a new approach to the challenges of science, which aims to greater transparency and better communication, always guided by the formula "as open as possible, as closed as necessary". Therefore, if a dataset contains personal data whose sharing is likely to generate potential harm to the individuals involved, such data will not be open, but, instead, made potentially accessible. Accessibility is one of the FAIR Data Principles[90] which is not absolutely equivalent to the concept of openness. The openness consists in archiving data related to the research project in a repository that is either institutional or domain-based and releasing it in open modality, i.e. freely accessible by entering the repository. Data accessibility, on the other hand, still requires the researcher to store data related to the research project in a suitable repository, but access is not free, rather restricted by a qualified request.

An example shows the difference between accessibility and openness in the Open Science, related to the processing of personal data. Suppose a researcher intends to publish the results of a research project in which an essential part is represented by the processing of personal data. Let us further assume that such personal data cannot be anonymised for research purposes, or that the researcher considers

[89] *Ibid.*
[90] See: Section 4.1.2.

that such personal data, even if anonymised, could pose risks to the rights and freedoms of data subjects. In such circumstances, according to the tenets of Open Science, the researcher is required to store in the institutional repository or in the domain-based repository her data in a closed manner, i.e., not freely available to the outside world. This storage ensures accessibility, i.e., research data (personal or not) available and accessible upon explicit request. This would satisfy a potential request for access during the peer review phase of the publication: the reviewers of the submitted scientific paper will be able to access the data underlying the research project, in order to replicate the results and allow a thorough and legitimate check by the scientific community. This approach satisfies the goal of Open Science, guaranteeing transparency, but at the same time respecting the rights of the data subjects involved. Bear in mind, however, that the choices made by the researcher depend on an informed management of research data at all levels, including the protection of personal data[91].

5.3.2.2 The Meta-Principle of Accountability

The second reason why the principle of data minimisation is also achievable in an Open Science scenario is related, once again, to the design of the processes. The underlying principle of the European data protection system is the principle of accountability, enshrined in the Article 5(2) of the GDPR: "The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1", where paragraph 1 sets out the six principles of personal data processing. Additionally, Recital 74 explains that "[…] the controller should be obliged to implement appropriate and effective measures and be able to demonstrate the compliance of processing activities with this Regulation, including the effectiveness of the measures". The controller has the duty to demonstrate the *an* of the adoption, the *quomodo* and

[91] On this aspect, see: Section 5.5.2. There, it is argued that in order to ensure effective protection of personal data, a broader strategy of aware governance of research data in general needs to be implemented.

the *quantum* of the measures: "The controller should be able to demonstrate not only that she has in abstract terms complied with the obligations imposed, but also that she has adopted effective measures in respect of the *concrete* characteristics of the data and of the processing, as well as of its impact and the risks that it may determine for the rights and freedoms of the data subjects (and, where appropriate, also of third parties)"[92].

In other words, the principle of accountability can be interpreted as a meta-principle, guiding the whole application of the provisions of the GDPR. As Buttarelli stated: "Being accountable for data processing is not a substitute for compliance with the applicable legal obligations. It should be understood as an ethical responsibility for activities that take place for a given purpose, whether profit making, law enforcement, social care, or research—or even a combination of them"[93].

From this perspective, it is suggested that the principle of data minimisation should be interpreted as an "obligation of diligence", rather than a "specific-result obligation"[94]. In other words, the controller is obliged to put in place all the necessary safeguards to protect the data subjects involved. In contrast, the controller is not obliged to perform an activity, in order to obtain a certain result from it: the aim is to reduce as much as possible the probabilities that the data will be re-identified.

Thus, it is suggested to adopt the interpretation of the principle of accountability as a meta-principle, and the interpretation of the principle of minimisation as an obligation of diligence. In doing so, the dynamic nature of the protection of personal data imposed by the

---

[92] GIOVANNI MARIA RICCIO, GUIDO SCORZA, ERNESTO BELISARIO, *GDPR e normativa privacy, op. cit,* p. 62. [Translation from the Italian original text and the word "concrete" not emphasised in the original text].
[93] GIOVANNI BUTTARELLI, *"The EU GDPR as a clarion call for a new global digital gold standard.", op. cit.,* p. 78.
[94] Here, the reference goes to the Italian legal constructs of "*obbligazioni di mezzo e obbligazioni di risultato*" and the similar "*obligation de moyen et obligation de résultat*", according to the French Civil Law. A classic example of an intermediate obligation is the commercial activity of professionals (e.g., lawyers, doctors, etc.).

GDPR clearly emerges: the controller, always keeping in mind the main *ratio*, i.e., the protection of the freedoms and rights of the individuals involved, must modulate the measures adopted.

This can be done by adopting a series of organisational and technical measures that make it too expensive for third parties to carry out a re-identification process. Consequently, in this specific case, where it is deemed that there is a risk of de-anonymisation, a dynamic protection approach should be adopted, designing a process that provides for a set of subsequent safeguards.

A practical suggestion can be the introduction of practices that require certain datasets to be labelled as "potentially risky", thereby modulating the degree of accessibility in the institutional or domain repository, with regular state of the art assessments.

In the field of data, the approach of the Open Science and the GDPR is not at all divergent[95]: both support a dynamic protection and an effective and informed management of data, the processing of which is now an essential aspect of many fields of knowledge.

5.3.2.3 The Reasonableness Test

The third reason why data minimisation is feasible even taking into account the risks associated with anonymisation, is represented by the "reasonableness test". I call the "reasonableness test" the evaluation that the controller should realise about the anonymisation techniques she intends to adopt. This test is inspired by the Article 104 of the Italian law on privacy and data protection[96], entitled "Scope of

[95] On the convergence between Open Science and data protection, see: Section 5.4.

[96] Decreto Legislativo 30 giugno 2003, n. 196 "*Codice in materia di protezione dei dati personali, recante disposizioni per l'adeguamento dell'ordinamento nazionale al regolamento (UE) n. 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE*", https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2003-06-30;196!vig, as modified by the Decree 101/2018 (Decreto Legislativo 10 agosto 2018, n. 101, Disposizioni per l'adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonchè alla libera circolazione di tali dati

application and identification data for statistical and scientific research purposes". This article states that, in order to understand if data should be defined as "identification data" (and, as a consequence, falling within the scope of the law), "[…] account shall be taken of all the means that can be *reasonably* used by the data controller or by others to identify the data subject, including the knowledge acquired in relation to technical progress"[97]. In this respect, the GDPR itself, on several occasions, specifically related to the security of processing activities, refers to the state of the art of the technology[98].

The anonymisation techniques adopted pass the reasonableness test if are able to hinder the re-identification of personal data according to the state of the art of the technological progress. The reasonableness test can also be described by the formula "*ad impossibilia nemo tenetur*": imposing the closure of data that, at the state of the art, are anonymised – and therefore no longer personal – out of concern that any future technology might be able to de-anonymise them is overreaching.

Not to mention that imposing the closure of anonymised data for concern that any future technological development could be able to de-

e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).

[97] The Article 104(2), D. Lgs. 196/2003. [Translation from the Italian original text and the word "reasonably" is not emphasised in the original text]. In addition, this provision of the Italian Law is in line with Recital 26 of the GDPR, which states: "To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments".

[98] The Article 32, GDPR and also Recital 83, GDPR: "In order to maintain security and to prevent processing in infringement of this Regulation, the controller or processor should evaluate the risks inherent in the processing and implement measures to mitigate those risks, such as encryption. Those measures should ensure an appropriate level of security, including confidentiality, taking into account the state of the art and the costs of implementation in relation to the risks and the nature of the personal data to be protected". Among scholars, a similar perspective is presented in SOPHIE STALLA-BOURDILLON, ALISON KNIGHT, "Anonymous data v. personal data-false debate: an EU perspective on anonymization, pseudonymization and personal data." *Wisconsin International Law Journal*, 34 (2016): 284-322, in which the Authors assume a "risk-based approach to anonymized data", p. 321.

anonymise them would be a potential occurrence of what Pagallo has defined as the "risk of under-use"[99].

By contrast, even more complex is the problem of the transfer of personal data to third countries or international organisations, to which the next section is dedicated.

5.2.5 Transfer to Third Countries or International Organisations

Chapter 5 of the GDPR regulates the "Transfer of personal data to Third Countries or International Organisations". A transfer of personal data to third countries is not expressly defined by the GDPR[100]. The Article 4(23) provides the definition of "cross-border processing", but this kind of processing activity doesn't concern the international transfer of personal data. Rather, the cross-border processing concerns the case in which the controller or the processor are established in multiple EU Member States.

The concept of international transfer of personal data can be derived from the *ratio* behind this discipline, i.e., the intention to protect freedoms and rights of data subjects also beyond the geographical boundaries of the European Union[101].

The Open Science paradigm is based on free circulation of ideas, researchers, knowledge, and data: as stressed several times, knowledge doesn't have, by its very nature, geographical boundaries and limits, even more if its dissemination is realised through ICTs. As claimed in a recent study on the legal and policy framework of the EOSC: "[…] the demand of cross-border use of research services clearly does exist and

[99] UGO PAGALLO, *Il dovere alla salute. Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario.* (Milano-Udine: Mimesis, 2022).

[100] CHRISTOPHER KUNER, CHRISTOPHER DOCKSEY, LEE BYGRAVE, *The EU General Data Protection Regulation: A Commentary, op. cit.,* p. 762.

[101] In CHRISTOPHER KUNER, CHRISTOPHER DOCKSEY, LEE BYGRAVE, *The EU General Data Protection Regulation: A Commentary, op. cit.,* p. 763, the Authors also underline that the concept of international transfer of personal data can be also derived from the jurisprudence of the ECJ, with specific regards to the Linquvist Case (C-101/01) and the Schrems I Case (C-362/14).

is likely to grow"[102]. In parallel, this growing demand is also including international transfers outside the EU.

Therefore, it seems fair to admit that the Open Science is a scenario in which all the concerns, limits, and issues of the international transfer of personal data emerge even more clearly.

The GDPR sets "[…] a three-tiered structure for legal bases for data transfers, with adequacy decisions at the top, appropriate safeguards in the middle, and derogations at the bottom"[103]. In other words, according to the Article 45 of the GDPR "[A] transfer of personal data to a third country or an international organisation may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection"[104].

In the absence of the adequacy decision according to the Article 45 of the GDPR, the transfer is only possible and lawful if "[…] the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available", as established in the Article 46(1) of the GDPR.

---

[102] EOSC Pillar Project, "Legal and Policy Framework and Federation Blueprint.", v. 2.2 (2021), p. 8, doi: 10.5281/zenodo.4486610. Frequently, the expression "cross-borders" indicates both, processing carried out by controllers or processors based on multiple Member States, and transfer of personal data beyond the European Union.

[103] CHRISTOPHER KUNER, CHRISTOPHER DOCKSEY, LEE BYGRAVE, *The EU General Data Protection Regulation: A Commentary, op. cit.,* p. 764.

[104] The Article 45(1), GDPR. Currently, "The European Commission has so far recognised Andorra, Argentina, Canada (commercial organisations), Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Switzerland, the United Kingdom under the GDPR and the LED, and Uruguay as providing adequate protection", as declared at: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en. The mentioned "LED" stands for Law Enforcement Directive, i.e. Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016, *on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA*, OJ L 119, 4.5.2016, p. 89–131, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016L0680.

At the bottom, the last case in which the international transfer can be realised is regulated by the Article 49 of the GDPR, which states that "[I]n the absence of an adequacy decision pursuant to Article 45(3), or of appropriate safeguards pursuant to Article 46, including binding corporate rules, a transfer or a set of transfers of personal data to a third country or an international organisation shall take place only on one of the following conditions", providing a list of specific situations. Yet, the EDPB, in the guidelines regarding the processing of health data for research purposes, during the COVID-19 Pandemic, clarified that "[…] the derogations of Art. 49 of the GDPR do have exceptional character only"[105], also for the scientific research field.

5.3.3.1 After Schrems II Case

The discipline set out in the GDPR has been made even more complex by the ECJ's ruling, *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems*, known as the Schrems II Case, issued in 2020[106]. This ruling nullified the agreement concluded between the European Union and the United States, in 2016, known as "Privacy Shield" agreement: this settlement established rules for commercial operations in which personal data where processed and transferred to third countries ensuring the respect of the European data protection requirements[107]. The ECJ, in Schrems II, refused to allow data controllers to have US data processors using the "Privacy Shield" agreement as a lawful legal basis for processing or transferring personal data. The reason at the basis of this decision is that the United States didn't guarantee a sufficient and adequate protection for

[105] EDPB, *Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak*, *op. cit.*, p. 14.
[106] Case C-311/18, *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems* (2020) ECJ, ECLI:EU:C:2020:559.
[107] At the time, this agreement became necessary because of another ruling, the so-called Schrems I judgment (Case C-362/14 *Maximillian Schrems v Data Protection Commissioner* (2015) ECJ, ECLI:EU:C:2015:650) which nullified the previous agreement on international data transfer, known as "Safe Harbor" agreement.

personal data transferred from the European Union. The ECJ declared the "Privacy Shield" agreement not compatible with the Article 45 of the GDPR, concerning the adequacy decision and the Articles 7 and 8 of the Charter of Fundamental Rights of the European Union, enshrining the right to privacy and the right to the protection of personal data.

The focus, both in Schrems II and in the "Privacy Shield" agreement, was on international transfers of personal data for commercial purposes. Yet, this ruling generated a notable impact on the whole European data protection framework, even on the field of research. Consider, for instance, the EOSC: this trusted and federated environment for the benefit of researchers will be open – sooner or later – to international players, public and private, as explicitly declared by the President of the European Commission in 2020[108].

As described by some scholars, "[T]ypical bases to transfer personal data in the research context include standard contractual clauses approved by the European Commission or consent of the data subjects"[109]. However, the practical provision of consent seems to be very difficult to apply, in this kind of situations. If the controller from outside the EU intends to process personal data of data subjects within the EU on the basis of the consent, the controller needs to have a representative established in the EU. According to the Article 4(17) of the GDPR the representative is "[…] a natural or legal person established in the Union who, designated by the controller or processor in writing pursuant to Article 27, represents the controller or processor with regard to their respective obligations under this Regulation". The Article 27 of the GDPR, entitled "Representatives of controllers or

---

[108] During the 2020 World Economic Forum, Ursula Von der Leyen mentioned exactly the intention to open the EOSC to international actors and players. See: EOSC Portal, "EC President Ursula von der Leyen talks EOSC in Davos.", 22 January 2020, www.eosc-portal.eu/news/ec-president-ursula-von-der-leyen-talks-eosc-davos.

[109] David Peloquin, Michael DiMaio, Barbara Bierer, Mark Barnes, "Disruptive and avoidable: GDPR challenges to secondary research uses of data." *European Journal of Human Genetics*, 28 (2020):701-702, doi: 0.1038/s41431-020-0596-x.

processors not established in the Union", doesn't apply to "a public authority or body"[110], thus excluding public research entities[111]. Otherwise, without a representative, the three-tiered structure mentioned before[112] applies in this context.

Similarly, also the identification of standard contractual clauses generates some problems. First, it is required the approval of the European Commission. Second, "[…] the standard contractual clauses are not viable when the recipient entity is an arm of the US government, such as the US National Institutes of Health (NIH) or public universities or academic medical centers, because the US government cannot agree to certain terms in the standard contractual clauses including dispute resolution in European courts, and US state entities (including state universities and public hospitals)"[113]. In the end, this uncertain situation "[…] may create confusion in terms of who is ultimately controlling the data that are at stake"[114].

In light of the analysis of the European discipline of international transfers of personal data for research purposes, it emerges a clear uncertainty. However, despite this legal uncertainty, scientific research is progressing, including through international cooperation and exchange (which is essential, for instance, in one of the greatest challenges of our time, i.e., the fight against the SARS-CoV-2 virus). In addition, consider that the international cooperation is not only a practice: rather, "the encouragement and development of international contacts and co-operation in the scientific and cultural fields" should be interpreted as a real obligation for Member States that have signed the International Covenant on Economic, Social and Cultural Rights,

---

[110] The Article 27(2)b, GDPR.

[111] By contrast, including private research entities, e.g., provider of services related to specific research projects, or private managers of biobanks, etc.

[112] CHRISTOPHER KUNER, CHRISTOPHER DOCKSEY, LEE BYGRAVE, *The EU General Data Protection Regulation: A Commentary, op. cit.,* p. 764.

[113] DAVID PELOQUIN, *et al., "Disruptive and avoidable", op. cit.*, p. 702.

[114] ANNA BERTI SUMAN, ROBIN PIERCE, "Challenges for Citizen Science and the EU Open Science Agenda under the GDPR." *European Data Protection Law Review*, 4.3 (2018): 292, doi: 10.21552/edpl/2018/3/7.

according to the Article 15(4) of the Covenant[115]. Currently, the international cooperation in science, mentioned in the Article 15 of the Covenant, is also realised through the international transfer, sharing and reuse of personal data for research purposes. Avoiding regulating the international transfer of personal data for research purposes can therefore be considered as an infringement of the Article 15 of the Covenant.

In addition, the described legal uncertainty doesn't stop transfers, but rather, risks only to affect freedoms and rights of data subjects involved. In other words, the concern is that, at some point, the empirical difference between law in the books and law in action will be overwhelming.

5.3.3.2 An Alternative Option: The EOSC Association

The gap generated by the Schrems II Case should be seen as an opportunity to finally tackle, at the European level, the transfer of personal data to third countries and international organisations, also with specific reference to the research domain.

It is needed a European intervention in this matter because we cannot expect Member States to address the issue of international transfer of personal data for research purposes, for two main reasons. First, national interventions would generate more fragmentation. Having 27 disciplines for international transfer of personal data for research purposes would be a real disincentive to develop research projects with European partners. Second, from a geopolitical perspective, it is difficult to imagine that every Member State of the EU has the necessary power to impose itself on foreign countries, such as the United States, China, Japan and so on. Actually, this seems to be exactly the case in which the European Union should exercises its coordinating role in the field of research, as set out in the Title XIX of

[115] On the human and fundamental rights framework of the Open Science paradigm, see: Section 3.3 and Section 3.3.1.

the TFEU, described in Section 5.2.2. The European scientific research can only be supported and effectively sustained by a concerted effort, not by national initiatives.

An alternative solution while waiting for institutional intervention could be to develop codes of conduct at international level specifically for research purposes. The Article 40 of the GDPR establishes that "[A]ssociations and other bodies representing categories of controllers or processors may prepare codes of conduct […] such as with regard to […] the transfer of personal data to third countries or international organisations"[116].

Some scholars have shown how "[…] universities and research institutions had neither the incentives nor the resources to create internationally binding codes of conduct and settled for declarative documents or internal guidelines"[117]. However, it is believed that the emergence of the Open Science paradigm may change the scenario of traditional research in this respect as well.

If individual research centres or universities are not adequate in establishing international codes of conduct, the EOSC Association might be. As argued in Section 2.2.7, the EOSC Association is a legal entity authorised to be part of binding contractual agreements, established in 2020. The EOSC association aims to become the focal point of the governance of the EOSC, i.e., the federated and trusted environment of services and data for researchers across Europe[118]. The EOSC association has more than 200 members[119], and a structure that allows it to coordinate the various and multiple national initiatives, through its bodies: president, board of directors and general assembly.

---

116 The Article 40(2)*j*, GDPR.

117 MICHAL KOŠČÍK, MATĚJ MYŠKA, "Data protection and codes of conduct in collaborative research." *International Review of Law, Computers & Technology* 32.1 (2018): p. 141, doi: 10.1080/13600869.2018.1423888.

118 For a more in-depth analysis of the EOSC see: Section 2.2.3 and Section 2.2.4.

119 Here, the list of members of the EOSC Association: https://www.eosc.eu/general-assembly.

It is a well-structured entity, supported by the European institutions but, at the same time, is the result of coordinated activity at local level (its members are universities, research centres, associations in the scientific field, etc.).

The EOSC Association might be the entity that, in compliance with the Article 40 of the GDPR, has the necessary strength and capacity to define a code of conduct for the transfer of personal data to third countries and international organisations, specifically for scientific research purposes.

For this reason, its concrete participation in the European scientific research governance would be outlined in line with the approach of legal cooperation, where different regulatory systems, levels of intervention and actors interact[120].

This proposal could temporarily be a solution for the transfer of personal data to third countries and international organisations that nonetheless take place, in order to avoid sanctions, as an application of the Schrems II judgment[121]. Where uncertainty emerges from the legal framework, the intervention of other systems, levels and actors in the described legal coordination mechanism is essential. This intervention can be introduced by codes of conduct, as proposed in the genetic field, since "an ethical approach is essential to leverage the potential of digital innovation to improve science and public health and to address ethical risks before these lead to social rejection and too strict regulation, which would eventually hamper scientific progress"[122].

---

[120] See: Section 2.4.3.

[121] On this regard, see the recent decision of the EDPS against the European Parliament for transfer of personal data to third Countries: EDPS, *Decision of the European Data Protection Supervisor in complaint case 2020-1013 submitted by Members of the Parliament against the European Parliament*, 5 January 2022.

[122] Giorgia Brambilla Pisoni, Mariarosaria Taddeo, "Apropos Data Sharing: Abandon the Distrust and Embrace the Opportunity.", *DNA and Cell Biology* 41.1 (2021): p. 13, doi: 10.1089/dna.2021.0501.

Going further, in our context, the adoption of codes of conduct, by the EOSC Association may represent a way to address not only ethical risks, but also legal one.

Thus, in light of our analysis, it seems that the Open Science paradigm doesn't appear as a barrier in itself. Rather, the Open Science paradigm points out more clearly the preexisting weaknesses of the system. For this reason, next section argues that, by contrast, there is a substantial convergence between Open Science and data protection, despite the empirical challenges still persisting.

5.4 Open Science and Data Protection: A Possible Convergence

The analysis of the European data protection discipline in the field of scientific research, both at the European and national level, made possible to identify three major issues: i.e., the definition of the legal basis; the compliance with the principle of data minimisation, with specific regard to the anonymisation limits; and finally, the concerns related to the international transfer of personal data.

This analysis made clear two key assumptions. On the one hand, the Open Science scenario doesn't generate the major problems of the data protection in the field of scientific research. Rather, the emergence of the Open Science paradigm highlighted the weaknesses already existing in personal data processing activities for research purposes. On the other hand, "[…] data protection rules, such as the GDPR, are fully compatible with and do not hinder genuine scientific research. At the same time, sharing of personal data always involves a degree of risk to the data subjects, including where the purpose is scientific research, especially in cases of sensitive data. Data protection rules are intended to serve as a *robust safety net* for individuals whose data are needed to support science, as well as a framework steering researcher toward innovation that reflects the European values"[123], as recently stated by the EDPS.

As a result, the Open Science paradigm and the European data protection discipline proceed (or should proceed) in the same direction. The final aims pursued by the GDPR is not divergent from what the Open Science paradigm intends to achieve, at least as defined and interpreted in this dissertation[124].

However, it is necessary to recognise that some barriers still persist for data protection in the research domain. Here the intention is to describe the reasons behind the declared convergence between data protection and Open Science[125], before identifying these still persisting barriers, approaching possible solutions[126].

A misleading interpretation of Open Science can lead to deem that the data protection discipline is "a legal limit of Open Science"[127]. In Chapter 4[128], the perspective of a legal discipline as a "barrier" or "limit", opposing to others interpreted as "enablers", has been rejected[129]. There, the focus was on the Open Data discipline, considered by some scholars[130] as an "enabler" for Open Science, while,

---

[123] EDPS, *Opinion 3/2020 European Strategy for Data*, 2020, p. 9, [Words "robust safety net" not emphasised in the original text], available at: https://edps.europa.eu/data-protection/our-work/publications/opinions/european-strategy-data_en; as also declared in EDPS, *A Preliminary Opinion on data protection and scientific research*, 2020, *op. cit.*, p. 27.

[124] On the definition of Open Science, as adopted in this dissertation, see: Section 3.2.

[125] Investigating the tension between "privacy" and "access to knowledge", similarly Paolo Guarda comes to a similar conclusion, arguing that "[…] to resolve this conflict, a non-traditional approach must be used", see: PAOLO GUARDA, "Privacy e fruizione della conoscenza scientifica", *Pubblicazioni scientifiche, diritti d'autore ed open access – Atti del Convegno tenuto presso la Facoltà di Giurisprudenza di Trento* (2008): p. 13.

[126] Subsequently, Section 5.5 is dedicated to the obstacles still persisting, hindering an effective protection of personal data in the scientific domain.

[127] MARK PHILLIPS, BARTHA M. KNOPPERS, "Whose Commons? Data Protection as a Legal Limit of Open Science." *The Journal of Law, Medicine & Ethics* 47.1 (2019): 106, doi: 10.1177/1073110519840489.

[128] See: Section 4.2.1; and, also, Section 4.2.4.1.

[129] In Chapter 4, this critical analysis started from the X-Officio interpretation according to which the European data protection and the copyright discipline were seen as barrier to Open Science, opposed to the Open Data Directive, considered as an enabler of the Open Science. See: Section 4.2.1.

[130] OHAD GRABER-SOUDRY, *et al.,* "Legal Interoperability and the FAIR Data Principles" (1.0). *Zenodo*, (2021): p. 5, doi: 10.5281/zenodo.4471312.

here, the attention is on the data protection discipline, which is often considered as a limit to the development of the Open Science paradigm.

This interpretation, which considers the Open Science and the European data protection discipline far apart, starts from some adequate premises, although leads to a rejected conclusion. The supposed divergence between Open Science and data protection is generated from, at least, four concerns, analysed below: (i) the development of inequalities; (ii) the risk of science privatisation; (iii) the use of untrustworthy technologies; (iv) and a clash of cultures.

(i) Development of inequalities. The concern of the development of inequalities stems from the polarisation that the technological advancement generated: "The potential pitfalls of open science include exacerbating existing inequalities, by supporting the development of expensive new diagnostics and treatments that are practically available only to the stratum of the population who can afford them, while putting already-disadvantaged individuals and groups at risk of harms, such as discrimination and stigmatization"[131].

This risk is real: digital technologies and innovation triggered a redistribution of power, in every aspect of our societies, including even science. Yet, as argued by Pagallo, "[T]he zero-sum logic of the beastly state of humans in the state of nature has been challenged by the new experiences of open source and free software, with forms of scientific and commercial collaboration based on the autonomy of individuals and trust, with the continuous emergence of spontaneous forms of order in cyberspace. […] so, the problem is to ascertain how power is today redistributed between new and old social actors, leaving open the further question of establishing the role that law plays in these cases"[132]. Therefore, starting from the analysis on *how* the power is

---

[131] MARK PHILLIPS, BARTHA M. KNOPPERS, *"Whose Commons?", op. cit.*, p. 107.

[132] UGO PAGALLO, *Il diritto nell'età dell'informazione: il riposizionamento tecnologico degli ordinamenti giuridici tra complessità sociale, lotta per il potere e tutela dei diritti* (Torino: G. Giappichelli Editore,

currently redistributed, all actors at any level involved have the ethical responsibility to design the Open Science as inclusive as possible.

(ii) Risk of science privatisation. Strictly related to the first concern, there is a risk of "[…] *de facto* privatization of personal data, by organizing data in a manner that benefits only those who possess sufficient resources to allow them to usefully analyze them, thus transforming public funding of open science into an indirect subsidy to private industry"[133]. In this dissertation, in several circumstances, an intermingling between private and public sector in the domain of scientific research has been pointed out[134]. The blurred boundaries between public and private spheres may be problematic to the extent that "[…] the reliance on market-oriented services […] raises additional concerns in terms of the awareness and freedom of the users in sharing"[135]. Yet, once again, it is a matter of design. As shown in Section 2.4.2, although the implementation of Open Science projects has been underway for some years now, the development is still in the early stages. For this reason, all the actors and institutions involved are responsible for the final development of the Open Science. This means, first of all, taking aware choices in the domain of infrastructures and services[136], in order to have a scenario reflecting a precise project of Open Science, designed to strive for transparency and inclusiveness.

(iii) Untrustworthy technologies. Some scholars claims that "[…] efforts towards more transparency and open science in the aftermath of the replication crisis increase tensions with privacy and data

2014): pp. XX – XXI.

[133] MARK PHILLIPS, BARTHA M. KNOPPERS, "*Whose Commons?", op. cit.*, p. 107.

[134] See: Section 4.2.3; Section 4.3.1; Section 4.2.4.2. This aspect will also be discussed in Section 6.3.

[135] ANNA BERTI SUMAN, ROBIN PIERCE, *"Challenges for Citizen Science", op. cit.*, p. 289. The Authors refers to the concerns related to the sharing of health data in a context of Citizen Science. Assuming – as it has been argued in Chapter 3 – that Citizen Science is one of the several dimensions of the Open Science paradigm, the reasoning presented by the Authors can be applied to the general discussion on Open Science.

[136] See: Chapter 6.

protection and anonymity of study participants is getting more at risk"[137] because the sharing of personal data which are anonymised today can be re-identified in the future, with technological improvements and progress.

The emergence of the Open Science paradigm is based on, and is generated thanks to, the use of new technologies, digital ICTs, new computational powers, huge amount of data, elaborated by machines, AI, and algorithms. The openness of the scientific research process, in every stage, most of the time depends on the technology. For this reason, the research process is facing all the concerns typical of a human activity strictly based on technologies. This doesn't mean that the Open Science is irreparably incompatible with the data protection discipline. On the contrary, this situation requires a more careful analysis. The Open Science paradigm requires a check and balance system constantly tested, avoiding a "one-fits-all" solution. As Guarda claimed: "[I]t is not possible to be overcome by a naive optimism, considering everything that comes from technical evolution as an (inevitable) progress that in an inescapable and linear way will lead humanity to new heights of well-being and pleasure. Nor is it possible to withdraw into the darkest obscurantism, condemning everything we do not know as intrinsically wrong. On the contrary, we must try to understand the direction of the (social, economic and technological) transformations: it is essential to understand how to govern processes in an ethical and socially sustainable way"[138]. Moreover, this is the same approach envisaged in the GDPR, based on accountability and on the analysis of actual risks at stake, with the "Data Protection Impact Assessment" (DPIA) pursuant to the Article 35 of the GDPR.

(iv) Clash of cultures. For some scholars, then, the "[T]ensions between scientific openness and data protection" may also be found in

---

[137] BENJAMIN ERB, *et al.*, *"Emerging Privacy Issues, op. cit.*, p. 4.
[138] PAOLO GUARDA, *Il regime giuridico dei dati della ricerca scientifica.* (Trento: Collana della facoltà di Giurisprudenza dell'Università di Trento, 2021), pp. 38-39.

a "[…] clash of two possibly conflicting cultures, that of opening and that of protecting research data"[139]. This assumption is understandable since, as claimed by Sarah Jones about the implementation of the EOSC: "[…] many of the challenges are more on the social and policy side than on the technical side"[140]. However, this consideration gives rise to a distorted vision of the openness of science. Underlying this clash of cultures – which does exist – results in two misinterpretations.

First, there is an ambiguity in what is meant by Open Science. Once again, in fact, it must be stressed that such openness is never indiscriminate and without criterion: science is intended to be "as open as possible, as closed as necessary", based on balancing of opposing interests, typical of law.

Second, there is a distortion in the vision of the European data protection framework. Once again, the view of the GDPR as a limitation or barrier to data circulation should be refused. Although the GDPR sets out the aim of "protecting the fundamental rights and freedoms of citizens, with particular regard to the right to the protection of personal data", also sets an objective in relation to the free movement of data, namely, not to restrict or prohibit such movement for reasons connected with the data protection[141].

In other words, the GDPR should rather be interpreted as a set of rules governing the circulation of personal data, aimed at safeguarding the fundamental rights of individuals: the aim is not to limit, but rather to regulate.

However, some scholars argued that "[D]espite the appearance that the GDPR strikes the proper balance between accommodating scientific research and securing individual rights and dignity, the tension between open science and data protection goes to the very core

[139] ANNA BERTI SUMAN, ROBIN PIERCE, *"Challenges for Citizen Science", op. cit.*, p. 288-289.
[140] Sarah Jones is the new EOSC Engagement Manager at GÉANT, an expert and long-time advocate of Open Science; see: VERONICA BERTACCHINI, *et al.*, *EOSC Symposium Report*, 2021, p. 13, Zenodo, doi:10.5281/zenodo.5176089.
[141] The Article 1, GDPR.

of the two movements"[142]. On the contrary, it is deemed that the problems that still persist in this context are not to be found in the essence of the GDPR *per se*, nor in the foundation of the Open Science.

Next section therefore investigates these remaining barriers, to clarify where they stem from and to suggest how to tackle them.

5.5 Persisting Barriers to Data Protection in the Scientific Domain

The data protection issues still persisting in the scientific context may lead to assume that the Open Science paradigm is unavoidably divergent from the tenets of the European data protection discipline.

However, it is deemed that the Open Science paradigm doesn't generates these challenges, rather it makes more evident previously existing weaknesses. Thus, the persisting barriers to data protection in the scientific domain should be found in other three main problems, analysed below: (i) the legal fragmentation generated by different disciplines, related to every Member State (Section 5.5.1); (ii) the lack of an all-encompassing data management strategy at local level, (Section 5.5.2); and (iii) a "closed science", not enough transparent, neither inclusive nor based on cooperation (Section 5.5.3).

5.5.1 National Fragmentation

As seen in Section 5.2, the national level plays a fundamental role in the definition of rules governing data protection in the scientific domain. This choice of the European institutions, although understandable considering the distribution of competences between the EU level and Member States, may represent a barrier. The fragmentation of the EU data protection discipline for scientific research has been broadly investigated by scholars[143]. As regards the

[142] MARK PHILLIPS, BARTHA M. KNOPPERS, "*Whose Commons?", op. cit.*, p. 109.
[143] Many scholars have investigated the consequences of this choice of the European lawmaker. See, *ex multiis*: PAOLA AURUCCI, "Legal Issues in Regulating Observational Studies: The Impact of the GDPR on Italian Biomedical Research." *European Data Protection Law Review* 5(2) (2019): 197-208, doi:

emergence of the Open Science paradigm, this fragmentation is problematic for two main reasons: (i) it generates legal uncertainty; and (ii) it represents a constraint on the establishment of a "European researcher" profile.

(i) Legal uncertainty. Today, there is a clear need for greater coordination of data protection in research. This need, claimed by many scholars, is also proven by several opinions released by the EDPB and the EDPS, in the field of research, with the aim of clarifying discussed aspects or outlining a commonly pursued approach[144]. The risk is that at a local level, universities and research centers will limit themselves in undertaking certain activities involving the processing of personal data, perhaps using cutting-edge technologies, out of fears related to (or supposed to be related to) the protection of personal data. In other words, the risk is that the uncertainty stemming from multiple and different disciplines, which sometimes even overlap, becomes a very strong disincentive to carry out such research activities: in this circumstance, as observed by Pagallo concerning the use of AI in the health domain, a problem of under-use may emerge[145].

The European Union, from a geopolitical perspective, presents itself as an efficient regulator, also inspiring a specific approach[146]. However, a regulatory framework is beneficial insofar as it protects an asset or

10.21552/edpl/2019/2/9; MASSIMO DURANTE, MONICA SENOR, "Report on the harmonization of Italian Law with the enforcement of The GDPR.", in KAREN MCCULLAGH, OLIVIA TAMBOU, SAM BOURTON (eds.) *National Adaptations of the GDPR* (2019), https://wp.me/p6OBGR-3dP; CHIH-HSING HO, "Challenges of the EU' general data protection regulation' for biobanking and scientific research." *Journal of Law, Information and Science* 25.1 (2017): 84-103; CHRISTOPHER F. MONDSCHEIN, COSIMO MONDA, "The EU's General Data Protection Regulation (GDPR) in a research context." in PIETER KUBBEN, MICHEL DUMONTIER, ANDRE DEKKER (eds.), *Fundamentals of clinical data science* (Cham: Springer Nature, 2019): 55-71, doi: 10.1007/978-3-319-99713-1.

[144] See, among others: EDPS, *A Preliminary Opinion on data protection and scientific research*, 2020, *op. cit.*; EDPB, *Document on response, op. cit.*; EDPS "Preliminary Opinion of the European Data Protection Supervisor on the European Health Data Space", March 2021, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021XX0323(02).

[145] UGO PAGALLO, *Il dovere alla salute. Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario.* (Milano-Udine: Mimesis, 2022).

[146] In Section 2.1.2, the risks connected to the so-called Brussels Effect have already been discussed.

an interest and strikes a balance between opposing interests. A regulatory framework in the field of research that is characterised by fragmentation, and which generates legal uncertainty, cannot be justified solely by the desire to guarantee leeway to the Member States, at the expense of the effectiveness of the system itself.

(ii) The constraint on the establishment of a "European researcher" profile. This national fragmentation (and the corresponding uncertainty it generates) stands in contrast to the EU policies on Open Science and to the European Research Area (ERA)[147], which aims to facilitate the circulation of knowledge and researchers as much as possible. The European Commission in 2020, proposing "a new ERA", stated that the intention pursued is "[...] to further progress on the free circulation of knowledge in an upgraded, efficient and effective R&I [*Research and Innovation*] system, in particular by moving from an approach of coordination towards deeper integration between national policies. The ERA will continue to promote adequate framework conditions and inclusiveness, help develop the skills that researchers need for excellent science, and connect all actors across Europe, including in education, training and the labour market"[148]. If the intention is to shape European, or even international, researchers' profiles, then identifying multiple research disciplines at national level will not be effective.

As an illustration, consider some examples. The French data protection discipline requires the legal basis of the consent for processing personal data with the purpose of research[149]. The Italian

---

[147] The European Research Area (ERA) is a space of free movement for scientific research in Europe, which aims to ensure an increasing movement of research personnel among Member States; and, in addition, it intends to promote a greater circulation of knowledge through the medium of the digital tools. On the ERA, its origin and development, see: Section 2.2.1; but also, Section 2.2.5.
[148] European Commission, *A new ERA for Research and Innovation*, COM/2020/628 final, p. 5, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:628:FIN.
[149] Decreto Legislativo 30 giugno 2003, n. 196, Article 110 "Ricerca medica, biomedica ed epidemiologica".

law, in the Article 110, precisely regulates when, in conducting processing activities of personal data for medical, biomedical and epidemiological research, the consent is not needed[150].

Furthermore, consider the national data protection law of Luxembourg[151]. The Article 65 provides a list of twelve measures the controller shall implement in order to process personal data for research purposes (e.g., anonymisation, definition of the DPIA, etc.)[152]. However, the last paragraph of the same Article 65 states: "The controller shall document and justify for *each* project for scientific or historical research or statistical purposes the exclusion, if any, of one or more of the measures listed in this Article"[153]. This disposition requires a considerable amount of effort to be invested in each single research project, resulting in an analysis that is almost comparable to a further data protection impact assessment.

Only from these three examples, related to France, Italy and Luxembourg[154], it clearly emerges the complexity that stems from different rules and disciplines, for the same domain. In addition, consider that the analysis on which dispositions and requirements must be applied to a specific research project is generally carried out by the researchers involved. This assessment represents a notable effort, in terms of time and also in terms of competences, taking into

[150] Loi n. 78-17 du 6 janvier 1978, *relative à l'informatique, aux fichiers et aux libertés*, ELI: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/2021-12-11/. The Article 75 states: "Dans le cas où la recherche nécessite l'examen des caractéristiques génétiques, le consentement éclairé et exprès des personnes concernées doit être obtenu préalablement à la mise en œuvre du traitement de données".

[151] Loi du 1er août 2018, *portant organisation de la Commission nationale pour la protection des données et du régime général sur la protection des données* (Luxembourg Data Protection Law) (2018), ELI: https://data.legilux.public.lu/eli/etat/leg/loi/2018/08/01/a686/jo.

[152] Here, again, it is used the simplification described at the beginning of the chapter: also in this case the data protection Law of Luxembourg refers to "des fins de recherche scientifique ou historique, ou à des fins statistiques".

[153] The Article 65, Luxembourg Data Protection Law, 2018. [Word "each" not emphasised in the original text].

[154] The reference to these three different and potentially contrasting dispositions from these three national frameworks (i.e., Italy, France, and Luxembourg) is merely illustrative, but it also evokes the three national legal frameworks more related to this doctoral research project.

account the fact that researchers don't necessarily have a legal background or familiarity with laws.

In an Open Science scenario, in which the cross-border collaboration (not to mention the international collaboration, analysed in Section 5.2.5) is promoted, problems deriving from this fragmentation become crystal clear.

Similarly, another problematic aspect made evident by the emergence of the Open Science paradigm is the lack of a coherent and common strategy for the governance of research data, which also exerts effects on the sphere of personal data protection: this discussion is developed in the following section.

5.5.2 Lack of Research Data Strategies

Beyond the national level and the major problem of fragmentation, at local level there is a problem of data governance. Currently, the strategy on the management and governance of research data is developed at local level, within the university or the research center. Sometimes, it lacks even the internal policy on the management of research data, which is left to the initiative of each researcher.

The compliance with the GDPR should necessarily be considered as a component of the broader data governance strategy. Many of the empirical challenges posed by the application of the EU data protection discipline to the field of research stems from lack of understanding of the bigger picture, i.e., the management of research data. Even scholars who are skeptical about the openness of science, in relation to the protection of personal data, acknowledge that "[A] possible solution could visualise data-sharing practices to build a greater awareness among both data users and (potential and existing) research participants of how data is being used, and to encourage more data sharing"[155]. As long as researchers continue to keep terabytes of their

[155] JUSAKU MINARI, GO YOSHIZAWA, NARIYOSHI SHINOMIYA, "COVID-19 and the boundaries of open science and innovation: Lessons of traceability from genomic data sharing and biosecurity." *EMBO reports* 21.11 (2020): p. 1, doi: 10.15252/embr.202051773.

research data on their own servers or in third-party cloud computing infrastructures without common guidelines and management[156], the issues, rules and practices related to sharing and reuse will be secondary[157].

The protection of personal data of data subjects involved in research projects needs to become part of a broader management of research data. For this reason, the emergence of the Open Science paradigm doesn't represent a further limit. On the contrary, an essential pillar of the Open Science paradigm is represented by the FAIR Data Principles[158], which are a set of guidelines exactly for data management. The adoption of the FAIR Principles is not devoid of challenges: e.g., these principles are not standards, thus they need to be conformed to the domain of application. However, the FAIR Data Principles provide a direction to follow, the infraethical infrastructure[159], in the definition of the most suitable data governance strategy.

An accountable analysis on the protection of personal data in the research context, developed case by case, should be part of the process of FAIRness of research data and, eventually – if and only if possible – of the process of openness (according to the well-known formula "as open as possible, as closed as necessary").

In Chapter 4[160], the role of data stewardship has already been investigated. Here, again, it is important to further emphasise this concept. The Open Science paradigm identifies the current way of doing science, strongly driven by new technologies, digital ICTs, computational power, AI and ML: if data have a significant role in today's science, then it is crucial to take into account their

[156] On the relevance of the research infrastructures, as a precondition for the establishment of the Open Science paradigm, see, Section 6.1.
[157] This situation is even more complicated when, in such circumstances, foreign disciplines apply, first and foremost the CLOUD Act, analysed in Section 6.3.
[158] The FAIR Data principles has been investigated in Section 4.1.2.
[159] See: Section 4.3.2.
[160] See: Section 4.2.3.1 and Section 4.2.3.2.

management. In addition, in designing the processes to manage research data, it is fundamental to envisage a supporting role of data stewards in analyising data protection issues[161].

So far, therefore, it seems that in light of these barriers to effective protection of personal data, the Open Science paradigm does not emerge as a limitation, but rather as a beneficial boost in promoting the overcoming of them. This aspect is made even more evident by the third barrier to the protection of personal data of individuals involved in science, namely the absence of transparency, which is discussed in the next section.

5.5.3 Closed Science

Currently, scientific research is increasingly characterised by the intermingling between private and public sector[162] and "[T]he risk is that behind the hype of a science more open to the citizen and thus more democratic, the 'winner' are the market actors rather than the people"[163]. This concern is real.

The use of digital technologies in conducting scientific research projects necessarily implies the involvement of the private sector. Hence, it is likewise necessary strike the right balance between public and private interests.

The only approach to define this relationship in a proper way is to be aware about the dynamics characterising the scenario. In other words, the main risk for the protection of personal data in the context of scientific research doesn't stem from the Open Science, but rather from a closed science, with no inclusiveness, cooperation and – especially – transparency. The lack of transparency in doing science is

---

[161] Here, again, the intention is to underline the relevance of having data steward with an interdisciplinary competence: technical, domain-based but also legal. On this aspect, see: Section 4.2.3.1.

[162] See: Section 4.2.3; Section 4.3.1; Section 4.2.4.2; Section 5.3.1. This topic is also further discussed in Section 6.3.1, entitled "Public-Private Interplay in Science", as part of the wider investigation concerning the role of e-infrastructures in scientific research.

[163] ANNA BERTI SUMAN, ROBIN PIERCE, *"Challenges for Citizen Science", op. cit.*, p. 290-291.

exacerbating the crisis of reproducibility and replicability of science; while the lack of transparency in defining the rules which regulates the scientific research (i.e., evaluation and careers advancements; allocation of funding; use of resources) represents a considerable barrier.

On the contrary, the Open Science approach is based on inclusiveness, cooperation and transparency. As regards data protection, it means to be accountable about the processing activities of the personal data involved and aware of the risks related to every single research project.

In other words, the Open Science paradigm, systematising the management of research data, helps to concretise the principle of informational self-determination in the research domain. The informational self-determination principle has been defined by the German Constitutional Court, in 1983[164]. According to the *Bundesverfassungsgericht,* "A social order – and the legal order enabling it – in which the citizens no longer can know who knows what, when, and on what occasion about them, would be incompatible with the right to informational self-determination. [...] If someone is uncertain whether deviant behaviour is noted down and stored permanent as information, or is applied or passed, he will try not to attract attention by such behavior. [...] This would not only impact his chances of development, but would also impact the common good, because self-determination is an elementary functional condition of a free democratic society based on its citizen capacity to act and to cooperate"[165].

---

[164] Bundesverfassungsgericht, Judgment of 15 December 1983, 1 BvR 209/83, ECLI:DE:BVerfG:1983:rs19831215.1bvr020983.

[165] MASSIMO DURANTE, *Computational Power, op. cit.*, p. 143; ANDREA GLORIOSO, "Un nuovo concetto di auto-determinazione informazionale come bussola concettuale per navigare il nuovo mondo digitale.", in UGO PAGALLO, MASSIMO DURANTE (eds.), *Manuale di informatica giuridica e diritto delle nuove tecnologie*. (Milano: Utet Giuridica, 2012), 383-394; and UGO PAGALLO, "On the principle of privacy by design and its limits: Technology, ethics and the rule of law." in SIMONA CHIODO, VIOLA SCHIAFFONATI (eds.), *Italian Philosophy of Technology*. (Cham: Springer, 2021): 111-127, doi: 10.1007/978-3-030-

Applying this principle to the topic under investigation, consider the case in which someone is uncertain whether her personal data, processed for research purposes, are also (or mainly) processed by private actors, with different purposes: this would impact on the inclination of individuals to allow the processing of their personal data for research purposes and, thus, the "common good" also referred to by the German Constitutional Court. Note that, in defining the informational self-determination, the *Bundesverfassungsgericht* emphasised the fact that "[H]owever, the sharing of data for scientific purposes (§9.4 of the 1983 Census Act) is compatible with the Basic Law"[166].

A more open and transparent scientific research process may be the means by which to guarantee the principle of informational self-determination of individuals in the context of science. "More generally, Open Science could again help to alleviate the ongoing problems with and serious challenges for the peer review and publishing system by encouraging transparency, accessibility and accountability"[167], but much will depend on how the Open Science paradigm will be implemented.

Accordingly, the EDPS recently stated that "[…] boundaries between public interest, academic freedom and private gain today are more blurred than ever. This uncertainty may create loopholes in the protection of fundamental rights, including the right to privacy and to personal data protection. The EDPS therefore strongly recommends that the Strategy and the envisaged legislation address specifically the definitions and the scope of the key notions such as scientific research,

54522-2_8.

[166] Bundesverfassungsgericht, Judgment of 15 December 1983 – 1 BvR 209/83, p. 2, ECLI:DE:BVerfG:1983:rs19831215.1bvr020983.

[167] JUSAKU MINARI, GO YOSHIZAWA, NARIYOSHI SHINOMIYA, *"COVID-19 and the boundaries of open science and innovation", op. cit.*, p. 2.

innovation, public interest, to avoid inconsistency with existing notions in the GDPR"[168].

This intermingling between public and private sector in the domain of scientific research has been analysed so far from different point of view. As regards the data protection perspective, this unavoidable interplay may be addressed with a transparent approach, in line with the Open Science attitude, instead of engaging in opaque management of research data, including personal data. However, this topic is further investigated in the following Chapter 6, as an aspect of the study on the role of e-infrastructures in the scientific domain.

5.6 Conclusive Remarks

Chapter 5 focused on the protection of personal data in the research domain, investigating whether and to what extent the emergence of the Open Science paradigm is complexifying the scenario. This study intended to provide an answer to RQ5:

> Which issues arise from the protection of personal data in the context of scientific research?
> Does the emergence of the Open Science scenario complexifies the compliance with the EU data protection discipline?

The analysis about the data protection issues that represent the main barriers to the development of scientific research led to argue that the Open Science paradigm doesn't appear as a barrier in itself. Rather, the Open Science paradigm brings out more clearly the preexisting weaknesses of the system.

The study about the relation between Open Science and data protection started identifying the kinds of processing activities carried out in the Open Science context. This resulted in the decision to focus on the case in which personal data of individuals are processed, by the scientific community, for research purposes.

[168] EDPS, *A Preliminary Opinion on data protection and scientific research*, 2020, *op. cit.*, p. 10.

In order to identify the major data protection issues related to this specific kind of processing activities, attention was drawn to the EU data protection framework, considering the European dimension, i.e., the GDPR, and the role of Member States.

The examination of the EU derogatory data protection discipline of processing activities for research purposes has enabled the identification of the main issues, represented by (i) the definition of the lawful legal basis; (ii) the compliance with the principle of data minimisation; and (iii) the challenges posed by the transfer of personal data to third countries and international organisations.

(i) Concerning the first issue, the chapter presented an analysis on the different legal bases which can be applied in scientific research. Following this, a set of guidelines was proposed to orientate the choice of the lawful legal basis in the case-by-case assessment that researchers are required to make when processing personal data for research purposes.

(ii) Regarding the principle of minimisation and its observance in the context of Open Science, a three-step reasoning was presented, described below.

- First, it is necessary to keep in mind the crucial distinction between the concept of openness (i.e., free, and immediate access) and that of accessibility (i.e., availability and potential access) of research data.

- Second, the application of the principle of accountability in its meaning of meta-principle is fundamental: the controller must adopt all organisational and technical measures that satisfy the principle of minimisation, through anonymisation techniques where possible. However, this obligation should be interpreted as a "diligence obligation" and not a "specific-result obligation". In other words, the controller has to minimise the likelihood of re-identification of personal data and be able to prove it.

- Finally, it is suggested to face the issue of minimisation and anonymisation in the field of scientific research with a test inspired by the Italian legislator, i.e., "the reasonableness test". In other words, in defining the technical and organisational measures to protect data subjects through anonymisation techniques, the controller must necessarily conform to the state of the art of the technological progress, at the same time taking into account the formula "*ad impossibilia nemo tenetur*".

(iii) The third issue, represented by the challenges posed by the transfer of personal data to third countries and international organisations has been described, considering both the legislative discipline, and the scenario transformed by the Schrems II Case. Then, an alternative option to face this issue has been envisaged, represented by the adoption of codes of conduct by the EOSC association.

Afterwards, in light of the suggested solutions to the major legal challenges of data protection in research, the convergence of data protection and Open Science was argued.

However, a number of still persisting barriers to the effective protection of personal data in research were acknowledged. These obstacles are represented by (i) the national fragmentation of the EU data protection discipline; (ii) the lack of data governance at local level; and (iii) a closed science, without transparency and inclusiveness.

In light of this described study, two remarks can be sketched, the first one *de lege data* and second one *de lege ferenda*.

Currently, *de lege data*, in the process of compliance with the GDPR, there is a lack of national coordination among the various research entities. The marker is the absence of an all-encompassing data governance strategy: the protection of personal data in research projects needs to become part of a broader management of research data. The progress of scientific research as a whole depends on the local approach on research data management.

Concerning *de lege ferenda* dimension, instead, the most consistent barrier is represented by the fragmentation of the discipline between

Member States, connected to an opaque intermingling of public and private sector in research. Since a more effective intervention of the EU institutions seems to be conceivable in light of the EU Treaties[169], it would be suitable a better co-ordination of the research domain at the European level.

However, to overcome these barriers (i.e., fragmentation and lack of transparency) it is needed to wonder how the Open Science paradigm will be implemented in Europe. If the European approach seems clear, the real challenge is the implementation[170]. Accordingly, it is essential to investigate the research infrastructures at the basis of the European Open Science implementation phase: for this reason, the Open Science e-infrastructures are topic of the next chapter.

---

[169] As argued in Section 5.2.2.

[170] The relevance of the implementation phase of the Open Science paradigm has been clarified in Section 3.4.

Chapter 6

Open Science e-Infrastructures

"While we were sleeping, computers and in particular virtual machines rapidly became our most important research assistants, but we continue to make their jobs miserable by spitting out narrative, PDF's and other file formats that are near-useless for computers"[1]. With this remark, Barend Mons, provocatively, aims to raise an issue that has been framed, from a philosophical point of view, by Luciano Floridi: in the scenario outlined by the digital revolution, human and artificial agents cooperate and coexist as similar "informational organisms" or "*inforgs*"[2]: "Today, we are slowly accepting the idea that we are not standalone and unique entities, but rather informationally embodied organisms (*inforgs*), mutually connected and embedded in an informational environment, the infosphere, which we share with both natural and artificial agents similar to us in many respects"[3]. Similarly, in science, researchers and machines coexist and collaborate in conducting research. Human agents, i.e., researchers, and artificial agents, i.e., machines, can be considered both as informational agents of scientific research or *inforgs* of science.

---

[1] BAREND MONS, *Data stewardship for open science: Implementing FAIR principles*. (Boca Raton, Florida: CRC Press, 2018), p. 2.

[2] "We are increasingly delegating or outsourcing to artificial agents our memories, decisions, routine tasks, and other activities in ways that will be progressively integrated with us. […] What I have in mind is rather a quieter, less sensational, and yet more crucial and profound change in our conception of what it means to be human. We are regularly outsmarted and outperformed by our ICTs. They 'reckon' better than we do. And because of this, they are modifying or creating the environment in which we live. We have begun to understand ourselves as inforgs not through some biotechnological transformations in our bodies, but, more seriously and realistically, through the radical transformation of our environment and the agents operating within it." in LUCIANO FLORIDI, *The fourth revolution: How the infosphere is reshaping human reality.* (Oxford: OUP, 2014), pp. 190, 193-194, EPub.

[3] LUCIANO FLORIDI, *Ethics after the information revolution.* (New York: Cambridge University Press, 2010), p. 11. But see, also: LUCIANO FLORIDI, "Artificial intelligence's new frontier: Artificial companions and the fourth revolution." *Metaphilosophy* 39.4-5 (2008): p. 651.

This consideration is at the basis of the emergence of the Open Science, as a new paradigm of science generated by the impact of new technologies on the scientific research process. For this reason, in order to assess the European approach to Open Science, it is essential to take into account the e-infrastructures, as frameworks in which the *inforgs* of science operate, both machines and humans[4].

Section 6.1 investigates the role of digital infrastructures in science, taking into account their rapid development in the last decades.

Section 6.2 focuses on a specific computing infrastructure, namely the High Performance Computing (HPC), in scientific research. In doing so, the investigation starts with the definition of HPC. Subsequently, the development of the European HPC policies from 2012 to the present are mapped. This analysis is instrumental in understanding the current European HPC strategies, i.e., the new Council Regulation of EuroHPC and its link to Open Science.

In light of the public-private sector interplay in this area that emerged from the analysis, Section 6.3 investigates the impact of the "Clarifying Lawful Overseas Use of Data Act" or CLOUD Act, enacted by the Congress of the United States, in 2018[5].

Once the outlines of this complex issue have been untangled, Section 6.4 is devoted to the case study of this dissertation, represented by the HPC facility of the University of Luxembourg, namely ULHPC[6]. The case study is described in three sections, respectively dedicated to: (i) the methodology adopted; (ii) the results of the technical analysis, i.e.,

---

[4] On the relationship between data and research infrastructures, the "Strategic Research and Innovation Agenda" (SRIA) is extremely clear: "Data without e-infrastructures to store, compute and connect are of no use to EOSC and can only exist on paper or in the researcher's head. On the other hand, e-infrastructures without any data (only 'zeros' or 'ones') are meaningless. Data- and e-infrastructures form what can be thought of as a 'Yin-Yang' relationship. One is not possible without the other.", see: EOSC Executive Board, "Strategic Research and Innovation Agenda (SRIA)", version 1.0, February 2021, https://www.eosc.eu/sria, p. 64.

[5] Clarifying the Lawful Use of Overseas Data Act, CLOUD Act, Pub. L. No. 115–141, 132 Stat. 348 (codified as amended in separate sections of 18 U.S.C.), 2018, available at: https://cli.re/BwPk5Q.

[6] https://hpc.uni.lu.

the mapping of the data flow in ULHPC; (iii) the findings of the legal analysis.

Section 6.5 concludes the chapter, summarising the major insights of the study on Open Science digital infrastructures and the role of HPC for research purposes.

6.1 The Role of Digital Infrastructures in Science

The methodological premise, i.e., one of the two foundational premises of this study, previously exposed[7], concerns the impact of digital ICTs and computational power on science. Digital technologies have, in fact, profoundly changed the way of doing science and the way of wondering about science[8]. As a result, the Open Science has been presented as a new paradigm of science that is emerging in these decades, represented by a new way of conducting science, precisely shaped by digital ICTs and computational power[9].

Since new technologies are at the basis of the Open Science (if not the very trigger), they gain fundamental relevance in designing and implementing the European strategies for scientific research: "[…] research can no longer be done without the use of machine-driven systems (hardware and software)"[10].

In this Chapter, the focus is on research infrastructures, and it is adopted the broad definition provided by the European lawmaker, in the recent Regulation EU 2021/695[11]:

> 'research infrastructures' means facilities that provide resources and services for the research communities to conduct research and foster innovation in their fields, including the associated human resources,

[7] See: Section 2.1.1.
[8] See: Section 2.1.1.2.
[9] On the definition of the Open Science paradigm, see: Section 3.2. In the provided interpretation the Open Science e-infrastructures are related to the Observable (O_4) "Tools".
[10] EOSC Executive Board, SRIA, *op. cit.*, p. 81.
[11] Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021, *establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013*, OJ L 170, 12.5.2021, p. 1–68, ELI: http://data.europa.eu/eli/reg/2021/695/oj.

> major equipment or sets of instruments; knowledge-related facilities such as collections, archives or scientific data infrastructures; computing systems, communication networks and any other infrastructure of a unique nature and open to external users, essential to achieve excellence in R&I [Research and Innovation]; they may, where relevant, be used beyond research, for example for education or public services and they may be 'single sited', 'virtual' or 'distributed'.

The "Strategic Research and Innovation Agenda" (SRIA), released in February 2021, by the EOSC Executive Board, proposed an *excursus* on the development of digital services and technologies available for researchers and scientists, since the 1970s[12]. The SRIA document illustrates that the first generation of the Internet infrastructures and its services (e.g., email, chat, files sharing, etc.) "[…] were only deployed within research communities […] to allow scientists to improve the way they collaborated"[13].

Then, a subsequent phase, with the spread of personal computers and the emergence of the UNIX operating system (1980s) was characterised by a mixture of public and private sectors: "[…] those software standards were developed within computer science laboratories both private and public (Unix, C and C++ at Bell Labs, TCP-IP at UC Berkeley, X Window System at MIT)"[14].

After, next crucial step was the birth of the World Wide Web, in the 1990s, at CERN in Geneva. The early Web 1.0 generated a "Web of documents"; later, in the 2000s, a "Web of data" began to be developed[15].

---

[12] EOSC Executive Board, SRIA, *op. cit.*, pp. 31-36.
[13] EOSC Executive Board, SRIA, *op. cit.*, p. 31.
[14] EOSC Executive Board, SRIA, *op. cit.*, p. 33.
[15] Here, with the emergence of Web 2.0, from a legal point of view, a strand of research starts to be developed, on the relationship between Web 2.0 and citizen political representatives, see: MARCO OROFINO, "The Web 2.0 and its impact on relations between citizens and political representatives." *Democratic Governance and Active Citizenship in the European Union, Saarbrucken: Lambert Academic Publishing* (2015): pp. 81-100.

Figure 6.1: "Table 1.1: EOSC in its technological context", in the SRIA Document (2021).

	1970	1980	1990	2000	2010	2020 >
European Infrastructures for Research	CERN (1954), ESO (1962), EMBL, ECMWF, ILL	NRENs, ESRF	DANTE, TERENA	ESFRI, ERICs, ESA, EPOS	GÉANT, EuroHPC, EGI, EUDAT, OpenAIRE, EIT Digital, ELIXIR, EUROfusion	E-ELT, AI-Data-Robotics, EOSC
Policy		Free Software		Open Access, Open Source, Open Data	Open Government	Open Science
Funding	Public	Public, Private	Public, Start-ups	Public, Private	Public, Private	Public, Private
Generic Services	Remote Login, Email, File Transfer, Chat	Office Automation, Teamwork (local), Shared File Systems, Databases	Browsers, Teamwork (Wiki), Discovery, Catalogues, Electronic Commerce	Encyclopedia, Social Networks, Telephone, Digital Photography, Television	Elastic Cloud, Electronic Publishing, Videoconferencing	
Standards	IETF	X Consortium	W3C	WAI	RDA	
Infrastructures	Internet	TCP-IP, X Window	Web of Documents, HTML, CSS, Mozilla,Apache, Internet Archive,arXiv	Web of Data, XML, Semantic Web, Web Services, Royalty-Free Policy, HAL	Linked Data Platform, Domain-Specific Standards	Next-Generation Internet, Distributed & Federated Clouds, Internet of FAIR Data, Solid, Software, Heritage
Technologies	Mainframes, Fortran, Cobol, Supercomputers, Leased Lines	Microprocessors, Workstations, Unix, C, C++, Graphics, Ethernet	PC, Mac, Linux, Java, JavaScript, ADSL	Mobile Phones, Fibre Optics (network), Wi-Fi, REST, Git, Python	Smartphones, Fibre Optics (last mile), Hadoop, Cloud Computing, Grid Computing, GitHub	Digital Continuum, Internet of Things, Machine Learning

The next decade, the 2010s, was marked by the emergence of smartphones and cloud computing, outlining unprecedented scenarios in terms of storage and services.

Finally, in the decade of the 2020s, the focus is on developments in machine learning (ML) algorithms, artificial intelligence (AI), Internet of Things (IoT) and Quantum Computing[16], advances that require more and more computational power.

Figure 6.1 represents a table, provided by the SRIA documents[17], that summarises technological development, highlighting the European research infrastructures involved at each stage. Among others, very relevant is ESFRI, the European Strategy Forum on

[16] As an illustration, consider the recent research project conducted by the Italian CNR (the national research council) and the University "Statale" of Milan and the Polytechnic Institute of Milan, about a quantum compiling based on a deep reinforcement learning method, see: LORENZO MORO, *et al.*, "Quantum compiling by deep reinforcement learning", *Communication Physics*, 4.178 (2021): 1-8, doi: 10.1038/s42005-021-00684-3.

[17] EOSC Executive Board, SRIA, *op. cit.*, p. 32.

Research Infrastructures, defined as "[…] a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach"[18]. ESFRI has been established in 2002, with a mandate from the European Council "[…] to support a coherent and strategy-led approach to policy-making on research infrastructures in Europe, and to facilitate multilateral initiatives leading to the better use and development of research infrastructures, at EU and international level"[19]. Recently the "EOSC Landscape Working Group" analysed the current projects conducted by ESFRI, which "[…] aims to contribute to an effective and efficient approach to e-infrastructure and services (vertical and horizontal) for European science and its competitiveness in the global scene building on existing research infrastructures and electronic infrastructures"[20].

The development of e-infrastructures for research has been defined by the different needs of the scientific community, without a centralised approach, similar to the establishment and development of the World Wide Web. As stated by Tim Berners-Lee, the inventor of the Web, "[P]hilosophically, if the Web was to be a universal resource, it had to be able to grow in an unlimited way. Technically, if there was any centralized point of control, it would rapidly become a bottleneck that restricted the Web's growth, and the Web would never scale up. Its being 'out of control' was very important"[21].

However, if a centralised control is not suitable, on the other hand a long-term strategy, capable of pointing the way forward, is essential. To some extent, this was the aim of the W3C, i.e., the Web Consortium, established in 1994, with the aim of promoting the Web and finding consensus and protocols to make it work on the Internet. Accordingly,

---

[18] https://www.esfri.eu/forum.

[19] https://www.esfri.eu/background.

[20] EOSC Executive Board Working Group (WG) Landscape, *Landscape of EOSC-Related Infrastructures and Initiatives*, v. 2 (Luxembourg: Publications Office of the European Union, 2020), doi: 10.2777/132181, p. 35.

[21] TIM BERNERS-LEE, *Weaving the Web. The original design and ultimate destiny of the World Wide Web*. (New York: HarperBusiness, 2000), p. 99.

"[…] a robust long-term vision is the most important prerequisite in order to successfully and sustainably build and operate a RI [Research Infrastructure]. As with any other vision, this vision also requires an adequate framework and must be embedded in a supportive policy driven environment to be successful. RIs [Research Infrastructures] are typically operational for several decades so they require continuous and stable support. Sufficient time and support must be given to the RI [Research Infrastructure] to fully unfold and develop its full potential"[22].

Among research infrastructures, an increasingly important role is played by computing infrastructures, which typically include High Performance Computing (HPC) facilities[23].

Considering the relevance of research infrastructures and HPC facilities, it is worthwhile to investigate whether and how the European institutions are promoting their development. In doing so, three remarks need to be pointed out.

First, the interventions of the European institutions have not always been linear. For this reason, the key measures in this field, both hard and soft law, are discussed below[24], with the aim of outlining the rather complicated evolution of the field.

Second, the problem is not only a lack of research infrastructures or digital capabilities, in the European Union. Rather, what is primarily problematic today is the mapping of all existing research facilities or platforms across Europe. As an illustration, recently a study on the EOSC National Structures has been released by the EOSC Secretariat[25]: the aim of this study was to assess the development of

---

[22] ESFRI Long-Term Sustainability Working Group, "Long-term sustainability of research infrastructures", *ESFRI Scripta Volume 2*, 2017, p. 1.
[23] EOSC Executive Board Working Group (WG) Landscape, *"Landscape of EOSC-Related Infrastructures and Initiatives", op. cit.*, p. 31.
[24] See: Section 6.2.2 and also Section 6.2.3.
[25] Sara Garavelli, Anu Märkälä, Iiris Liinamaa, "EOSC National Structures: an overview of the national EOSC coordination and engagement mechanisms in Europe", EOSC Secretariat, 2021, doi: 10.5281/zenodo.5602949.

EOSC at national level, mapping the current European research ecosystem and the national entities involved. This study indirectly shows that the first concern is not a lack of research facilities: the first and foremost action to be taken is to map and enhance existing ones[26]. Therefore, there is a need to coordinate existing facilities, rather than (or before) creating new ones. Consider that also from a legal point of view, all these EU research facilities, developed at local level, are facing the same problems.

Third, developing effective policies and strategies requires focusing on the technology, understanding its mechanisms, and its limits. This understanding is essential to enhance the link between technological development and Open Science. The comprehension of the technologies involved underpins the effectiveness of the human right to science: "The digital age, the most recent stage in an evolving *continuum* of ways in which technology has supported science, presents an opportunity to improve the conduct of research in multiple directions, including with regard to openness, speed of access to scientific results, reproducibility and multi-disciplinarity. This should result in better science, increased trust in science, and an improved ability to meet global challenges. However, this potential will only be realised if research infrastructures evolve to allow scientists to exploit, in an easy-to-use and integrated environment, the (vast amounts of) relevant data being produced"[27].

The analysis on the European institutions interventions' in promoting the development of research infrastructures focuses on the High Performance Computing capabilities, which are explored in next section.

---

[26] In this regard, the task of "[M]apping of the existing research infrastructures in Europe which are candidates to be part of the EOSC federation" is also one of the main purpose of the EOSC Landscape Working Group, see: https://www.eoscsecretariat.eu/working-groups/landscape-working-group.
[27] EOSC Executive Board, SRIA, *op. cit.*, p. 11.

6.2 High Performance Computing and Scientific Research

The analysis of the evolution of e-infrastructures in science has brought out the need to investigate the current strategies put in place by European institutions, as an essential aspect of the Open Science paradigm.

The landscape of the European e-infrastructures for research is so vast and complicated that it requires narrowing down the scope of research to provide an effective legal analysis. For this reason, the focus is on the High Performance Computing (HPC) facilities. The HPC is becoming an increasingly valuable asset for scientific research and much seems to be happening at European level in this domain.

In order to assess the European HPC strategies for scientific research, the investigation starts focusing on the concept of HPC, from a technical point of view (Section 6.2.1). Then, the aim is to map the evolution of the European projects on HPC (Section 6.2.2). Afterwards, the current European regulation on EuroHPC is analysed, clarifying the link between the use of HPC in research and the Open Science paradigm (Section 6.2.3).

6.2.1 High Performance Computing (HPC)

The concept of High Performance Computing (HPC) does not have a standardised and commonly accepted definition. The main reason lies in the very nature of the technology. HPC refers to the maximum computing power that can be achieved, so it is difficult, if not counterproductive, to crystallize a unique definition, taking into account the unstoppable technological development. The computational power increases, in a very short time, and as a consequence it would make any quantitative definition of HPC quickly obsolete.

In 2012, the European institutions linked HPC to supercomputing or high quality computing, representing it as "[…] multiple processors (tens, hundreds or even thousands) connected together by a network to achieve the performance well above that of a single processor"[28]. Similarly, the National Institute for Standards and Technology (NIST) identified "HPC systems in academic environments" as systems "[…] used to transfer data and to perform fast, parallel, and repetitive mathematical calculations for scientific problems"[29].

This dissertation adopts the following general definition of HPC:

> High Performance Computing (HPC) generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation, in order to solve large problems in science, engineering, or business.[30]

Three different types of HPC can be distinguished: (i) HPC can be achieved through supercomputers; or (ii) HPC is obtained through clusters of multiple small processors; and, (iii) an HPC environment can be also provided in a cloud computing architecture.

(i) HPC Supercomputers. HPC can be achieved through very powerful supercomputers or mainframes, which involve the provision of very expensive hardware. The measurement of a supercomputer's performance, in order to define it, varies considerably over time, in relation to the technological advancement. While in 2010 the machine with the highest computing power was able to perform 2.57 petaflops/s (quadrillions of calculations per second[31]), after a stagnation in development around 2014, in 2019 the world's most powerful supercomputer boasts 148 petaflops/s[32]. It is further discussed below

[28] European Commission, *High-Performance Computing: Europe's place in a Global Race*, COM/2012/45 final, p. 2, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012DC0045.
[29] NIST Working Draft, "An Action Plan for High Performance Computing Security", November 2016, p. 3, https://www.nist.gov/system/files/documents/2018/03/15/working_draft_actionplanhpc.pdf.
[30] Ludovica Paseri, Sébastien Varrette, Pascal Bouvry, "Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes." *Annual Privacy Forum* (Cham: Springer, 2021), p. 125, doi: 10.1007/978-3-030-76663-4_7.
[31] Flops, floating-point operation per second, represents row compute capability.
[32] https://www.top500.org/lists/2019/11/.

how the strategy of making massive investments in hardware is the one recently pursued by the European institutions[33].

(ii) HPC Clusters. Historically, around the 1990s, starting from a strong demand for computational power in scientific research, when machines reached the limit dictated by physics of 1 nanosecond, alternative solutions needed to be explored: such as parallel computing[34]. In this scenario, a multitude of processors, cooperating with each other to compute a specific task, has been able to significantly increase performances, represented by the clock speed of every single-processor machine.

(iii) HPC Cloud. Since the demand for computational power for scientific research continued to increase over the years, the solution adopted by some was to migrate to the cloud[35], since "[T]he scalability, flexibility, cost-effectiveness and relative user-friendliness of various cloud services make it also an attractive model to address computational challenges in the scientific community"[36]. In this scenario, virtualisation determines a separation of the services offered by the actual underlying physical hardware. This separation implies a reduction in the expenditure for the use of the same services in the cloud environment. HPC cloud facilities represent, precisely, the possibility to take advantage of the architecture of cloud computing to run traditional HPC applications[37].

HPC is a valuable asset for many different sectors. In business, HPC is becoming a fundamental resource[38]. On this dissertation the focus is

---

[33] See: Section 6.2.2.

[34] On the concept of parallel computing, see: SCOTT L. RIDGWAY, TERRY CLARK, BABAK BAGHERI, *Scientific parallel computing*. (Princeton: Princeton University Press, 2021). See, also: DAVID PADUA (ed.), *Encyclopedia of Parallel Computing*. (Heidelberg, Springer: 2011), doi: 10.1007/978-0-09766-4.

[35] A recent and successful example is provided by the University of North Carolina at Chapel Hill (UNC-Chapel Hill), http://herald.web.unc.edu/2019/11/how-cloud-computing-supports-green-technology/.

[36] ROMAN LEDYAYEV, HARALD RICHTER, "High performance computing in a cloud using openstack." *Cloud Computing* (2014): 108.

[37] MARCO A. NETTO, *et. al.*, "HPC Cloud for Scientific and Business Applications: Taxonomy, Vision, and Research Challenges." *ACM Computing Surveys*, 51.1 (2018): 8.2, doi: 10.1145/3150224.

[38] THOMAS ALSOP, "Revenue from broader high performance computing (HPC) market worldwide from

on the HPC usage in the research domain: "High-Performance Computing (HPC) is one of the strategic priorities for research and innovation worldwide due to its relevance for industrial and scientific applications"[39]. For this reason, next sections investigate the evolution of the EU projects on HPC and its relationship with Open Science.

6.2.2 European Projects on HPC

The European strategy to foster the development of HPC in Europe dates back to 2012: the European institutions believed that HPC could be a key element for innovation, making the EU a world leader in the race towards exa-scale computing[40]. The need to outline a strategy at the European level also emerged from the great demand for HPC, which came from many directions: from public administrations in matters of national security; from the public research sector, i.e., universities and research organisations; and, also from industry[41]. These requests, due to the scarcity of HPC, addressed beyond the EU borders[42].

2015 to 2024, by segment (in million U.S. dollars)", *Statista* (2021): "In 2020, the high performance computing (HPC) market generated total revenues of 23.98 billion U.S. dollars, of which the server market generated 11.85 billion U.S. dollars in revenue. The HPC storage market saw revenues reach 4.77 billion U.S. dollars.", available at: https://www.statista.com/statistics/724916/worldwide-broader-hpc-revenue-by-segment/. Less recently, see: EARL C. JOSEPH, *et al.*, "A Strategic Agenda for European Leadership in Supercomputing: HPC 2020," *IDC Final Report of the HPC Study for the DG Information on Society of the European Commission* (2010): p. 14, in which the Authors show that: "[…] 97% of large companies worldwide that had adopted HPC said they could no longer compete or survive without it".

[39] MARCO ALDINUCCI, *et al.*, "The Italian research on HPC key technologies across EuroHPC." *Proceedings of the 18th ACM International Conference on Computing Frontiers* (2021): p. 178, doi: 0.1145/3457388.3458508. Precisely in light of the relevance of HPC, in this article the Authors describe the Italian initiative of the National Interuniversity Consortium for Informatics (CINI), about the establishment of a laboratory which intends to enhance HPC developments, and also "[…] to propose a coordinated approach to HPC research within the EuroHPC Joint Undertaking, participating in the calls 2019-20 to five successful proposals for an aggregate total cost of 95M €".

[40] European Commission, *High-Performance Computing, op. cit.,* 2012*,* p. 12.

[41] European Commission, *High-Performance Computing, op. cit.,* 2012*,* p. 3.

[42] Note that in 2011 two leading HPC Countries, the United States and Japan, individually held greater HPC capacity than all EU Member States. See: European Commission, *High-Performance Computing, op. cit.,* 2012*,* p. 3 and https://www.top500.org/lists/2011/11/. See, also: GUSTAVE KALBE, "The European Approach to the Exascale Challenge." *Computing in Science & Engineering*, 21.1 (2019): pp. 43, doi: 10.1109/MCSE.2018.2884139, where the Author explains: "The use of HPC is growing rapidly for research and development purposes in both science and industry. Concrete industrial and commercial

In this scenario, the European institutions launched a strategy to foster the development of HPC in Europe, foreseeing its implementation within the Horizon 2020 project. This strategy pursued three strands, described below, which are: (i) the increase of the HPC infrastructures, for industry and public sector; (ii) the development of a new generation of HPC technologies; and (iii) the achievement of an applicative competence through the establishment of the so-called "Centres of Excellence" (CoEs)[43].

(i) HPC infrastructures. Regarding HPC infrastructures, there are two main drivers: PRACE and GEANT. PRACE Research Infrastructure (that stands for "Partnership for Advanced Computing in Europe") is a legal entity established in 2010, with the purpose of creating a single infrastructure meant to pool all the state-of-the-art information systems, for the entire European academic sector[44]. GEANT is another essential element in the e-infrastructures system made available in the EU, which has enabled the interconnection of national research and education networks, representing "[…] the largest and most advanced R&E [Research and Education] network in the world, connecting over 50 million users at 10,000 institutions across Europe and supporting all scientific disciplines"[45].

(ii) HPC technologies. The second strand of the 2012 European strategy on HPC was to develop new HPC technologies, in order to meet the challenge of the exa-scale and become competitive with the world's

applications are also emerging across a broad range of sectors, such as the automotive industry, renewable energy, climate change, and health. The availability of vast amounts of data supported by new techniques benefiting from artificial intelligence will mean that HPC resources will be more and more in demand in the coming years. Today the demand in Europe for HPC and data services already far exceeds what European public and private operators can supply. This imbalance has increasingly led researchers, industry professionals, and SME owners to rely on non-European service providers, in some cases located outside the EU, to carry out simulations and process their data".

[43] On project deployment see: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/high-performance-computing-hpc.

[44] https://prace-ri.eu.

[45] https://www.geant.org/About.

major HPC powers. The development of the HPC technologies was considered the means by which to deploy the necessary capabilities to cover the full range of processor architectures and applications, crucial to achieve the benefits of HPC. A very important role in this scenario is played by the "European Technology Platform for HPC" (ETP4HPC), a private and non-profit association under Dutch law, based in Amsterdam, whose members come from industry. The aim of the ETP4HPC Association is to promote the European HPC research and innovation through multiple support and development initiatives[46]. Another relevant actor, related to this second strand, is the "European Processor Initiative Consortium" (EPI Consortium): a consortium, composed of 27 partners from 10 EU Member States, which has signed an agreement with the European Commission, committing itself to accomplish the first phase of the implementation of the European ecosystem for HPC and Big Data, started in December 2018[47].

(iii) Application Expertise. Pursuing the goal of coordinating support in the application of HPC, specifically in the field of scientific and university research, the creation of the so-called "Centres of Excellence" (CoEs)[48] has been established. The purpose of these CoEs is to federate capabilities across Europe, taking advantage of interdisciplinarity. The aim was to have entities able to work in synergy with national and local programs, embodying a cooperative approach. The need to pursue this type of coordination is believed to be one of the main concerns of this domain[49]. However, once again, this endeavour is linked – also from a semantic point of view – to the concept

[46] https://www.etp4hpc.eu/who-we-are.html.
[47] https://www.european-processor-initiative.eu.
[48] Information on CoEs, and the related progress of European projects, can be found at: https://cordis.europa.eu/en, which is the website of CORDIS, the Community Research and Development Information Service, i.e., the European Commission's main source of information on the results of projects funded by the EU Framework Programmes for Research and Innovation.
[49] Oh this aspect, see: Section 6.2.3.

of 'excellence' in science, which has been much debated among scholars[50].

The first concrete step towards the implementation of the European strategy for HPC was the approval of the "Important Project of Common European Interest on HPC and Big Data enabled Applications" (IPCEI-HPC-BDA), in 2015. This ambitious project involved four Member States: Luxembourg (as leader), France, Italy, and Spain. This project took up the three strands already identified by the European Commission's 2012 HPC strategy[51]. Concerning the Centres of Excellence, the IPCEI-HPC-BDA Project declared the intention to "[…] develop and test HPC-enabled applications in specific strategic sectors at the regional, national and pan-European scale"[52].

The most significant development in promoting the potential of HPC in Europe was, however, certainly achieved in 2018, with the approval of the EuroHPC Joint Undertaking (EuroHPC JU) project. This project aimed to establish a legal and funding entity, with the aim to empower pooling of the EU and national resources in HPC. As set out in the

---

[50] On this debate, see: GEOFFREY BOULTON, "University rankings: Diversity, excellence and the European initiative." *Procedia-Social and Behavioral Sciences* 13 (2011): 74-82, doi: 10.1016/j.sbspro.2011.03.006; as regards the link between evaluation of research and excellence, see: ROBERTO CASO, *La rivoluzione incompiuta: La scienza aperta tra diritto d'autore e proprietà intellettuale*. (Milano: Ledizioni, 2019), p. 118; MARIA CHIARA PIEVATOLO, "La bilancia e la spada: scienza di stato e valutazione della ricerca." *Bollettino Telematico di Filosofia Politica* (2017), https://btfp.sp.unipi.it/it/2017/05/libric/; MATHIEU ALBERT, SUZANNE LABERGE, WENDY MCGUIRE, "Criteria for assessing quality in academic research: the views of biomedical scientists, clinical scientists and social scientists." *Higher Education* 64.5 (2012): 661-676, doi: 10.1007/s10734-012-9519-2; on the risk of limiting the development of the Italian university system, to the detriment of small university centres, generated by an overemphasis on the concept of excellence in science see: JUAN CARLOS DE MARTIN, *Università futura: tra democrazia e bit*. (Torino: Codice Edizioni, 2017), pp. 77, 182-188.

[51] In doing so, the IPCEI-HPC-BDA Project identified three main pillars: technology; infrastructures; and large-scale pan-european pilots. See: European Strategic Positioning Paper, "Important Project of Common European Interest on HPC and Big Data enabled Applications (IPCEI-HPC-BDA)", available at: http://knowledgebase.e-irg.eu/documents/243153/299805/IPCEI-HPC-BDA.pdf.

[52] European Strategic Positioning Paper, *"IPCEI-HPC-BDA", op. cit.*, p. 4. The declared aim was to address the so-called "Grand Societal Challenges", ranging from digitisation to climate change, embracing all fields of knowledge and the most diverse aspects of our societies.

Council Regulation 2018/1488 of 28 September 2018[53], the EuroHPC JU project is related to the European Cloud Initiative "[…] in order to gather the necessary resources and capabilities […] to close the chain from research and development to the delivery and operation of the exascale High Performance Computing system"[54].

The EuroHPC Joint Undertaking has been formed by the European Commission, as representative of the EU institutions, 33 European countries (but not limited to the EU[55]) and two associations, namely the aforementioned "European Technology Platform for High Performance Computing Association" (ETP4HPC) and the "Big Data Value Association" (BDVA)[56].

The EuroHPC JU has been established as a legal entity, based on the Article 187 of the TFEU[57], and its governance has been composed of three entities: (i) the Governing Board, (ii) the Industrial and Scientific Advisory Board, and (iii) the Executive Director.

(i) The Governing Board, composed of representatives of the EU and the 33 participating States, is responsible for strategic policy and funding decisions.

---

[53] Council Regulation (EU) 2018/1488 of 28 September 2018, *establishing the European High Performance Computing Joint Undertaking*, in OJ L 252, ELI: http://data.europa.eu/eli/reg/2018/1488/oj.

[54] Recital 8, Council Regulation (EU) 2018/1488. Concerning the choice of the legal instrument of the Joint Undertaking, Recital 11 states that "A Joint Undertaking represents the best instrument capable to implement the goals of the European High Performance Computing Strategy as set out in the European Cloud Initiative". About the "European Cloud Initiative", see: Section 2.2.3.

[55] Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Montenegro, the Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and Turkey.

[56] The Big Data Value Association (BDVA) is a non-profit association, made up of representatives from industry and research, which pursues the aim to "[…] develop the Innovation Ecosystem that will enable the data and AI-driven digital transformation in Europe delivering maximum economic and societal benefit, and, achieving and sustaining Europe's leadership on Big Data Value creation and Artificial Intelligence", avaliable at: http://www.bdva.eu/about. In 2014 BDVA has also set up a contractual public-private partnership with the European Union, with the aim of supporting European policies on Big Data.

[57] Consolidated version of the Treaty on the Functioning of the European Union (TFEU), OJ C 326, 26.10.2012, p. 47–390, http://data.europa.eu/eli/treaty/tfeu_2012/oj.

(ii) The Industrial and Scientific Advisory Board is formed by the "Research and Innovation Advisory Group" (RIAG) and the "Infrastructure Advisory Group" (INFRAG): the former is in charge of the regular updating of the draft of the multi-annual strategic agenda for research and innovation; while the latter has the task of providing advice, in an independent manner, to the Governing Board in the acquisition and implementation of petascale and pre-exascale supercomputers; the members of both Advisory Groups are private actors, stakeholders in the field of supercomputers.

(iii) The Executive Director is the chief executive as well as the legal representative, who manages the EuroHPC JU in its daily operations, in accordance with the decisions of the Board.

The EuroHPC JU aimed to build 5 petascale systems[58], 3 pre-exascale systems[59], 2 exascale systems and, subsequently, a post-exascale system by 2027. In doing so, the EuroHPC JU relies on the initial co-investment of the 33 participating States and the European Commission for a total of about EUR 1 billion, out of which EUR 486 million come from the EU. The private members will also provide additional contributions to the value of over EUR 400 million, through participation in the Joint Undertaking's activities. The Joint Undertaking provide financial support in the form of procurement or research and innovation grants to participants following open and competitive calls[60].

---

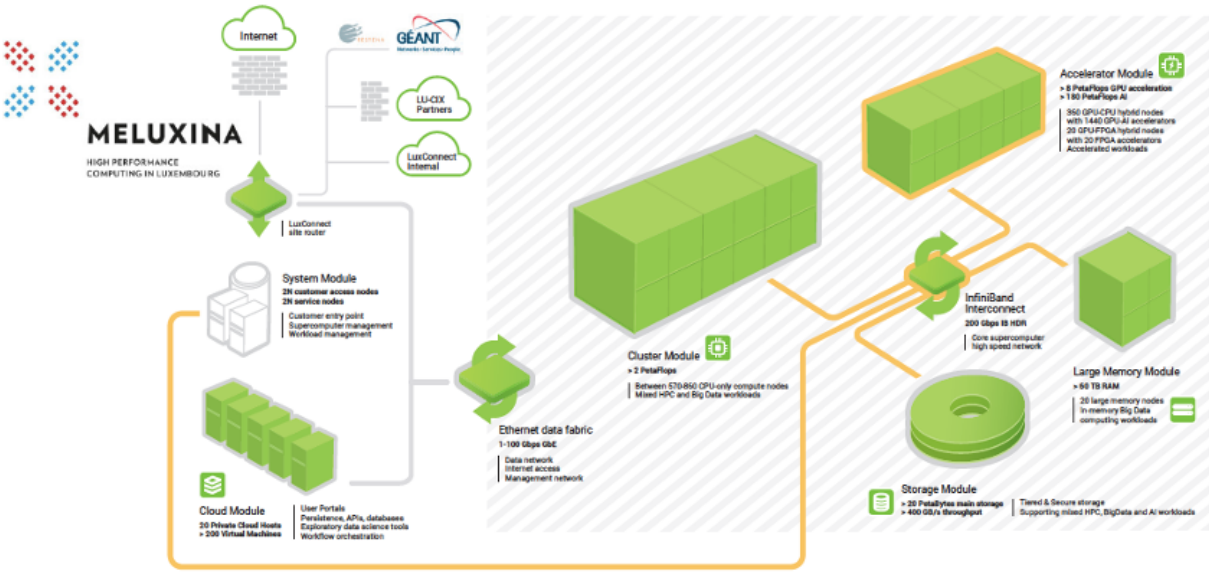
[58] Currently, they are located in Luxembourg ("Meluxina" in Bissen); Slovenia ("VEGA" in Maribor); Bulgaria ("Discoverer" in Sofia); Czech Republic ("Karolina" in Ostrava); and Portugal ("Deucalion" in Minho). See: https://eurohpc-ju.europa.eu/discover-eurohpc-ju.

[59] The three EuroHPC pre-exascale supercomputers are located in Finland ("LUMI" in Kajaani), Spain ("MareNostrum5" in Barcelona) and Italy ("LEONARDO" in Bologna). See: https://eurohpc-ju.europa.eu/discover-eurohpc-ju.

[60] The EuroHPC JU has launched in February 2019 two calls for hosting entities i.e., countries meant to host EuroHPC supercomputers: a call for Precursors to exascale systems (i.e., with 200 PFlops computing capacity), attributed to 3 consortium candidates (from Finland, CSC; Italy, Cineca; and Spain BCS); a call for petascale systems (i.e., between 2 and 100 Pflops), which was attributed to 5 sites (Czech Republic, Luxembourg, Portugal, Slovenia and Bulgaria). A current call is open for the procurement of two Exascale systems, with probable candidatures emanating from France and

Luxembourg applied for the “EuroHPC call for Petascale Supercomputers”, through a Consortium led by Luxconnect[61], in which are taking part many relevant actors, such as: the “Luxembourg Institute of Science and Technology” (LIST); the “Jülich Supercomputing Centre” (JSC); the “ParTec Cluster Compentence Center”; and the HPC team of University of Luxembourg, i.e., the ULHPC. The aim pursued by this Consortium was to implement a 10 PFlops system, called Meluxina, located in Bissen, Luxemborug.

Figure 6.2. The Structure of Meluxina (Varrette, Bouvry, Plugaru, 2018)



Germany. Subsequent calls for post-Exascale systems are expected by 2027.

[61] Luxconntect is a private company set up on the initiative of the Luxembourg Government in 2006, for the management of the fibre network and national data centres, see: https://www.luxconnect.lu.

On 7 July 2021, the Meluxina supercomputer became operational[62]. The structure of the HPC petascale system of Luxembourg, Meluxina, is described in Figure 6.2[63].

The participation of the ULHPC, an academic HPC facility developed for the benefit of researchers, reveals an intention to strengthen scientific research in the European Union. In addition, it shows also a strong synergy and interplay between public and private sector. The ambitious EU HPC strategy, which intends to establish a European supercomputing ecosystem is, from the outset, deeply connected with the world of scientific research. According to the SRIA document, the EOSC would be the link between the European HPC strategy and the domain of publicly funded scientific research: "EOSC will bridge this separation and help address the question of the relation between centralised and federated e-infrastructures"[64].

This scenario of technological developments represents a real paradigm shift: "Open Science is a transformative driver that will shape the research and innovation policies for a renewed European Research Area"[65]. The principles of openness and collaboration, which are at the basis of the Open Science paradigm, should always be taken into account as ordering and inspiring criteria in the implementation of the technical infrastructures. It seems fair to wonder whether this actually happens in practice. To further explore this aspect, attention should be drawn to the new HPC regulation, discussed in next section.

6.2.3 New HPC Council Regulation and Open Science

On 13 July 2021, a new Council Regulation establishing the European High Performance Computing Joint Undertaking was issued, repealing

---

62 https://eurohpc-ju.europa.eu/press-releases/meluxina-live-eurohpc-ju-supercomputer-luxembourg-operational.

63 SÉBASTIEN VARRETTE, PASCAL BOUVRY, VALENTIN PLUGARU, *et al.*, "Overview and Challenges of the UL HPC Facility at the EuroHPC Horizon," *University of Luxembourg, 7th High Performance Computing School*, Luxembourg (2018).

64 EOSC Executive Board, SRIA, *op. cit.*, p. 51.

65 EOSC Executive Board, SRIA, *op. cit.*, p. 47.

the previous Regulation 2018/1488[66]: the European institutions identify a new legal basis for the establishment of the EuroHPC JU.

This new Council Regulation is in line with the previous repealed Regulation. Therefore, the governance of the EuroHPC JU, as described in the previous Section 6.2.2[67], remains unchanged. In addition, the Article 38, entitled "Repeal", expressly states that actions carried out under the previous Regulation shall not be prejudiced, including specifically the annual implementation plans and the financial obligations[68].

The need to issue a new Council Regulation on this subject arose mainly in connection with the adoption of the new Regulation (EU) 2021/695 defining the framework of the European research funding, i.e., the Horizon Europe programme. In fact, the previous Council Regulation 2018/1488 deployed the EuroHPC Joint Undertaking under the previous European scientific research funding programme, i.e., the Horizon 2020. As a result, the new European funding programme made it necessary to innovate the legal basis of the EuroHPC JU. This need is spelt out in Recital 13, which states that "[T]he revision would also allow for the alignment of the EuroHPC Joint Undertaking's rules with the new legal framework, in particular Regulation (EU) 2021/695, as well as Regulations (EU) 2021/694 and (EU) 2021/1153"[69].

---

[66] Council Regulation (EU) 2021/1173 of 13 July 202, *on establishing the European High Performance Computing Joint Undertaking and repealing Regulation (EU) 2018/1488*, OJ L 256, 19.7.2021, ELI: http://data.europa.eu/eli/reg/2021/1173/oj.

[67] The governance is still composed of three entities: (i) the Governing Board, (ii) the Industrial and Scientific Advisory Board, and (iii) the Executive Director, pursuant to the Article 4 and Article 5 of the Annex to the Council Regulation (EU) 2021/1173. An addition among the participating associations is noted: the "Data, AI and Robotics" (DAIRO) association, registered under Belgian law with its registered office in Brussels, as established in the Article 2 of the Council Regulation (EU) 2021/1173.

[68] The Article 38, the Council Regulation (EU) 2021/1173.

[69] Recital 13, the Council Regulation (EU) 2021/1173. The reference goes to the Regulation (EU) 2021/695, establishing the Horizon Europe Programme; the Regulation (EU) 2021/694, establishing the Digital Europe Programme; and the Regulation (EU) 2021/1153, establishing the Connecting Europe Facility.

Going beyond the motivation to realign with the new legal framework, this new HPC regulation can be interpreted from three perspectives, further clarified below: (i) the evolution of the European HPC ecosystem; (ii) the recent political stance of the European Commission, in the technological field; and (iii) a remarkable emphasis on Open Science.

(i) EU HPC ecosystem evolution. Compared to the HPC strategy put in place in 2012, a lot has changed and been done, both from an organisational and a technical point of view. In 2019, in light of a survey involving about a hundred of the most relevant European HPC stakeholders, "the HPC in Europe portal"[70] has been introduced. The need for a specific dedicated portal arose precisely from the great development that the various HPC initiatives at European level had undergone in just a few years. Therefore, the survey revealed the need for clarity and coordination at European level. The aim was to facilitate access to the various types of resources as much as possible: "[T]he HPC in Europe portal has been designed to be complementary to the classification of HPC services and related activities according to different target audiences (i.e.: researchers, students, industry and projects)"[71].

In a short time, then, further profound developments have also taken place in the field of HPC research. As an illustration, consider the Italian initiative of the National Interuniversity Consortium for Informatics (CINI)[72], about the establishment of a laboratory which intends to enhance HPC developments, and also "[…] to propose a coordinated approach to HPC research within the EuroHPC Joint

[70] http://www.hpc-in-europe.eu.
[71] Florian Berberich, *et al.*, "European HPC Landscape." *2019 15th International Conference on eScience (eScience)*, 2019, pp. 478, doi: 10.1109/eScience.2019.00062.
[72] This initiative has already been mentioned, emphasising the value of HPC in scientific research. See, Section 6.2.1.

Undertaking, participating in the calls 2019-20 to five successful proposals for an aggregate total cost of EUR 95M"[73].

Another example of the development of the research on HPC is represented by a strand related to e-waste and eco-friendly HPC, suggesting that the EU should focus more on HPC systems where it is lacking, rather than on further hardware acquisition[74]. The Authors explain that "Europe could lead this transformation towards a more eco-friendly HPC approach: while in Europe we do consume a lot of the over all available HPC computing power, we are less active in the production of HPC systems. Europe could alleviate its dependency by extending the system lifetime and by investing more in services and software development than in hardware acquisition. Such a shift will not emerge instantaneously, it would need to be pushed and pursued actively by the different stakeholders within the HPC ecosystem"[75].

(ii) European Commission's stance on technology. Starting from the year 2020, a precise political stance of the European Commission in the technological field started to be identified: in Chapter 2, I defined it as the "von der Leyen Doctrine"[76]. The European Commission intends to put in place a strategy in the technological domain able to regain the EU lost ground from a geopolitical perspective. This purpose, already declared at the beginning of 2020 with the Communication n. 66 of 2020, entitled "A European Strategy for Data"[77], was then resulted in the release of a set of proposed legislative acts: the suite of proposals

[73] MARCO ALDINUCCI, *et al.*, *"The Italian research on HPC key technologies across EuroHPC.", op. cit.*, p. 1.

[74] MAIKE GILLOT, *et al.*, "Towards Sustainable HPC", (2021): 1-11, http://people.irisa.fr/Francois.Bodin/wp-content/uploads/2021/03/20210330Sustainable_HPC.pdf. On the difference between HPC systems and hardware, see: Section 6.2.1.

[75] MAIKE GILLOT, *et al.*, *"Towards Sustainable HPC", op. cit.*, p. 8.

[76] See, Section 2.2.5.

[77] European Commission, *A European strategy for data*, COM/2020/66 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066.

for regulations related to the governance of data[78], services[79] and markets[80], concluded with the already widely debated Artificial Intelligence Act[81].

In addition, equally relevant was the announcement of the release of a proposal for a Regulation on microprocessors, the so-called "European Chips Act", expected in 2022[82].

---

[78] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on European data governance (Data Governance Act)*, COM/2020/767, 25.11.2020, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767

[79] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, COM/2020/825 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN. On the implications of this proposal for a regulation, see: AINA TURILLAZZI, MARIAROSARIA TADDEO, LUCIANO FLORIDI, FEDERICO CASOLARI, "The Digital Services Act: An Analysis of Its Ethical, Legal, and Social Implications." *Legal, and Social Implications* (2022): 1-22.

[80] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on on contestable and fair markets in the digital sector (Digital Markets Act)*, COM/2020/842 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN.

[81] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM/2021/206, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206. On the ongoing debate about this proposal of Regulation, see: MARTIN EBERS, *et al.*, "The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)." in UGO PAGALLO, MASSIMO DURANTE (eds.) Special Issue. The Impact of Artificial Intelligence on Law, J 4.4 (2021): 589-603, doi: 10.3390/j4040043; MICHAEL VEALE, FREDERIK ZUIDERVEEN BORGESIUS, "Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach." *Computer Law Review International* 22.4 (2021): 97-112, doi: 10.9785/cri-2021-220402; and also: PHILIPP HACKER, "A legal framework for AI training data—from first principles to the Artificial Intelligence Act." *Law, Innovation and Technology* 13.2 (2021): 257-301, doi: 10.1080/17579961.2021.1977219.

[82] As announced in European Commission, *Commission work programme 2022. Making Europe stronger together*, COM(2021) 645 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2021:645:FIN: "Despite many challenges and disruptions, Europe came through the crisis in large part due to its innovative skills, its strong industrial base and its diversified and competitive supply chains. However, in a few strategic sectors, it has been vulnerable due to high dependency on a very limited number of non-EU suppliers, especially in relation to raw materials. This is particularly apparent when it comes to semi-conductors. Supplies of these chips which power Europe's digital solutions have become a real concern for EU industry, with cases of production being slowed down. Against this background, we will adopt a European chips act to promote a state-of-the-art European chip ecosystem to boost our innovative capacity, security of supply and develop new markets for ground-breaking European tech.", p. 4-5.

At the same time, there has been a considerable economic investment in funding, which is needed to make the initiatives operational. Regarding the EuroHPC JU, the Article 15 of the Council Regulation establishes the allocation of EUR 150 million for the period 2021-2027 to the HPC strategy[83].

This wide-range and structured set of European initiatives seems designed to compensate the EU lack of power in the technological sphere, lost after the birth of the Web[84], mainly to the advantage of the United States. The concept behind this attitude, sometimes expressed directly, sometimes indirectly, is the so-called "technological sovereignty"[85].

It seems fair to wonder whether this approach can or will be effective. Some remarks on this aspect have already been expressed previously[86]. In addition, for instance, consider the context of cloud computing. Recent statistics show that "[…] since the beginning of 2017 the European cloud market has grown almost fourfold, reaching EUR 7.3 billion (US$8.8 billion) in the second quarter of 2021"[87]. However, it is worthwhile to underline that the market share of the European cloud services providers has decreased by 27%, to the benefit of three major actors, such as Amazon, Microsoft and Google. This doesn't represent a concern in itself, but in light of the explicit political stance declared by the European Commission, it denotes that – at the moment – the EU has much to do. Moreover, it seems that this scenario requires more effort than economic investment.

(iii) New attention on Open Science. As previously said, the development of the European strategy on HPC is deeply connected with

[83] The Article 15(1), the Council Regulation (EU) 2021/1173. Consider that this amount of the EU investment is only for the upgrade of EuroHPC supercomputers.
[84] See: Section 6.1.
[85] The concept of the so-called "technological sovereignty" has already been discussed, in Section 2.2.3.1, investigating the relationship between EOSC and Gaia X project.
[86] See, again, Section 2.2.3.1, where the Gaia X project was discussed.
[87] Synergy Research Group, "European Cloud Providers Double in Size but Lose Market Share", 2021, https://www.srgresearch.com/articles/european-cloud-providers-double-in-size-but-lose-market-share.

the domain of research and universities, from the outset. A benchmark in the European research domain is represented by the Regulation (EU) 2021/695 defining the framework of European research funding, i.e., the Horizon Europe programme[88], for the period 2021-2027. In particular, the Article 14, entitled "Open Science", establishes that the research funding programme Horizon Europe shall be based on Open Science, understood as "[…] an approach to the scientific process based on cooperative work and diffusing knowledge". This "approach" or, as understood in this dissertation, this real paradigm shift represented by the emergence of Open Science, also has consequences on the project related to the EuroHPC JU and, more generally, on the entire European HPC strategy. As a result, the Recital 38 of the new Council Regulation states that "[T]he Joint Undertaking should also ensure the interconnection of the federated, secure supercomputing, and quantum computing service and data infrastructures with the common European data spaces, including the European Open Science Cloud, and federated, secure cloud infrastructures announced in the Communication from the Commission of 19 February 2020 entitled 'A European Strategy for Data', for seamless service provisioning to a wide range of public and private users across Europe"[89].

Already in the previous Council Regulation on the EuroHPC JU, a precise connection with the EOSC already emerged[90], but the real novelty is found in the new Council Regulation, precisely in its Recital 51. This Recital provides indications on the allocation of users' time on HPC facilities. This aspect is crucial, because it is the organisational

[88] On the Horizon Europe programme, see: Section 2.2.6.

[89] Recital 38, the Council Regulation (EU) 2021/1173.

[90] Recital 8, the Council Regulation (EU) 2018/1488, stated: "[…] a mechanism should be set up at Union level to combine and concentrate the provision of support to the establishment of a world-class European High Performance Computing infrastructure and for research and innovation in High Performance Computing by Member States, the Union and the private sector. This infrastructure should provide access to the public sector users, users from industry, including small and medium-sized enterprises (SMEs), and users from academia, including the scientific communities of the emerging European Open Science Cloud".

criterion by which computational power is provided in HPC ecosystems: the user sends a request after which she is put in a queue; in her turn, she is granted the required time, corresponding to the computational resources needed for her operations. Recital 51, chiefly, states that "[U]ser allocation of access time to the supercomputers of the Joint Undertaking should be free of charge for public users. It should also be free of charge for private users for their applications related to research and innovation activities funded by Horizon Europe or the Digital Europe Programme, as well as for private innovation activities of SMEs, where appropriate"[91]. In stating this, it effectively establishes a bridge to scientific research, allocating part of the computational power of national supercomputers to research, by default. In addition, Recital 51 precises that "[…] all users benefiting from free-of-charge access time to the supercomputers of the Joint Undertaking should adopt an open science approach and disseminate knowledge gained through that access, in accordance with Regulation (EU) 2021/695"[92]. As a result, the Open Science tenets represent the guiding principles: the Open Science goes beyond the strictly public research sphere to become a cornerstone of the European scientific and technological research and innovation system[93].

Although the EuroHPC JU project is of central importance, it must be stressed that it is only one part of the HPC ecosystem at the European level: there is, in fact, strong development at the local level, i.e., in universities and research centres. Attention is drawn specifically to a thriving example of these local entities, i.e., the University of Luxembourg's HPC platform (ULHPC), the case study of this dissertation. However, before proceeding in this direction, a

[91] Recital 51, the Council Regulation (EU) 2018/1488.
[92] *Ibid.*
[93] Here, the connection established in 2017 seems to be taking shape. The reference goes to the mid-term review on the implementation of the Digital Single Market Strategy, see: Section 2.2.4.

challenging issue needs to be dwelt: the US regulation of the CLOUD Act.

6.3 Public-Private Interplay and the CLOUD Act

The analysis of the European HPC strategy shows that the development of the European HPC ecosystem is based on a strong interplay between private and public sector. Similarly, as stated by the EOSC Symposium 2021, "[A] crucial component of EOSC will be services provided by commercial organisations, such as commercial cloud services"[94]. Section 6.4.1 clarifies the extent to which this relationship between private and public sector may be problematic. Subsequently, Section 6.4.2 sheds light on the consequences of this public-private interplay when the private partners are US-based companies.

6.3.1 Public-Private Interplay in Science

In this dissertation, in several circumstances, an intermingling between private and public sector in the domain of scientific research has been pointed out[95]. As has already been observed, this relationship between the public and private sectors in research has always existed and is even essential to some extent[96].

---

[94] VERONICA BERTACCHINI, *et al.*, *EOSC Symposium Report*, 2021, p. 19, Zenodo, doi:10.5281/zenodo.5176089. Here, in particular, the participation of the private sector was taken into account from an economic point of view, related to the cost of the services for end-users, e.g., researchers: "One mechanism for delivering these free at the point of use is the 'service voucher' provided to individual researchers for purchase of such services".

[95] See: Section 4.2.3; Section 4.3.1; Section 4.2.4.2. But also, related to the data protection issues, see: Section 5.1 and Section 5.4.

[96] See: Section 3.4.2, in which the case of the failure of scientific research in the former Soviet Union, due to the absence of a corresponding market, was described, mentioning the interpretation of Nielsen: "The importance of the market to the role of science is vividly illustrated by what happened when the market was suppressed in the Soviet Union. Although the Soviet Union had one of the best scientific research systems in the world, without a market system it was mostly unable to make scientific innovations available to its citizens.", in: MICHAEL NIELSEN, *Reinventing discovery. The new era of networked science.* (Princeton: Princeton University Press, 2012), p. 158.

However, the emergence of the Open Science paradigm and the COVID-19 pandemic made this reciprocal interaction even more evident[97].

Currently, private actors can offer a very high-level infrastructural services, also essential for publicly funded research. As argued in a recent study on health and medical research, that can be easily generalised, "[…] traditional stakeholders in the field of health research […] have control of highly specialized and sophisticated health data sets, which represent the very core asset of scientific enquiries. Conversely, big tech companies, such as Google, appear to offer the algorithmic infrastructure needed for the treatment of these sophisticated data sets, the generation of new digital information, and the enactment of statistical analyses and predictions. The complementary nature of such differently owned assets triggers health data sharing agreements, gathering together various types of health data – ranging from more sophisticated clinical data to 'real world' health data – under a common processing technology"[98].

Sometimes the contribution of the private and public sectors is easily identifiable. Other times, however, this collaboration is less evident. Consider a case related to the COVID-19 pandemic. In March 2020, in the early stages of the pandemic, the "Allen Institute for Artificial Intelligence" in Seattle, in collaboration with other entities, released the first version of "CORD-19", an acronym for "COVID-19 Open Research Dataset". CORD-19 is a collection of scientific publications on the SARS-CoV-2 virus and related topics[99]. The initiative is presented

---

[97] The COVID-19 pandemic and the involvement of pharmaceutical companies in vaccine research and development have renewed this debate also among the public at large, to the extent that has also become a topic in cinema, such as the Oscar-nominated movie "Don't look up" (2021).

[98] GIULIA SCHNEIDER, "Disentangling health data networks: a critical analysis of Articles 9(2) and 89 GDPR." *International Data Privacy Law*, 9.4 (2019: 270, doi: 10.1093/idpl/ipz015. The Author, then, starting from this assumption, concludes by arguing for greater coordination and networking between the actors involved: "Against this backdrop, the analysed case studies confirm the statements made by some scholars, stressing how the diversity of (scientific) 'knowledge requirements and the more complex technology frontiers imply a need for networks'", p. 270.

[99] On CORD-19, see: LUCY LU WANG, *et al.*, "Cord-19: The covid-19 open research dataset." *ArXiv* (2020);

in an OECD Report on Open Data as a public health project[100]. Such a platform is certainly a valuable resource for carrying out scientific research, related to the domain of public health, making possible collaboration between different fields of knowledge and public authorities, but presented in this way it somehow gives the idea that CORD-19 is a public initiative. However, there are many partners in this project, including the Chan Zuckerburg Initiative (CZI), Microsoft's research department and the Alphabet-owned company Kaggle[101].

The risks arise in cases in which the interplay is opaque. Public and private actors traditionally have different legal regimes and in some cases the latter are not much freer than the former[102].

However, the field of scientific research often tends to reduce the difference between the public and private sectors. Consider the case of personal data protection, investigated in Chapter 5. According to the General Data Protection Regulation (GDPR)[103], the definition of scientific research is very broad and also includes research carried out by the private sector[104]. In the case of data protection, processing of personal data conducted for research purposes may benefit from a

---

and also, GIOVANNI COLAVIZZA, *et al.*, "A scientometric overview of CORD-19." *Plos one* 16.1 (2021): 1-18, doi: 10.1371/journal.pone.0244839.

[100] OECD, *Open Data in action. Initiatives during the initial stage of the COVID-19 pandemic*, March 2021, p. 11, https://www.oecd.org/gov/digital-government/use-of-open-government-data-to-address-covid19-outbreak.htm.

[101] "[…] the Allen Institute for AI (AI2), in collaboration with our partners at The White House Office of Science and Technology Policy (OSTP), the National Library of Medicine (NLM), the Chan Zuckerburg Initiative (CZI), Microsoft Research, and Kaggle, coordinated by Georgetown University's Center for Security and Emerging Technology (CSET)", see: LUCY LU WANG, *et al., "Cord-19: The covid-19 open research dataset.", op. cit.*, p. 1.

[102] In partial opposition to a strand of research related to the so-called "surveillance capitalism", see: SHOSHANA ZUBOFF*, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019).

[103] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, *on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ L 119, 4.5.2016, p. 1–88, ELI: http://data.europa.eu/eli/reg/2016/679/oj.

[104] See: Section 5.2.1, in which the reference goes to Recital 159 of the GDPR.

considerable privileged legal framework, due to the derogatory discipline that the GDPR establishes for scientific research. Therefore, bringing activities within the scope of research can be a competitive advantage for companies.

Another risk of the opacity of the public-private interplay is that the logic of the market is obviously and justifiably different from the logic of science[105]. Consider, as an example, the case of the company Slideshare. Slideshare is a company that provides storage services for slide presentations. This company, owned by Microsoft, was sold in 2020 to the company Scribd. From September 2021, Slideshare contents became accessible only upon payment. Suppose the case of a researcher who for years has been using Slideshare's services for conference presentations and her research and teaching activities. Suddenly, she would have to pay to access her content, or similarly, people interested in her presentations would have to pay to access them, despite the fact that these presentations are the result of publicly funded research. Admittedly, relying on a proprietary service to generate research-related content is not the best strategy. But it is also true that the fact that many online services are free of charge, even the most basic ones, has for years led to the belief that these services were not private, and therefore neither could them be subject to the business logic.

For this reason, in scientific research more than in many other fields, it is crucial to make the interplay between the public and private sectors as transparent as possible. The aim of the EU strategies in scientific research should avoid an *opaque* interplay, *a fortiori*, considering that a large part of the private actors, very efficient and able to offer the most innovative infrastructures, are not European.

The problem should not be framed in the alternative between participation or exclusion in science of private actors. Rather, the

[105] On the logic of science and the meaning of science in society see: Chapter 1, with references to Merton, Polanyi, and Habermas.

trade-off is between transparent or opaque interplay between public and private actors in scientific research[106].

6.3.2 CLOUD Act

From a legal point of view, the public and private sector interplay in the scientific domain may arise some challenges when it is opaque and involves private partners located outside the EU, especially in the US. In that case, in fact, it should be considered the so-called "Clarifying Lawful Overseas Use of Data Act" or CLOUD Act, enacted by the Congress of the United States, in 2018[107], by amending the Chapter 121 of the Title 18 of the United States Code and the Stored Communication Act (SCA)[108].

This piece of legislation is relevant to the European perspective because supports the US public authorities to overcome "[…] inability to access data stored outside the United States that is in the custody, control, or possession of communications-service providers that are subject to jurisdiction of the United States"[109]. At the basis, the *ratio* of the CLOUD Act was to solve a conflict generated in the judicial case *Microsoft Ireland v United States*[110]. The case was raised in 2013, when the US federal agents issued a warrant requiring Microsoft to disclose all e-mails and other information associated with a user's account, on the grounds that the individual was suspected to be involved in an illegal drug trade. Microsoft denied access, claiming that the material

[106] Similarly, in MARIAROSARIA TADDEO, LUCIANO FLORIDI, *The Moral Responsibilities of Online Service Providers*. (Cham: Springer, 2017), p. 13, doi: 10.1007/978-3-319-47852-4_2, the Authors argued that "[…] there is a lack of an ethical framework that can (a) define OSPs' [*Online service providers*] responsibilities, and (b) provide the fundamental sharable principles necessary to guide OSPs' conduct within the multicultural and international context in which they operate." This is *a fortiori* even more true and evident regarding the activities of private service providers in the field of scientific research.
[107] Clarifying the Lawful Use of Overseas Data Act, CLOUD Act, Pub. L. No. 115–141, 132 Stat. 348 (codified as amended in separate sections of 18 U.S.C.), 2018, available at: https://cli.re/BwPk5Q.
[108] The Stored Communication Act (SCA) is the Title II of the Electronic Communications Privacy Act enacted in 1986.
[109] CLOUD Act, §2201 (13-17).
[110] US Supreme Court, *Microsoft Ireland v United States*, 583 U.S. (2018), https://supreme.justia.com/cases/federal/us/584/17-2/.

was stored in Microsoft's data centers located in Europe, precisely in Ireland. The Supreme Court got a chance to address the case[111].

In the meantime, in March 2018, the Congress – with bipartisan approval[112] – enacted the CLOUD Act.

The CLOUD Act introduced the following requirement: "[A] provider of electronic communication service or remote computing service shall comply with the obligations of this chapter to preserve, backup, or disclose the contents of a wire or electronic communication and any record or other information pertaining to a customer or subscriber within such provider's possession, custody, or control, regardless of whether such communication, record, or other information is located within or outside of the United States"[113]. In this manner, the US service providers are obliged, upon a warrant, to disclose data, despite the fact that they are located outside the US territory, regardless of whether they are personal data or not[114].

---

[111] The court case *Microsoft Ireland v United States* ends with a Supreme Court ruling, issued on 16 April 2018: "The Supreme Court vacated, finding the case moot. No live dispute remains between the parties over the issue with respect to which certiorari was granted; a new warrant replaced the original warrant" US Supreme Court, *Microsoft Ireland v United States*, 583 U.S. (2018), https://supreme.justia.com/cases/federal/us/584/17-2/. Consider that, specifically, Microsoft has set up a system that would indeed seem to elude the CLOUD Act: "[…] Microsoft has introduced a 'data trustee' model, whereby it puts customer data beyond the jurisdictional reach of US authorities by partnering with foreign companies. In 2015, Microsoft announced that it put Deutsche Telekom in charge of its German datacenters. According to this strategy, Deutsche Telekom's subsidiary, T-systems, operates and transacts Microsoft's data-centers in Germany independent of Microsoft. Microsoft hands over control of cloud customer data to this 'trustee' and customers need to sign a contact with T-systems, not with Microsoft. This arrangement could create a situation where personal data concerning a US person and required for US domestic crime investigation purpose is neither located in the USA nor effectively controlled by a US company", in, HALEFOM H. ABRAHA, "How compatible is the US 'CLOUD Act' with cloud computing? A brief analysis." *International Data Privacy Law* 9.3 (2019): 208, doi: 10.1093/idpl/ipz009. It would be interesting to assess the application of this Microsoft solution to other situation, for example, EOSC or Gaia X.

[112] On the legislative history of the CLOUD Act, see: MIRANDA RUTHERFORD, "The CLOUD Act: Creating executive branch monopoly over cross-border data access." *Berkeley Technology Law Review*, 34 (2019): pp. 1186-1193, doi: 10.15779/Z387940V34.

[113] CLOUD Act, §2202-2203 (24-25, 1-7).

[114] Consider that while the definition of personal data is very broad in the EU, the situation is different in the US, where the concept of "Personal Identifiable Information" does not have a single commonly accepted definition. On this aspect, see: PAUL M. SCHWARTZ, DANIEL J. SOLOVE, "Reconciling personal

Given the empirical interplay between public and private sector in the field of scientific research and the impact of CLOUD Act on the EU, I identify two problematic issues, that need to be further analysed below: (i) the conflict between the CLOUD Act and the GDPR; (ii) the role played by service providers for scientific research, in light of the CLOUD Act.

(i) Conflict between CLOUD Act and GDPR. The enactment of the CLOUD Act has an impact on the management of electronic evidence and on data sharing beyond the EU. From a legal point of view, however, it does not play an innovative role, since the CLOUD Act "[…] represent a classic case of international lawmaking via domestic regulation, as mediated by major multinational corporations that manage so much of the world's data"[115]. The aim is to bypass the international law mechanism of Mutual Legal Assistance Treaty (MLAT)[116].

The most relevant aspect is that the CLOUD Act generates some tension with the European data protection rules established in the GDPR, with specific reference to the issue of transfer of personal data to third countries[117]. The Article 48 of the GDPR, entitled "Transfers or disclosures not authorised by Union law", expressly provides that there may be transfers to third countries if they are based on orders from foreign courts, i.e., in the context of judicial proceedings in third countries. However, the GDPR specifies that these orders are subject to the following condition: "Any judgment of a court or tribunal and any

information in the United States and European Union." *California Law Review* 102 (2014): 877-916. On the extension of the EU definition of personal data, also see: Case C-434/16, *Peter Nowak v Data Protection Commissioner* (2017) ECJ, ecli:EU:C:2017:994.

[115] JENNIFER DASKAL, "Microsoft Ireland, the CLOUD Act, and International Lawmaking 2.0." *Stanford Law Review Online* 71 (2018): p. 9.

[116] The CLOUD Act introduces a new system that is more streamlined and immediate than MLATs. It provides for so-called 'executive agreements' between the US and foreign states to facilitate the collection of electronic evidence held by US service providers located in third countries. See: CLOUD Act, §2209(3-10). In particular, in 2019, US and UK established the first executive agreement under the CLOUD Act.

[117] This issue has been already analysed in Chapter 5, in particular, in Section 5.3.3.

decision of an administrative authority of a third country requiring a controller or processor to transfer or disclose personal data may only be recognised or enforceable in any manner *if* based on an international agreement, such as a mutual legal assistance treaty, in force between the requesting third country and the Union or a Member State"[118]. Currently, however, there is no international agreement in place between the US and the EU, following the ruling of the European Court of Justice (ECJ), Schrems II[119]: at the moment, considering the protection of personal data, there is a real legal vacuum[120].

However, the Article 49 of the GDPR exceptionally allows the transfer for "the establishment, exercise or defence of legal claims"[121]. It remains unclear under what conditions these transfers to third countries for legal claims are legitimate, e.g., whether a warrant is required; whether it is already sufficient at the investigation stage; whether a MLAT is not requested, etc.

Among scholars, in the aftermath of the Schrems II judgment, the recommendation for companies was to adopt technical solutions that would allow data to be physically located in data centres within the EU[122]. These kind of measure is suitable to limit the flow of personal data transfers to third countries and as a consequence to comply with the GDPR in the post-Schrems II uncertainty. However, these

[118] The Article 48(1), GDPR.

[119] Case C-311/18, *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems* (2020) ECJ, ECLI:EU:C:2020:559. On this aspect see, Section 5.3.3.1.

[120] Xavier Tracol, " 'Schrems II': The return of the Privacy Shield." *Computer Law & Security Review* 39 (2020): p. 8, doi: 10.1016/j.clsr.2020.105484. Consider that, in general, about the legal collaboration, see: European Union, *Agreement with the United States on mutual legal assistance* (2019), available at: https://eur-lex.europa.eu/summary/EN/jl0052, (developed under the CLOUD Act).

[121] The Article 49 (1) e, of the GDPR. Some scholars identify also the Article 49(1) let. d "public interest" and let. g "legitimate interest". However, it is not believed these two cases can represent a possibility in our case, related to the request of a foreign Country. In this regard, see: Jennifer Daskal, *"Microsoft Ireland, the CLOUD Act, and International Lawmaking 2.0."*, *op. cit.*, p. 12.

[122] Or where not possible to opt for "[…] the use of a strong and secure system of encryption as an appropriate technical and organisational measure to protect the personal data.", in Xavier Tracol, *"Schrems II: The return of the Privacy Shield."*, *op. cit.*, p. 11.

measures can do little against the CLOUD Act. This federal law has been appropriately defined a "legal shortcut"[123].

(ii) The role of service providers for scientific research. A further complexity is related to the identification of the actual recipient of the search warrant. Some scholars have pointed out that in a cloud computing architecture, there may be a multiplicity of actors involved, corresponding to a different degree of control over the infrastructure or data: "[T]his complicates the task of law enforcement agencies in identifying who is in 'possession, custody, or control' of personal data in view of serving a subsequent order"[124]. If this type of problem arises in a cloud environment, it can only be amplified in a cloud connected HPC ecosystem or in a federated resource environment such as the EOSC.

Indeed, the role played by service providers becomes essential in this context: "[S]ervice providers have to assess every request on a case-by-case basis and decide whether a request from qualifying foreign government complies with the CLOUD Act, with domestic laws of the requesting government, and of course with their internal policies"[125]. Similarly to what the ECJ established in the Google Spain case[126], regarding the right to be forgotten[127], the US Congress also seems to

---

[123] MARCIN ROJSZCZAK, "CLOUD act agreements from an EU perspective." *Computer Law & Security Review* 38 (2020): p. 13, doi: 10.1016/j.clsr.2020.105442.

[124] HALEFOM H. ABRAHA, *"How compatible is the US 'CLOUD Act' with cloud computing?", op. cit.*, p. 208.

[125] HALEFOM H. ABRAHA, *"How compatible is the US 'CLOUD Act' with cloud computing?", op. cit.*, p. 211.

[126] Case C-131/12, *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* (2014) ECJ, ECLI:EU:C:2014:317. On this ruling, see: MARCO OROFINO, "Trattamento dei dati personali e libertà di espressione e informazione." in LICIA CALIFANO, CARLO COLAPIETRO, *Innovazione tecnologica e valore della persona*. (Napoli: Editoriale Scientifica, 2017), pp. 531-533.

[127] MASSIMO DURANTE, *Computational Power: The Impact of ICT on Law, Society and Knowledge*. (New York: Routledge, 2021), pp. 61-62. The Author describes the intervention of the ECJ as a "private enforcement of a right", explaining that "[I]n the case of the right to be forgotten, the European Union has not only ordered Google to comply with European law; it has essentially handed off enforcement of the right in the first instance to Google. In doing so, it has in fact outsourced a task and given rise to a singular private enforcement of the right to be forgotten, due to the fact that the large private

grant a "quasi-judicial power"[128] to service providers. It seems legitimate to wonder whether or to what extent this kind of assessment should be in the hands of a private actor. If "[…] the CLOUD Act left users with no option but to trust that the service providers will protect them from illegitimate government requests"[129], the situation is even more complex in the domain of scientific research. In this context, the users involved are researchers, who in turn rely on service providers for the processing of personal data of data subjects, for whom they (or the institution or university under which they work) act as controllers, pursuant to the Article 4(7) of the GDPR. In this delicate scenario, trust plays a fundamental role, even more relevant than in a relationship between a private company-controller and user- or consumer-data subjects[130].

Admittedly, the scope of the CLOUD Act is limited to the "United States person", establishing that "[A] provider […] may file a motion to modify or quash the legal process where the provider reasonably

---

enterprises that constitute the digital infrastructure have the technical and bureaucratic capacity to regulate and govern speech, through blocking, filtering, and removing content, through otherwise controlling access to their facilities, and through digital surveillance". Similarly, this aspect has been illustrated in: MARIAROSARIA TADDEO, LUCIANO FLORIDI, "[T]he debate on the moral responsibility of online service providers." *Science and Engineering Ethics* 22 (2016): 1575-1603, doi: 10.1007/s11948-015-9734-1. The Authors underline the fact that: "[T]he academic interest in these topics stems from the pressing need felt by society to regulate OSPs' conduct in order to ensure the respect of the public good and the fostering of societal welfare. Such a need is often addressed by endorsing an *ad hoc* approach and by delegating to OSPs normative decisions. A good example of the case in point is offered by Google, which is currently both the "judge and the jury" with respect to the application of the right to be forgotten in Europe.", p. 1597.

128 HALEFOM H. ABRAHA, *"How compatible is the US 'CLOUD Act' with cloud computing?", op. cit.*, p. 211.

129 *Ibid.*

130 In the case under investigation in this section, i.e., the scenario outlined by the CLOUD Act, the first type of fiduciary relationship that emerges is between human agents, i.e., the controller, in terms of a university or research centre, who relies on the service providers, i.e., the processors, in order to protect the data subjects involved. But there is also another fiduciary relationship, that of the data subjects, towards the research organisation that processes their data for research purposes. In this second relationship, the trust may depend on the technologies involved, or on the anonymisation and encryption techniques envisaged. On trust related to the use of digital technologies see: MARIAROSARIA TADDEO, "Trusting Digital Technologies Correctly." Minds & Machines 27 (2017): 565-568, doi: 10.1007/s11023-017-9450-5.

believes that that the customer or subscriber is not a United States person and does not reside in the United States"[131]. Nevertheless, in the first place, this assessment of reasonableness is left entirely to the provider, who, as a private actor, proceeds with its own evaluations, and to some extent without having to account for its choices[132]. In the second place, the extraterritoriality of the GDPR means that the its scope of application is not limited solely to the European citizens, pursuant to the Article 3 of the GDPR.

One element of risk is related to the breadth of the notions involved. The incipit of the CLOUD Act states that "[T]imely access to electronic data held by communications-service providers is an essential component of government efforts to protect public safety and combat serious crime, including terrorism"[133]. Although in principle this is an entirely supportable statement, from a legal point of view it raises concerns related to the definition of the terms "serious crime" and "terrorism", which are not defined in the federal law. The "war on terror" of the early 2000s, conducted by the administration of the then President Bush, with the suite of measures linked to the Patriot Act and the related threats to the privacy and personal data protection of the US citizens and others, represent an evident precedent[134].

---

[131] CLOUD Act §2204 (9-21)

[132] Here the intention is to point out the distinction between the obligation to provide the motivation imposed on an administrative or public authority or a judge, compared to the free initiative that characterises the action of a private individual. In any case the user has always the possibility to appeal in court, to ask the private plaintiff to account for her actions, if she considers that damage has occurred. However, in that case the user must be able to prove such damage.

[133] CLOUD Act, §2201 (7-11).

[134] As explained by the Boniolo, "[…] one aspect we have become directly familiar with since 11 September 2001: it seems that we are increasingly willing to surrender personal and collective freedoms in exchange for security." in GIOVANNI BONIOLO, *Il virus dell'idiozia. Sette scrittini su Covid-19, scienza, intellettuali e cittadini.* (Milano-Udine: Mimesis, 2021), p. 36 [Translation from the Italian original text]. See, also: UGO PAGALLO, ELEONORA BASSI, "The future of EU working parties' "the future of privacy" and the principle of privacy by design." *An Information Law for the 21st Century* (2011): p. 305; UGO PAGALLO, *La tutela della privacy negli Stati Uniti d'America e in Europa: modelli giuridici a confronto.* (Milano: Giuffrè Editore, 2008), pp. 157-196.

The CLOUD Act itself provides a set of safeguards to counter-balance indiscriminate access, for example "commitment to the rule of law and the protection of privacy and civil liberties"[135]. Their effectiveness will be assessed in the next future. In any case, the evaluation on a case-by-case basis, and thus also the interpretation of these broad and vague concepts, is left to the service providers, at least in the first instance.

In light of these considerations, it is fair to wonder about the reaction in the EU. However, for the time being, the situation is characterised by an institutional stalemate. So far, the only feedback on the participation of private US actors in higher education and research comes from the French supervisory authority, the "Commission Nationale Informatique & Libertés" (CNIL). The French authority gave its opinion in May 2021, on a request by the "Conférence des grandes écoles" (CGE) and the "Conférence des présidents d'université" (CPU), arguing that there is a risk of illegal access by US authorities to data held by US-providers operating in the EU. Pending further developments, the CNIL said it would assist in identifying "alternative solutions", as well as support higher education institutions, universities and research centres in their adaptation[136].

As mentioned in the previous chapter[137], data transfers to third countries and international organisations do take place, *despite* a complex and uncertain legal framework. Thus, this happens especially in the field of scientific research, where data transfer is often essential to progress in research projects and the advancement of knowledge, despite the uncertain legal framework. The institutional *impasse* does not limit the transfer, it only makes it less secure. Similarly, this

---

[135] CLOUD Act, §2202 (12-14). Moreover, in the case of the definition of 'executive agreement', a reciprocity clause is also provided for, between the parties to the agreement.
[136] CNIL, "La CNIL appelle à des évolutions dans l'utilisation des outils collaboratifs états-uniens pour l'enseignement supérieur et la recherche", May 2021, https://www.cnil.fr/fr/la-cnil-appelle-evolutions-dans-utilisation-outils-collaboratifs-etatsuniens-enseignement-superieur-recherche.
[137] See: Section 5.3.3.

applies to the regulation of electronic evidence, made more complex and obscure by the CLOUD Act.

In this context, the possibilities appear to be twofold: either exclude the US providers from participating in the European research projects, primarily the EOSC (or similarly, the Gaia X project); or, if this option does not seem feasible (due to the quality of the services of these private actors; or the economic model that has made them pervasive; or the intention of the institutions to express inclusion, etc.), then it is urgent and essential to develop a solution at the international level. As argued about the collection of the electronic evidence: "[…] from a pragmatic point of view, the search for a supra-regional e-evidence regulation should be based on a multilateral agreement. In principle, bilateral agreements in the field of regulating events occurring in cyberspace serve the function of *lex specialis* and thus complement the more general regulations arising either from national law or other international law mechanisms"[138]. Following the direction laid down by international law appears to be the only way to effectively protect the rights and freedoms of those involved indirectly (e.g., individuals) or directly in their work, i.e., researchers.

Bearing in mind this scenario, attention should now be drawn to the case study of this dissertation: the HPC platform of the University of Luxembourg, i.e., ULHPC.

6.4 ULHPC: The Case Study

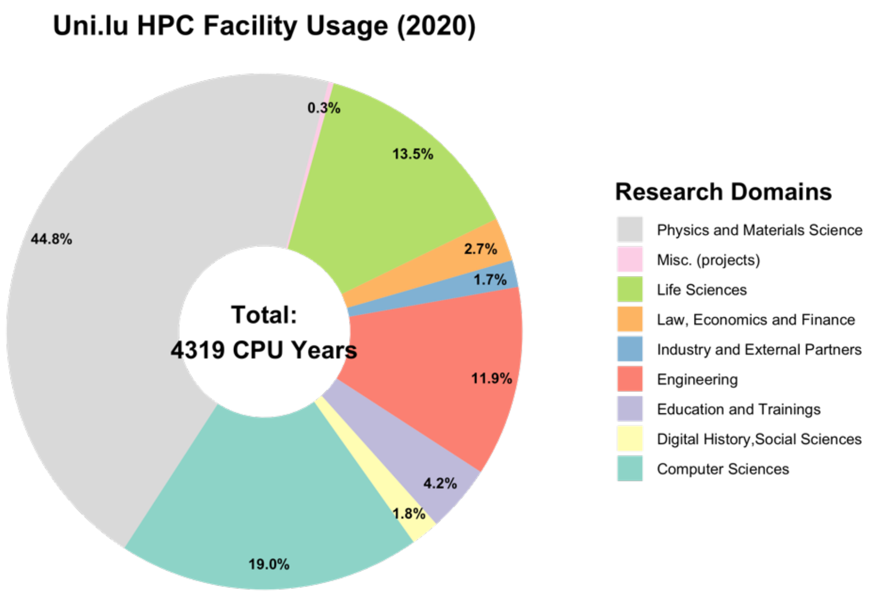
Across the European Union, many thriving research e-infrastructures and HPC facilities have emerged and are successfully operating in the field of public scientific research. These local experiences are essential to the development of the European scientific research. The European Open Science projects – first and foremost the EOSC – have a federated nature. The purpose is not to build a European research environment

[138] MARCIN ROJSZCZAK, *"CLOUD act agreements from an EU perspective.", op. cit.*, p. 14.

from scratch, but instead should be to enhance already existing and successful local resources.

The University of Luxembourg (UL), founded in 2003, holds an HPC platform, i.e., ULHPC, operating since 2007, led by led by Prof. Pascal Bouvry and Dr. Sebastien Varrette[139]. The ULHPC currently features a total of 690 nodes, namely 11228 computing cores, for a cumulative computing power estimated at 1262,869 TFlops.

Figure 6.3: "Computing usage of the ULHPC facility per research domain", in Paseri, Varrette, Bouvry (2021).



This private cloud, representing a Platform as a Service (PaaS) model, without the use of virtualisation, provides resources mainly for research purposes to researchers and companies acting as private partners in research projects, and in a small part, represented by 20%

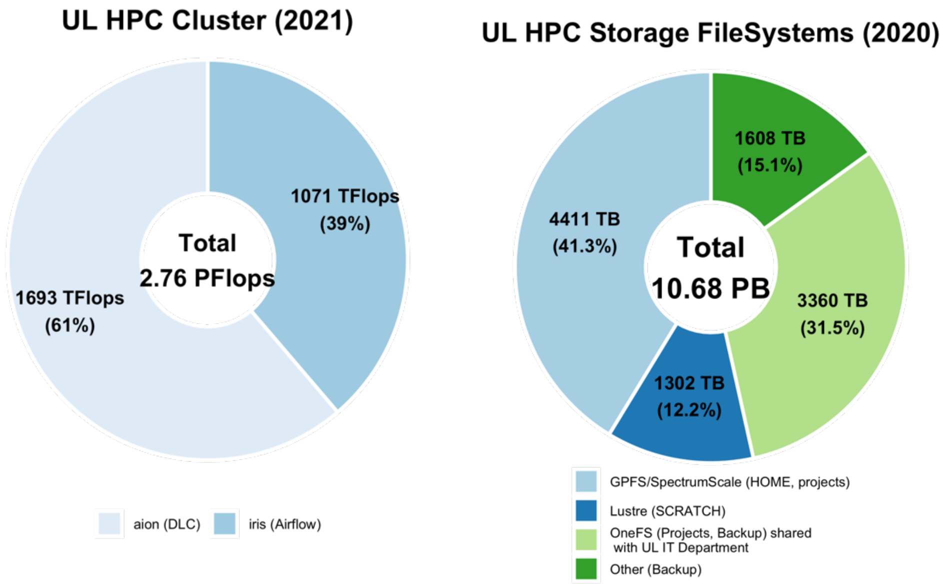
[139] https://hpc.uni.lu.

of the total resources available, for private actors, pursuing private interests.

In Figure 6.3[140] is represented the usage of the ULHPC platform, among the different areas of knowledge, during the year 2020. An interesting fact is that not only STEM disciplines benefit from the HPC platform. The humanities (e.g., social sciences, education sciences, history, etc.) stand to benefit from the usage of the ULHPC facility to conduct their research project.

The following Figure 6.4, then, provides a sketch, on the one hand of the computing capacity, measured in Flops, related to the year 2021; on the other hand, of the storage capacity, measured in bytes, related to the year 2020.

Figure 6.4: "Overview of the current computing and storage capacity of the ULHPC facility", in Paseri, Varrette, Bouvry (2021).



[140] Ludovica Paseri, Sébastien Varrette, Pascal Bouvry, *"Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes", op. cit.*, p. 127.

The case study of the ULHPC is set out below. Initially, in Section 6.4.1, the methodology adopted, and the research questions are illustrated. Then, the results are presented, distinguishing between the two strands of investigation: Section 6.4.2 presents the outcome of the technical analysis; Section 6.4.3 illustrates the finding of the legal analysis.

6.4.1 Methodology

The study conducted in Luxembourg, with Prof. Bouvry and Dr. Varrette and their team, represents the case study of this dissertation. The purpose was to realise an "[...] in-depth, multi-faceted explorations of complex issues in their real-life settings"[141], in light of the theoretical study developed in the previous chapters on the European approach to Open Science and the challenges related to the legal framework of the research data management. In other words, the aim was to bring the general issue to the concrete case[142]: understanding the strength of the theoretical analysis; identifying which aspects actually require intervention; proposing recommendations, both at operational level, i.e., addressed to those who manage the HPC platform for research purposes, but also at institutional and policy level, whether local, national or European.

Hence, the Research Question that guided the case study of the investigation was the following:

> *Which are the main legal challenges that an HPC platform used primarily for scientific research purposes has to face, in the Open Science scenario?*

---

[141] SARAH CROWE, *et al.*, "The case study approach." *BMC medical research methodology* 11.1 (2011): 1-9, doi: 10.1186/1471-2288-11-100.

[142] See, for example, ROBERT K. YIN, *Case study research. Design and methods*, 4 eds. (London: Sage Publication, 2009), p. 18: "[I]n other words, you would use the case study method because you wanted to understand a real-life phenomenon in depth, but such understanding encompassed important contextual conditions – because they were highly pertinent to your phenomenon of study".

In order to answer this RQ, the study was structured in four phases, described below: (i) technology study; (ii) identification of the disciplines involved; (iii) legal analysis; (iv) outcome presentation.

(i) Technology study. The first phase of the case study investigation consisted of an in-depth study of the technology involved. This first phase was mainly carried out in two ways: on the one hand, through the study of HPC and its technical functioning; on the other hand, through a series of interviews with the practitioners of the ULHPC team at the University of Luxembourg.

Two important aspects emerged at this stage. First, this prodromic analysis turned out to be fundamental in an interdisciplinary research project aimed at investigating the impact of new technologies on the scientific research environment[143]. Second, a typical issue of interdisciplinarity emerged[144], represented by the lexical differences between the two different areas of knowledge involved, i.e., law and computer science.

As an illustration, consider the basic concept of "personal data". The European lawmaker defines the notion of personal data in the Article 4(1) of the GDPR. Let us consider, then, the case of a researcher in the

[143] On the essential need to learn technical aspects in interdisciplinary studies, see: GIOVANNI BONIOLO, RAFFAELLA CAMPANER, "Life Sciences for Philosophers and Philosophy for Life Scientists: What Should We Teach?" *Biological Theory* 15 (2020): 1-11, doi: 0.1007/s13752-019-00333-7. The Authors refer to the case of studies conducted by philosophers of law, in the field of life sciences, pointing out that such interdisciplinary projects "[…] need to be implemented in an environment in which students with a background in philosophy can work side by side with scientists in the labs, thus learning technical aspects of experimental practice in great deal.", p. 8.

[144] See, among others, JULIE THOMPSON KLEIN, "A Conceptual Vocabulary of Interdisciplinary Science." *Practising interdisciplinarity* (Toronto: University of Toronto Press, 2018): 3-24, in which the Author explains that "[T]here is no universal interdisciplinary language. Even powerful cross-fertilizing languages, such as mathematics and general systems, have limits. Interdisciplinary work requires […] 'horizontal communication' within an 'interdisciplinary/integrated culture'. A working language emerges through the negotiation of meanings", p. 18. But, see also ANN BRUCE, *et al.*, "Interdisciplinary integration in Europe: the case of the Fifth Framework programme." *Futures* 36.4 (2004): 457-470, doi: 10.1016/j.futures.2003.10.003, in which the Authors show that "[M]any in our survey stressed problems of language and communication caused by a range of factors. Different disciplines use different languages and the same word may mean different things in different disciplines, resulting in a great deal of frustration until this is clarified. Communication problems were found in all types of interdisciplinary collaboration", p. 467.

bio-medical field who everyday processes personal data, in conducting her research project, carried out through the use of HPC facilities. The researcher actually handles, processes, and elaborates personal data. Although, she may not agree on the boundaries of the concept of personal data with the law expert, who follows the wording of the GDPR. On the contrary, it is not straightforward for the legal expert to understand the technical concept of data and data flow in a technologically complex environment such as the HPC ecosystem.

This tension between different domain of knowledge has always traditionally existed. However, it has become more evident in relation to the use of new technologies, such as HPC facilities, which necessarily involve different field of knowledge. For this reason, every interdisciplinary project requires a prodromic activity, leading to the establishment of a common terminology.

(ii) Identification of the disciplines involved. In light of the technical analysis, the second phase concerned the legal disciplines involved in the processing of research data through the ULHPC facility. The most complex legal issue related to the usage of the ULHPC platform has been identified in the protection of personal data and the compliance with the GDPR. Secondly, concerning the emergence of the Open Science paradigm, the issue of greatest interest to this case study has been identified in the process of making the research data FAIR, i.e., Findable, Accessible, Interoperable and Reusable[145].

(iii) Legal analysis. The third stage in the development of the case study concerned the outlining of the legal analysis: the focus was on the protection of personal data and the compliance of ULHPC with the GDPR. The legal analysis was carried out by mapping all the major issues and weaknesses of the HPC platform's GDPR compliance process. In addition, this analysis included a comparison with similar HPC platforms at a local level: in particular, the HPC platform of the

[145] For an in-depth analysis of the concept of FAIRness of research data, see: Chapter 4, specifically, Section 4.1.2.

University of Turin (Italy), called “HPC4AI”[146], and the infrastructure of the Research Center “Area Science Park”, in Trieste, which holds an integrated environment of cloud computing and HPC capabilities, called “Orfeo Ecosystem”[147].

(iv) Outcome presentation. The results deriving from the legal analysis were then presented in two events: in the “Annual Privacy Forum” (APF) Conference, organised by the European Network and Information Security Agency (ENISA), in June 2021[148], with a related publication, entitled “Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes”[149]; and in a Seminar of the series “IncontrinRete”, organised by the Area Science Park, in Trieste, in December 2020.

In doing so, feedback and insights has been gathered from a legal perspective, in the APF Conference, and from a technical perspective in the Area Science Park Seminar.

Thus, in next sections the main findings of the ULHPC case study are exposed.

6.4.2 Technical Analysis Outcome

The fundamental HPC computing hardware is represented by the Central Processing Unit (CPU) and the Graphics Processing Unit (GPU). The former represents the traditional cores that, working in synergy, identify the processor, traditionally used for the various types of operations, given the maximum flexibility of the software; on the other hand, GPUs are the so-called accelerators, able to perform more complex operations (rotations, transitions, etc.). GPUs are very useful for machine learning (ML) or deep learning workloads, due to their ability to process vectors.

---

146 https://hpc4ai.unito.it.

147 https://www.areasciencepark.it.

148 https://www.enisa.europa.eu/events/apf-2021/annual-privacy-forum-2021.

149 LUDOVICA PASERI, SÉBASTIEN VARRETTE, PASCAL BOUVRY, “Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes.” *Annual Privacy Forum* (Cham: Springer, 2021), p. 125, doi: 10.1007/978-3-030-76663-4_7.

The essential HPC components are: (i) memory; (ii) interconnect; (iii) operating system; (iv) stack software; (v) file system; (vi) data center[150].

(i) Memory. Three types of memory can be identified, i.e., cache memory, Dynamic Random Access (DRAM) memory and disk memory. These three different types of memory represent a hierarchy: the closer you are to the CPU, the faster you can write to it. For this reason, cache memory is the fastest, as well as the narrowest (the size varies from 64 KB to 8 MG, for a speed ranging from 1 or 2 cycles to 20 cycles); while DRAM memory represents an intermediate way with a size of 1 GB and a speed of hundreds of cycles; and, finally, disk memory represents the slowest memory, on which it is more difficult to write, but with more space (the size is 1 TB, while the speed is tens of thousands cycles).

(ii) Interconnect. In order to manage a multitude of servers as if it were only one, it becomes fundamental to establish an effective method of transferring data between the various servers. In this regard, three technologies are generally used, i.e., Ethernet[151], Infiniband or Omni-Path Interconnect[152]. Infiniband is the most used interconnection in the HPC platforms and is also adopted by the ULHPC facility. It is extremely effective: the bandwidth does not vary much compared to the Ethernet, but the time necessary to send a message, i.e., the latency, is below 1 micro-second.

(iii) Operating System. The operating system represents the basic software, which makes all operations possible, managing the hardware

---

[150] SÉBASTIEN VARRETTE, PASCAL BOUVRY, VALENTIN PLUGARU, *et al.*, "Overview and Challenges of the UL HPC Facility at the EuroHPC Horizon," *7th High Performance Computing School,* Luxembourg (2018).

[151] Ethernet is the traditional protocol used by the Internet network, whose maximum quantity of data communicated per unit, namely the bandwidth, varies between 1 GB/s and 125 MB/s, while the network latency, i.e., the time necessary to transfer data between two locations, is decidedly minimal.

[152] Omni-Path Interconnect, supplied by Intel, represents a competitor of the Infiniband, since the two are very similar from a technical point of view.

and other software elements of the machine. The ULHPC platform is Linux-based: this option allows stability and development flexibility.

(iv) Software Stack. The software stack is essentially identified by the following elements: the remote connection to the platform, represented by the SSH (Secure Socket Shell), the protocol that allows a secure connection in order to access the remote machine; the identity management method, represented by the SSO (Single Sign-On), the property by which access is granted to a single user, through a single authentication, allowing the specific machines to perform, obtaining scalability; and the resource management, a work scheduling system, which stands between the single user's request and the platform status, to allocate computational resources; specifically, the ULHPC platform adopts SLURM Workload Manager.

(v) File System. One of the most important components is the file system, since it allows logical storage, organisation, and access to data. The typical file system for HPC is a parallel/distributed file system, due to its ability to increase the system, both in terms of capacity and performance, orchestrating the work of different servers operating in synergy. Among the most used are SpectrumScale and Lustre.

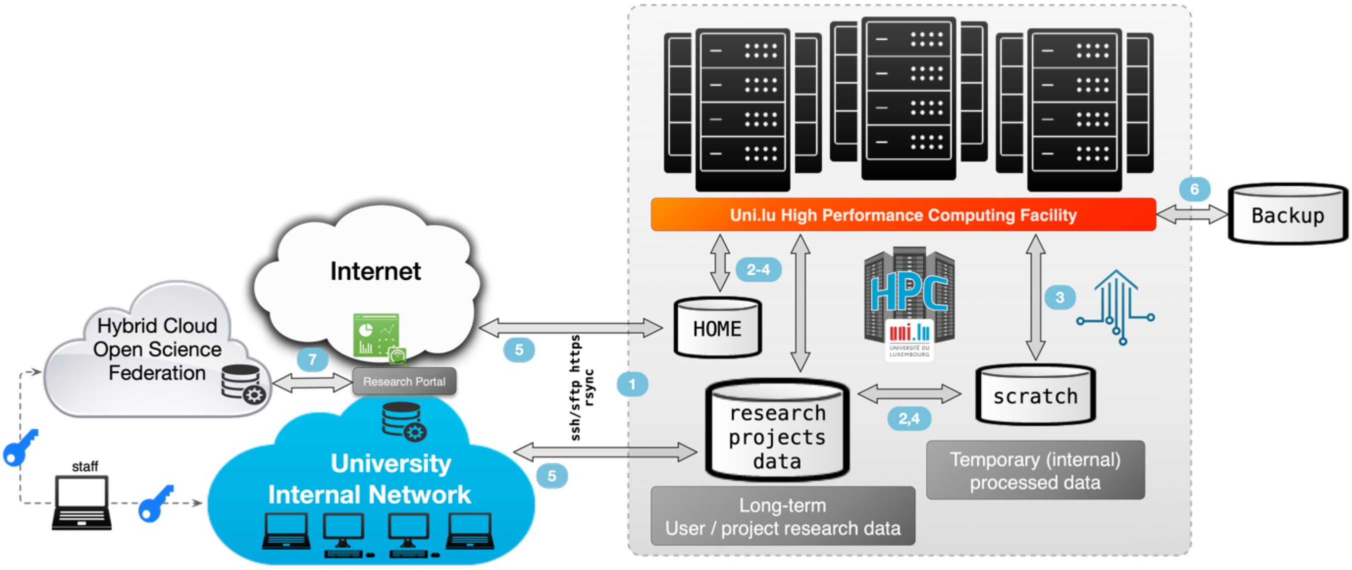
(vi) Data Center. The various elements analysed so far need, then, a physical house, represented by the data center, i.e., the structure specifically designed to host the various components necessary to make the platform operational, organised in racks, each with a height of 42 rack unit (RU).

The Figure 6.5 represents the data flow in the ULHPC environment[153].

---

[153] LUDOVICA PASERI, SÉBASTIEN VARRETTE, PASCAL BOUVRY, *"Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes."*, *op. cit.*, p. 134.

Figure 6.5: "Data processing interactions for HPC workflows on the ULHPC facility", in Paseri, Varrette, Bouvry (2021).



The numbered arrows in Figure 6.5 indicate the various steps in the data flow processed in the ULHPC. The data flow is summarised below.

1. Input data, here, pass from external sources to the long-term storage area of HPC. This phase marks the entry of data into the ULHPC environment.
2. This is the pre-processing phase of the research data: they are prepared for subsequent processing by data analysis techniques. This operation may require internal and temporary transfer to the scratch area.
3. In this phase, the data are actually processed: the computing facilities process the various jobs submitted; in doing so, intermediate and final data components are generated.
4. This is the post-processing phase: the aim is to extract the results from the research data just processed. In this phase metadata are generated. Metadata provide a description of the

data itself and facilitate its retrieval, fundamental for the FAIRness.

5. This phase indicates the output of processed data. They exit the HPC environment and are directed towards external resources, such as researchers' personal computers or laboratory servers.
6. The output data, processed, are archived, since a long-term backup of the storage area is guaranteed.
7. Finally, this is the phase in which research data, processed in the ULHPC environment, move externally. The cases here are manifold: they may be stored in a domain-based database; or they may be shared with external parties participating in the same research project; or where possible, they are openly released in specific open datasets, to allow them to be shared and reused by third parties.

Bearing in mind the results of this data lifecycle mapping in the ULHPC environment, it is time to expose the results of the legal analysis.

6.4.3 Legal Analysis Outcome

The legal analysis covered two strands of investigation: the protection of personal data and the application of the European Open Science policies.

Regarding the data protection issues, the Luxembourg Centre for Systems Biomedicine (LCSB) is making great efforts in the process of compliance with the GDPR, outlining a model of Data Protection Impact Assessment (DPIA), also related to research project conducted using the ULHPC facilities[154]. Performing the DPIA is a task that acquires even more relevance in light of the Luxembourg Data

[154] PINAR ALPER, REGINA BECKER, *et al.*, "Provenance-enabled stewardship of human data in the GDPR Era." *Provenance and Annotation of Data and Processes* (Cham: Springer, 2018): 266-269 doi: 10.1007/978-3-319-98379-0_33; MATTHIAS GANZINGER, ENRICO GLAAB, *et al.*, "Biomedical and clinical research data management" *SystemsMedicine: Integrative, Qualitative and Computational Approaches*, 3 (2021): 532-543, doi: 10.1016/B978-0-12-801238-3.11621-6.

Protection Law[155]. As explained above[156], the Article 65 provides a list of twelve measures the controller shall implement in order to process personal data for research purposes, which includes the definition of the DPIA. The last paragraph of the Article 65 states: "The controller shall document and justify for *each* project for scientific or historical research or statistical purposes the exclusion, if any, of one or more of the measures listed in this Article"[157]. This disposition requires a considerable amount of effort to be invested in each single research project, resulting in an analysis that is almost comparable to a further data protection impact assessment.

In addition to the DPIA, the main GDPR compliance issues arise in relation to (i) the principle of accountability; (ii) the information duties imposed by the Articles 13 and 14 of the GDPR[158]; (iii) the right of access to data by the data subjects; (iv) the principle of data minimisation; and (v) certain aspects related to the movement of data within the ULHPC environment[159].

For the first three issues, i.e., accountability, communication of information and fulfilment of the right to access, a two-pronged approach was suggested. On the one hand, the standardisation of internal procedures should be enhanced. The ULHPC team should establish internal practices to regulate the various types of interaction that may take place with data subjects involved in research projects. Essential from this point of view is the coordination between the

---

[155] The Article 65, Loi du 1er août 2018, *portant organisation de la Commission nationale pour la protection des données et du régime général sur la protection des données* (Luxembourg Data Protection Law) (2018), ELI: https://data.legilux.public.lu/eli/etat/leg/loi/2018/08/01/a686/jo. [Word "each" not emphasised in the original text].

[156] See: Section 5.5.1.

[157] The Article 65, Luxembourg Data Protection Law, 2018.

[158] Regarding the relevance of information duties, within the GDPR see: MASSIMO DURANTE, "Commento all'art. 13 GDPR. Informazioni da fornire qualora i dati personali siano raccolti presso l'interessato." in ENRICO GABRIELLI (ed.), *Commentario del Codice Civile Utet. Modulo delle Persone, Vol. II.* (Milano: Utet Giurdica, 2019), pp. 218-234.

[159] For a full description of the issues, see: LUDOVICA PASERI, SÉBASTIEN VARRETTE, PASCAL BOUVRY, *"Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes."*, *op. cit.*, p. 134-136.

ULHPC team and the different research teams of the University that utilise the infrastructure.

On the other hand, the reinforcement of training activities was suggested. This aspect is closely linked to the respect of the principle of accountability: all researchers, from any research field, who are involved in the processing of personal data should have a basic knowledge of the European data protection regulation[160]. Thus, it is suggested to interpret the principle of accountability as a "meta-principle"[161]. This requirement is often not easy to meet, as researchers are frequently burdened with bureaucratic and administrative tasks. Making data protection training mandatory, according to the GDPR, risks being seen only as another bureaucratic formality to be fulfilled. In this way, the *ratio* behind the GDPR obligations is totally undermined.

For this reason, a suggestion might be to develop a policy of incentives related to good data management. In Chapter 4, dealing with research data, and in Chapter 5 dealing with the legal challenges of data protection in the Open Science context, a proposal was developed, which may be applied here. Starting from the assumption

---

[160] This obligation is related to the task of the data protection officer, pursuant to the Article 39(1)*b* "[T]he data protection officer shall have at least the following tasks […] to monitor compliance with this Regulation […] including the assignment of responsibilities, awareness-raising and training of staff involved in processing operations". In addition, the Article 32(4) of the GDPR about the security of processing, states that "[T]he controller and processor shall take steps to ensure that any natural person acting under the authority of the controller or the processor who has access to personal data does not process them except on instructions from the controller". A general mandatory duty of training is also related to the principle of accountability, pursuant the Article 5(2) "[T]he controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability')".

[161] "The GDPR, in fact, uses the term "principle" to regulate different moments of the processing of personal data: there are principles that determine whether or not it is possible to implement a particular processing; principles that, on the other hand, are intended to determine how certain processing can be conducted; and finally, there are the meta-principles, helping to interpret others.The principle of accountability can be considered as a meta-principle. […] This interpretation of the accountability principle is extremely appropriate in a scenario like ours, characterised by a multitude of different subjects and by the possible ambiguity in the identification of the different roles within the domain of scientific research" in LUDOVICA PASERI, SÉBASTIEN VARRETTE, PASCAL BOUVRY, *"Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes.", op. cit.*, p. 135. This aspect has been discussed in Section 5.3.2.2.

that research data are fundamental, it is necessary to understand that, therefore, their proper management is also fundamental. A good management of research data should be envisioned as consisting of many aspects[162]. Data protection should be considered an essential part of the many elements related to good research data management[163].

However, good management of research data in general – and of personal data processed for research purposes in particular – cannot be limited to a bureaucratic burden. Alternative solutions should be conceived. One suggestion would be to intervene at the university policy level, designing a system of incentives linked to good data management. The advantage would be twofold: on the one hand, the effort of researchers in data management, in terms of time and knowledge acquired, would be rewarded; on the other hand, the university would obtain a very valuable research data asset.

If such an intervention is not feasible in many realities, Luxembourg is certainly not one of them. Investment in research is steadily increasing, and the University, although young, attracts many

[162] The topic has already been discussed in Section 4.3. One essential aspect is the process of making research data FAIR. The state of the art related to FAIRness of research data processed in ULHPC has been investigated in LUDOVICA PASERI, SÉBASTIEN VARRETTE, PASCAL BOUVRY, "*Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes.", op. cit*., p. 136-137. Regarding the FAIRness of research data in ULHPC environment, a problem of reproducibility emerged. In order to guarantee the reproducibility, it is essential to be able to provide multiple versions of software involved. The same concern has been discussed in EOSC Executive Board, SRIA, *op. cit*., p. 26: "In order to be usable by scientists, research software archives need to comply with specific requirements. They have to keep multiple versions in order for scientists to be able to use the version that will ensure reproducibility. Research software uses generic components such as operating systems, compilers, scientific libraries, etc. Therefore, in order to allow reproducibility, these generic components also need to be kept. As a consequence, archiving of research software has to be part of general-purpose software archives".

[163] The fact that data protection is already part of some templates in the Data Management Plan (DMP) documents is a very good start. See: PAWEŁ KAMOCKI, VALÉRIE MAPELLI, KHALID CHOUKRI, "Data management plan (DMP) for language data under the new general data protection regulation (GDPR)." *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (2018): 135-139; SAGAR BHIMRAO GAJBE, AMIT TIWARI GOPALJI, RANJEET KUMAR SINGH, "Evaluation and analysis of Data Management Plan tools: A parametric approach." *Information Processing & Management*, 58.3 (2021): 1-17, doi: 10.106/j.ipm.2020.102480.

researchers every year[164]. In addition, such an initiative could be a local implementation of the Regulation (EU) 2021/695, establishing the Horizon Europe programme. In fact, the Article 39 states that "[T]he work programme may provide for additional incentives or obligations for the purpose of adhering to open science practices".

Regarding the last two issues, i.e., the principle of minimisation and security measures related to data movement, the recommended approach is different. Data minimisation is a fundamental principle of the GDPR, but as seen in Section 5.3.2, in scientific research it is not always guaranteed, just as the related anonymisation techniques are not always feasible[165]. For this reason, the three-step reasoning proposed in Chapter 5 is recommended to address the issue of compliance with the minimisation principle[166]. Anonymisation techniques related to phase 2 of the data flow described above could be implemented[167].

Yet, where the anonymisation option is not feasible, the concern arises in phase 7 of the data flow described above[168]. In fact, the ULHPC is based on a local infrastructure, physically managed by the University, which itself owns and operates the data center. In this case, there is no outsourcing of services, so there are no external transfers of data *per se,* within the ULHPC environment.

Such transfers would take place in phase 7: either in the case of ULHPC participation in infrastructure federation projects (e.g., EOSC)

---

[164] Consider that Luxembourg ranked number 1 in Europe for attractive research system in 2018, 2019 and 2020, see: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_1150. In addition, according to a Eurostat Study, on data extracted in April 2021: "The share of people employed in a science and technology occupation peaked at 54.8 % in Luxembourg", https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Research_and_development_statistics_at_regional_level#Research_and_development_expenditure, and see, also: https://wwwen.uni.lu/university/about_the_university/rankings_accreditations, about the 2021 rankings of the University of Luxembourg.

[165] See: Section 5.3.2.

[166] See: Section 5.3.2.

[167] See: Figure 6.5 and Section 6.3.2

[168] *Ibid.*

or in the context of collaborative research projects. In the first case, i.e., the participation in infrastructure federation projects, it is essential to guarantee a secure authentication procedure for entry and authorisation for access to internal resources. In the second case, i.e., collaborative research, once again, training plays a fundamental role: in this scenario the researcher autonomously oversees the management of research data. For this reason, it is crucial to have previously invested in the accountability. In this way, the two typical and extreme behaviours can be avoided, i.e., the total and indiscriminate openness of data with risks to the freedoms and rights of the data subjects involved, or the opposite total closure of data, due to excessive protection. Both of these polarisations may be detrimental to scientific research.

The situation is different as regards the security issues within the ULHPC environment, linked to the tracking of data movement. The legal context is made more complex, once again, by the national discipline. The aforementioned Article 65 of the Luxembourg Data Protection Law imposes a list of twelve measures to safeguard data processing for research purposes. Among them, it provides for: "[…] the use of a log file enabling the reason, date and time that data is consulted and the identity of the person collecting, modifying or deleting personal data to be retraced". As previously stated, "[T]echnically speaking, parallel and distributed filesystems used in HPC environments such as the ULHPC are not yet fully able to account and log internal data movements"[169]. There are circumstances in which "[…] the effectiveness of the right is […] strongly challenged by the technical rule"[170]. Specifically, the difficulties emerge in phases 2 and

[169] Ludovica Paseri, Sébastien Varrette, Pascal Bouvry, *"Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes.", op. cit.*, p. 136.

[170] Paolo Guarda, *Il regime giuridico dei dati della ricerca scientifica.* (Trento: Collana della facoltà di Giurisprudenza dell'Università di Trento, 2021), p. 282, where the Author points out that: "In this context the real problem becomes not so much, or not only, the control of the data per se considered, as that of the infrastructure capable of managing it". [Translation from the Italian original text].

4 of the data flow described above. Instead, in phases 1, 5 and 7, organisational and security measures could be implemented to overcome the problem connected to the tracking of data, adopting measures which guarantee by design and by default greater control over the movement of data[171].

The results of the technical-legal analysis carried out on the ULHPC structure, summarised above, were then discussed with similar local experiences of HPC platforms for scientific research[172]. A relevant final consideration that emerged is that such local experiences face very similar problems, both in relation to the GDPR compliance and to the adoption of practices related to Open Science. What is lacking, therefore, is coordination between these multiple local experiences, which are operating with excellent results in their respective environments. An attempt to fill this gap of coordination is taking place in Italy, with the already mentioned "Italian initiative of the National Interuniversity Consortium for Informatics (CINI), about the establishment of a laboratory which intends to enhance HPC developments"[173]. Initiatives of this nature should aim to provide for interdisciplinary working groups, capable of dealing also with the practical legal issues arising in different scenarios.

6.5 Conclusive Remarks

Chapter 6 investigated the digital infrastructures in the Open Science scenario, starting from the following RQ6:

[171] The reference goes to the notion of "Data protection by design and by default", as set out in the Article 25 of the GDPR. The issue must be further investigated from the point of view of technical feasibility. This aspect would require further in-depth analysis and could represent a future development of the research presented here.

[172] The HPC platform of the University of Turin (Italy) , called "HPC4AI", see: https://hpc4ai.unito.it; and the infrastructure of the Research Center "Area Science Park", in Trieste, which holds an integrated environment of cloud computing and HPC capabilities, called "Ecosystem Orfeo", see: https://www.areasciencepark.it.

[173] MARCO ALDINUCCI, *et al.*, *"The Italian research on HPC key technologies across EuroHPC.", op. cit.*, p. 1.

What role do e-infrastructures for research play in the Open Science paradigm?

In answering RQ6, the recent considerable evolution of research e-infrastructures in the last decades was examined. Furthermore, the dissertation chose to focus, in particular, on one computing infrastructure, namely the High Performance Computing (HPC) facility. HPC was discussed from the European perspective, i.e., the definition and development of European policies; and from a local perspective, by presenting the case study of ULHPC.

In light of this analysis, drawing on the argument recently expressed by Pagallo[174], it seems fair to admit the existence of a real duty to take the maximum advantage of existing technologies. This duty on the one hand is grounded on the framework of human and fundamental rights at the basis of the Open Science paradigm, first of all the human right to science[175]; on the other hand, it is based on the awareness that currently human agents (i.e., researchers) and artificial agents (i.e., machines) cooperate and coexist in science, as *inforg* of science.

The European HPC strategy seems intended to take the advanteges from technologies, specifically for the benefit of scientific research. The investigation showed that at the basis of this strategy there is a notable interplay between the public and private sector. For this reason, it was argued that the problem should not be framed in the alternative between participation or exclusion in science of private actors. Rather, the trade-off is between transparent or *opaque* interplay between public and private actors in scientific research.

If this strong interplay between the public and private actors in HPC will be managed properly, in line with the Open Science tenants, this

---

[174] UGO PAGALLO, *Il dovere alla salute. Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario*. (Milano-Udine, Mimesis: 2022).
[175] On this aspect, see: Section 3.3, and also, GIULIA PERRONE, MARCO PERDUCA (eds.), *Così San Tuttə. Diritto alla scienza: istruzioni per l'uso*. (Roma: Fandango libri, 2021): pp. 111-119.

sector could become a benchmark of public-private collaboration in favour of scientific research.

Chapter 7

Conclusions

Our word is marked by inequalities in the access to opportunities and distribution of resources[1]. Historically, most of these have been economic resources consisting of material goods. Today, the inequalities connected to the distribution of resources increasingly concern the intangible goods, such as knowledge. Different limits and opportunities in accessing knowledge are therefore likely to deepen existing inequalities or generate new ones. In this perspective, law also plays a crucial role in the distribution of resources, as it helps to draw or redefine the limits and opportunities of access to knowledge[2].

Today we are going through a digital revolution: new technologies, ICTs, AI, ML, etc., have generated an enormous and various impact on every aspect of our lives. Law is no longer the only regulatory system. Different normative systems, such as technology, economics, or social norms, compete with the law, and institutions must be fully aware of this phenomenon[3]. Sometimes, this competition results in a delegation of public functions or quasi-judicial powers to private actors, such as in

---

[1] "At the end of 2019, our figures indicate that millionaires around the world – which number exactly 1% of the adult population – accounted for 43.4% of global net worth." in Credit Suisse, "Global Wealth Report", 2020, p. 29, https://www.credit-suisse.com/about-us/en/reports-research/global-wealth-report.html, but see also OECD, *Inequalities in household wealth and financial insecurity of households*, (2021), https://www.oecd.org/social/income-distribution-database.htm.

[2] KATHARINA PISTOR, *The code of capital. How the law creates wealth and inequality*. (Princeton: Princeton University Press, 2019), pp. 109-110.

[3] See: Section 2.4.3. Recently, on this aspect see: UGO PAGALLO, "Sovereigns, Viruses, and the Law." *Law in Context. A Socio-legal Journal* 37.1 (2020): pp. 18-19. Here, the Author focuses on the governance of multi-agent systems of information societies during the pandemic, but the reasoning can be adapted to the domain of the scientific research. In addition, many sectors over the years developed multi-level governance. Among the many, consider the emblematic example of telecommunications, see: MARCO OROFINO, *Profili costituzionali delle comunicazioni elettroniche nell'ordinamento multilivello*. (Milano: Giuffrè Editore, 2008), pp. 79-106.

the case of the CLOUD Act, exposed in Chapter 6[4]. Therefore, it seems fair to wonder about the suitability and consequences of this type of interventions.

In the field of science, the European Union has an incredible opportunity: building the conditions to promote the new paradigm of science, i.e., the Open Science paradigm, bridging different actors, systems and phases of the scientific research process[5]. The EU can be the most suitable space to advance science, taking full advantage of new technologies, promoting the fifth European freedom, i.e., the free movement of knowledge, going beyond the strictly academic environment[6].

Open Science encompasses and promotes the openness of knowledge in all its dimensions: for instance, circulation of research data, open educational practices, a widespread and greater societal engagement, open access to scientific publications, etc. In other words, Open Science policies may be considered the means by which to enforce, today, the human right to science as enshrined in the Article 27 of the Universal

---

[4] See: Section 6.3.2.

[5] On this aspect see: ROBERTO CASO, *La società della mercificazione e della sorveglianza: dalla persona ai dati: casi e problemi di diritto civile.* (Milano: Ledizioni, 2021), p. 71. Similarly, on this aspect, see also: MASSIMO DURANTE, "Re-designing the Role of Law in the Information Society." *European Journal of Legal Studies* 2.3 (2010): 1-18.

[6] "The possibility of building a true knowledge society, in which all citizens have the opportunity to be part of it not only formally but substantially, is not so trivial. It is not sufficient to fight for *isegoria*, but one should fight against *parrhesia*, against speaking out of turn. This implies that, on the one hand, there must be sharing and transmission of knowledge, but, on the other hand, there must also be the intention to share the knowledge of others and to be ready to receive knowledge. I find quite limiting the idea that knowledge should be addressed only by considering the school-university side, when it is obvious that this transmission is (should be) practised by those who have it (the teachers) towards those who do not have it (the students). I believe it is limiting to ignore life outside schools and universities, where there is a clear and serious division of knowledge between those who know and those who do not know, between those who want to know and those who do not want to know, between those who exploit the ignorance of others and those who are exploited because of their ignorance, between those who live better because they know what to do and where to go (especially in the case of illness) and those who do not know and live worse and, if they are unlucky, even die of it." in: GIOVANNI BONIOLO, *Conoscere per vivere: Istruzioni per sopravvivere all'ignoranza.* (Milano-Udine: Mimesis, 2018) p. 156-157. [Translation from the Italian original text].

Declaration of Human Rights and the Article 15 of the International Covenant on Economic, Social and Cultural Rights.

Section 7.1 sums up the main findings of this dissertation. Section 7.2 exposes some final remarks and policy considerations on the European approach to the governance of Open Science. In this perspective, Section 7.3 finally identifies some open issues and possible future research developments.

7.1 Main Findings

This dissertation intends to participate in the debate on the European scientific research governance, with a specific focus on the legal challenges of the new emerging paradigm of science, i.e., the Open Science. The emergence of the Open Science paradigm has been investigated for about a decade[7]. However, Open Science in recent years is receiving – and increasingly in the years to come will receive – considerable attention, essentially for two reasons.

First, we are experiencing an ever-increasing pervasive presence of technologies in every sphere of our lives and in every aspect of our societies, including the scientific research domain[8]. The COVID-19 pandemic has amplified this dynamic, shifting online many of the activities originally carried out in presence[9], some of which represent the exercise of fundamental rights, such as the right to work or the right to education. The "*onlife*" dimension[10] regards also the field of science[11].

---

[7] An emblematic analysis is SÖNKE BARTLING, SASCHA FRIESIKE, *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing.* (Cham: Springer Nature, 2014). For further investigation, see: Section 3.1.

[8] This factor is investigated in Section 2.1.1 as the methodological premise of this dissertation.

[9] MARIAROSARIA TADDEO, "Virus e tecnologia: perché sono importanti scelte etiche (Virus and Technology: Why Ethical Decisions Matter).", *Aspen Institute Italia, Aspenia* 89 (2020): p. 2, https://ssrn.com/abstract=3632977.

[10] LUCIANO FLORIDI, *The onlife manifesto: Being human in a hyperconnected era.* (Cham: Springer Nature, 2015).

[11] See, as an example: DAVID LESLIE, "Tackling COVID-19 through responsible AI innovation: Five steps

Second, European institutions have clearly expressed their intention to embrace the Open Science paradigm in the development of the European scientific research policies[12]. Open Science, in its meaning of free circulation of knowledge, is implicitly embedded in the very foundations of the European Union, as an expression of the fifth European freedom of movement. Recently, in addition, the intention to embrace Open Science has been made explicit: for example, with the project of the European Open Science Cloud (EOSC), a federated and trusted environment for access and sharing of data and services for the benefit of the European researchers[13]; or, with the establishment of the new research grant programme, i.e., the Horizon Europe programme, laid down in the EU Regulation 2021/695[14], which links research funding to the adoption of the Open Science tenets.

For these reasons, this dissertation investigates the emergence of this new paradigm of science, focusing on the approach adopted by the EU institutions. The following sections expose the main findings, which are briefly listed as follow: (i) identifying and addressing preliminary legal gaps to the emergence of the Open Science paradigm, suggesting a conceptual framework of the EU Open Science policies' evolution; (ii) proposing a governance approach able to encompass all the challenges of Open Science; (iii) defining the Open Science in its entirety; (iv) clarifying the concepts at the basis of the openness of the research data, such as FAIRness and legal interoperability; (v) exploring the hurdles of the "open by default" clause as an hard law provision; (vi) offering an ethical and infraethical framework for open research data; (vii) showing the feasibility of a win-win scenario for data protection and

in the right direction." *Harvard Data Science Review* (2020): 1-58, doi: 10.1162/99608f92.4bb9d7a7.

[12] This factor is investigated in Section 2.1.2 as the institutional premise of this dissertation.

[13] PAUL AYRIS, *et al.,* "Realising the European open science cloud", *Publications Office of the European Union* (2016): 1-24, doi: 10.2777/940154.

[14] Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021, *establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013*, OJ L 170, 12.5.2021, p. 1–68, ELI: http://data.europa.eu/eli/reg/2021/695/oj.

Open Science; (viii) presenting the Open Science actors as *inforg* of science.

7.1.1 A Conceptual Framework for the EU Open Science

The implementation of the Open Science paradigm is very complex and brings with it several issues. This complexity derives, primarily, from the fact that tackling the implementation of this new paradigm requires an interdisciplinary approach. In fact, Open Science involves many domains: notably, science and the way it operates; but also, technology, which is the real trigger of this paradigm shift; and law, establishing the rules and co-participating in the governance[15].

From a legal point of view, in this scenario, two preliminary gaps are identified. On the one hand, there are many initiatives, projects, and policies, developed at the European and national level, in the field of Open Science[16], and there is a lack of analysis providing a comprehensive overview of the European policies in this domain. This is a problem since it generates legal uncertainty, and it limits the development of new policies or the implementation of existing ones.

On the other hand, there is a lack of attention to the legal issues arising from the adoption of the Open Science strategies. These policies do not emerge from scratch (i.e., in a normative vacuum), and it is necessary to assess their impact on the existing legal framework of the European legislative provisions.

---

[15] Governance is "[…] understood as the set of formal and informal rules through which decisions are made and political authority is exercised", in UGO PAGALLO, "Good Onlife Governance: On Law, Spontaneous Orders, and Design." in LUCIANO FLORIDI (ed.), *The Onlife Manifesto: Being Human in a Hyperconnected Era.* (Dordrecht: Springer, 2017): pp. 161-177, doi: 10.1007/978-3-319-04093-6_18.

[16] Similarly, this is one of the problems that the EOSC is facing: "There are many monitoring mechanisms in Europe, offered by, for example, OpenAIRE, DCC, SPARC Europe, GÉANT, etc. Nevertheless, none of them provides a complex view of the landscape with a particular focus on EOSC. WG [*Working Group*] Landscape provided a snapshot of the EOSC – related environment in MS […] at the national and institutional level." in EOSC Executive Board, "Strategic Research and Innovation Agenda", February 2021, https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf.

For this reason, the evolution of the European Open Science policies has been mapped out, from 2015[17] to the present. In order to fill the first preliminary gap, a conceptual framework for the multiple interventions of the European institutions in the field of Open Science has been proposed[18]. In doing so, the essential steps of this evolution have been highlighted, striving for a consistent interpretation of the European vision.

Subsequently, concerning the second gap, attention was drawn to the European regulations most involved and problematic in the implementation of the Open Science paradigm. The focus was on: (i) the discipline of Open Data; (ii) the discipline of data protection; (iii) the set of rules on copyright and the ownership of data; and (iv) looking forward, the Regulation proposal, presented by the European Commission in 2020, called "Data Governance Act"[19].

7.1.2 Legal Coordination Approach

In light of this complex European framework, the need to investigate the most suitable approach to the governance of Open Science emerged. Two opposed and polarised positions were identified: on the one hand, the supporters of a specific discipline for Open Science, able to regulate it in an all-encompassing discipline; on the other hand, the deniers of the feasibility of Open Science based on the assumption that this new paradigm is incompatible with the current EU legal framework.

Between these two polarised positions, a third way has been proposed to address the Open Science governance: the legal cooperation

[17] Precisely, a formal introduction of the Open Science in the EU policies has been identified in 2015. However, it was argued that the substantial introduction of the Open Science in the EU architecture is found in the establishment of the European Research Area (ERA) in the early 2000s and the identification of the fifth European freedom of movement, i.e., the free circulation of knowledge. In other words, it has been claimed that the concept of Open Science appears strongly embedded in the foundations of the EU system.

[18] See: Section 2.2.

[19] European Commission, Proposal for a Regulation of the European Parliament and of the Council, *on European data governance (Data Governance Act)*, COM/2020/767, 25.11.2020, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767.

approach. This approach intends to take into account the multiplicity of regulatory systems (e.g., law; practices of the scientific community; market forces; technology, etc.); the levels of intervention (i.e., international, European, national, and local level related to specific research centers or universities); and the actors (e.g., public or private researchers; students; scientific publishers; institutions, database managers; etc.). In this scenario, the European institutions should coordinate different instances, constantly striking a balance between the multiple opposing interests and rights at stake[20].

The report of the European Commission entitled "Monitoring the open access policy of Horizon 2020" showed that 83% of scientific publications and 95% of research data from projects funded by the Horizon 2020 complied with its open access policies[21]. This result was recognised as a real success since it has been achieved in "[…] a decentralised European environment in which Member and Associate countries have different policies and infrastructures (or lack thereof)"[22]. These results also show that the Open Science paradigm, striving for greater transparency and communication, is indeed feasible, despite the unfavourable conditions.

This means that if these unfavourable conditions were addressed, the progress of the European scientific research would be enormous.

[20] Consider that, within an information society such as the one in which we live, this legal conflict between opposing interests also requires an ethical balancing act; on this aspect, i.e., the meaning that this balancing operation should have, and on the role of information in this balancing action, see: MASSIMO DURANTE, "Dealing with legal conflicts in the information society. An informational understanding of balancing competing interests." *Philosophy & Technology*, 26.4 (2013): pp. 437-457, doi: 10.1007/s13347-013-0105-z. Going further, on the assumption that "[…] such a balance can only be reached once individual rights are clearly defined", see: MARIAROSARIA TADDEO, "The Struggle Between Liberties and Authorities in the Information Age.", *Science and Engineering Ethics volume* 21, (2015): 1125-1138., doi: 10.1007/s11948-014-9586-0. The Author develops the reasoning with regard to the information society, but it is also applicable to the governance of scientific research and the balancing of opposing interests that it requires today.

[21] European Commission, *Monitoring the open access policy of Horizon 2020: final report*, Publications Office (2021), 1-122, doi: 10.2777/268348.

[22] European Commission, *Monitoring the open access policy, op. cit.,* p. 10.

Therefore, coordination between the various systems, levels and actors should precisely be addressed to redesign these starting conditions[23].

7.1.3 Open Science Definition

This initial analysis brought out multiple dimensions of the Open Science paradigm, which required to untangle the problematic definition of the Open Science paradigm. To this aim, an analysis of the concept of Open Science has been proposed, from its origins in Modern Science to the various meanings it has assumed in recent years.

Several definitions have been investigated, but none of them resulted adaptable to the aims pursued in this dissertation. In fact, in order to develop considerations on the policy of a phenomenon it is necessary to have in mind the constituent elements of that phenomenon.

Open Science is by nature a dynamic phenomenon: an overly analytical definition of its elements would have led to the exclusion of some dimensions that might emerge in the future, with the potential risk of thwarting projects or strategies. Based on these needs, therefore, a definition of Open Science has been developed by adopting Floridi's method of Levels of Abstraction (LoA)[24]. The LoA adopted to investigate the phenomenon was that of the process, identifying Open Science as the Open Scientific Research Process. The five observables detected represent the constitutive elements of the Open Science paradigm and are: (O_1) Resources; (O_2) Actors; (O_3) Methods; (O_4) Tools; (O_5) Benefits.

This enabled the identification of Open Science as follows:

---

[23] At the basis of this approach there is the recognition that the digital revolution has generated a multi-agent system. If this has occurred in general, it has similarly happened in the sphere of scientific research. The resulting coordination has also been identified as a guiding principle between traditional and electronic media in the public discourse, which is also applicable to the topic under investigation. See: MASSIMO DURANTE, "E-democracy as the Frame of Networked Public Discourse Information, Consensus and Complexity." *25th IVR World Congress: Law, Science and Technology* 20 (2012): 1-28.
[24] See: Section 3.2.

> the Open Science describes the Open Scientific Research Process in which openness regards (i) the inputs, i.e., the underlying resources, such as data and economic funding; (ii) actors, including public at large; (iii) methods, in terms of evaluation and verifiability; (iv) tools; and (v) benefits deriving from this process, including the dissemination of the results, through publication, teaching activities or more.

In order to complete this study about the notion of Open Science, the foundations of this new paradigm have been grounded in the framework of human and fundamental rights. This analysis led us to argue that Open Science policies should be considered the means by which effectively enforcing the right to science as enshrined in the Article 27 of the Declaration of Human Rights, reinforced by the Article 15 of the International Covenant on Economic, Social and Cultural Rights.

As a result, the right to science is not only at the basis of the current Open Science paradigm but may be interpreted as a real right to *open* science.

7.1.4 Research Data, FAIRness and Legal Interoperability

Focusing on one of the triggers of the Open Scientific Research Process, i.e., research data, the need to clarify many different and sometimes overlapping concepts emerged. In particular, providing a definition of research data, instrumental for the purposes of our dissertation, resulted necessary[25]:

> Research data are factual records, (i) used for research purposes; (ii) as part of a process of analysis traditionally called scientific method; (iii) enabling the reproducibility of experiments, in order to ensure the review of the research output by the scientific community.

Second, the concept of data FAIRness, a pillar of the Open Science, was investigated. FAIR is an acronym that stands for findable,

[25] See: Section 4.4.1.

accessible, interoperable, and reusable. The FAIR Data Principles are a set of guidelines called upon to standardise the management of data in scientific research, from a technical point of view. However, having FAIR data does not mean having open data, but it does mean having good research data management.

In addition, attention was drawn to the vague concept of legal interoperability. The term "legal interoperability" is often mentioned in the implementation process of the EOSC and in other major EU Open Science projects to refer to the whole range of legal-cultural issues related to the openness of research data.

Due to this vagueness, the notion of legal interoperability has been redefined emphasising two dimensions, i.e., the coordination and the harmonisation: the legal interoperability as a co-ordination process aims to enable the legal reuse of data, taking into account the multiple disciplines involved (e.g., copyright, database regulation, data protection, etc.); and the legal interoperability as an harmonisation process relates to the barriers resulting from the possible multiple jurisdictions involved in sharing and reuse of research data[26].

7.1.5 Openness by Default in Hard Law

Having clarified the main concepts at stake, related to open research data, attention was drawn to the EU Open Data framework, with specific reference to the recent Open Data Directive (ODD)[27].

The openness of research data is understood in terms of sharing and reuse. The European lawmaker, in the recent ODD, established the openness by default of research data, as the standard option. Adopting an opt-out approach, the EU institutions provide for a list of reasons that can justify the decision to close the research data. For the first

[26] See: Section 4.2.1.

[27] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019, *on open data and the re-use of public sector information*, OJ L 172, 26.6.2019, 56–83, ELI: http://data.europa.eu/eli/dir/2019/1024/oj.

time in hard law, research data closure becomes an exception and openness the rule.

Three main obstacles to the effective openness of research data have been investigated[28]: (i) the limitations arising from national transpositions and the consequent fragmentary nature of the discipline; (ii) the enforceability difficulties of the "open by default" clause of research data; (iii) the choice of the suitable licence to apply to sharing and reuse of such data.

Finally, going further, the analysis of the legal challenges of the open research data was integrated by investigating the mechanism of "data altruism" introduced by the "Data Governance Act" (DGA), the recent proposal of Regulation of the European Commission[29].

The data altruism mechanism is represented by the consent of the data subject or the permission of the data holder to allow the use of their data (personal data or proprietary data), without compensation, for purposes of general interest, "[…] such as scientific research purposes or improving public services"[30]. The aim was to assess the impact of the DGA provisions on the field of scientific research. A number of concerns have been raised, mainly related to the actual feasibility of this mechanism and the supposed centralised approach of data and power that seems to result from it.

7.1.6 An Infraethical Framework for Research Data

Regarding the ethical issues related to the research data in the Open Science context, the focus was on data quality. It is difficult to define the notion of high quality research data, but it is essential *to have* high quality research data. Too often the relevance of the quality is forgotten, but it is necessary and indispensable not only from the point of view of scientific research but also from that of the public deliberation and decision-making process that may rely on it. In this

[28] See: Section 4.2.4.
[29] See: Section 4.4.
[30] The Article 2(10), DGA.

perspective, according to Leonelli, data quality requires to be characterised by four main features, i.e., updated data; reliability; awareness of bias; and accuracy.

While the European lawmaker can impose the openness of research data by default (albeit with the limitations discussed), it is not its competence to define quality standards. This is to be found above all in the deontological charters and codes of conduct, in the forms of self-regulation of the scientific community. The intention was to offer an interpretation of the link between the three fundamental concepts of research data management, which are substantially different: the openness of research data; the FAIRness of such data; and their quality. The relation of these three concepts has been described adopting Floridi's notion of infraethics: the FAIRness of data concerns the infraethical level; the data quality concerns the ethical level; while the openness of research data – expressed in terms of sharing and reuse – is the goal to achieve[31]. The most careful and meticulous respect of the FAIR Data Principles (infraethics) may improve the dissemination and sharing of research data but does not improve – *per se* – the very quality of the research data shared (ethics). By contrast, research data created or collected and processed with the highest quality are wasted if they do not respect those technical guidelines, i.e., the FAIR Data Principles, that allow them to be found, accessed, be interoperable and eventually reused. This interpretation of the three fundamental concepts of quality, openness and FAIRness, may help the definition of the guidelines and the deontological charters of universities and research centers, dealing with the best practices for the research data management.

7.1.7 Data Protection and Open Science: A Win-Win Scenario

A crucial part of the management of research data is represented by the protection of personal data processed in the research domain. For

[31] See: Section 4.3.2.

this reason, it has been investigated whether and to what extent the emergence of the Open Science paradigm is complexifying the scenario.

After an overview of the data protection regulation at European level, i.e., the GDPR, and about the role of the Member States, the main data protection issues in the field of scientific research have been identified[32]. They are: (i) the adoption of the lawful legal basis for processing activities pursuing research purposes; (ii) the compliance with the principle of data minimisation and the subsequent problems related to the anonymisation techniques; and (iii) the challenges posed by the transfer of personal data to third countries and international organisations. This analysis showed that the major data protection issues in the context of research are not generated by the emergence of the Open Science paradigm. Rather, in the Open Science scenario, the previously existing weaknesses of the system tend to be more evident. The purpose here was to provide some suggestions to tackle these weaknesses, guided by the tenets of the Open Science.

In doing so, concerning the first issue related to the legal bases, a set of guidelines has been proposed: the aim was to orientate the choice of the lawful legal basis in the case-by-case assessment that researchers are required to make when processing personal data for research purposes.

Regarding the principle of minimisation and its observance in the context of Open Science, a three-step reasoning has been presented. It concerned: (a) the distinction between the concepts of openness and accessibility; (b) the application of the principle of accountability as a meta-principle; (c) and finally the adoption of the reasonableness test, inspired by the Italian legislator.

About the third issue, represented by the transfer of personal data to third countries and international organisations after Schrems II, an

[32] See: Section 5.3.

alternative option for the research domain has been envisaged: the adoption of codes of conduct, provided by the EOSC association.

In light of this analysis, an additional reasoning in support of the convergence between data protection and Open Science was carried out, focusing on the barriers still persisting today for an effective protection of personal data in the field of Open Science. These barriers have been found in (i) the national fragmentation of the EU data protection discipline; (ii) the lack of data governance at local level; and (iii) a closed science, without transparency and inclusiveness.

In particular, the lack of data governance strategies at local level is urgent. If data have a significant role in today's science, then it is crucial to take into account their management and stewardship. The protection of personal data of data subjects involved in research projects needs to become a part of a broader management of research data.

7.1.8 The *Inforg* of Science

The trigger of the emergence of the Open Science paradigm has been the impact of digital ICTs on the scientific research domain. Therefore, technologies and e-infrastructures, i.e., the tools of Open Science according to the definition proposed in this dissertation, must be carefully considered in designing the implementation of the Open Science paradigm.

The analysis of the e-infrastructures in the context of Open Science, started from the assumption that today human agents (i.e., researchers) and artificial agents (i.e., machines) cooperate and coexist in science: for this reason, the Open Science actors has been defined as *inforg* of science[33].

After an analysis on the evolution of research e-infrastructures in the last decades, the attention was drawn to one computing infrastructure, namely the High Performance Computing (HPC)

[33] See: Section 6.1.

facility. After defining the HPC, the investigation initially adopted a European perspective, focusing on the development of the policies in this domain; and then, adopted a local perspective, by presenting the case study of the ULHPC.

The European HPC strategy seems intended to take full advantage of technologies, specifically for the benefit of scientific research. At the basis of this strategy, the interplay between the public and private sector is even more evident.

For this reason, it was argued that the problem should not be framed in the alternative between participation or exclusion in science of private actors. Rather, the trade-off is between transparent or opaque interplay between public and private actors in scientific research[34]. This need became even clearer in light of the analysis of the “Clarifying Lawful Overseas Use of Data Act”, i.e., the CLOUD Act, enacted by the Congress of the United States, in 2018[35].

This piece of legislation is relevant to the European perspective because supports the US public authorities to overcome “[…] inability to access data stored outside the United States that is in the custody, control, or possession of communications-service providers that are subject to jurisdiction of the United States”[36]. In other words, the CLOUD Act allows US authorities to access data, including personal data, of individuals prosecuted in legal proceedings, even if such data are located in the EU.

The main issues, raised by the CLOUD Act, analysed in Chapter 6 are: (i) the conflict with the discipline provided by the GDPR; (ii) and the delegation of quasi-judicial power, in favour of private service providers.

[34] See: Section 6.3.1.
[35] Clarifying the Lawful Use of Overseas Data Act, CLOUD Act, Pub. L. No. 115–141, 132 Stat. 348 (codified as amended in separate sections of 18 U.S.C.), 2018, available at: https://cli.re/BwPk5Q.
[36] CLOUD Act, §2201 (13-17).

Against the backdrop of the major findings of this dissertation, what is the gist of the analysis on the European approach to Open Science? Next section intends to expose it, by highlighting some final remarks and policy considerations.

7.2 The Gist of the Analysis: Final Remarks and Policy Considerations

The European Union carried out many initiatives in the field of Open Science over the years. At international level, the EU, with its policies on Open Science, is perceived as a leading force in this domain.

The scenario is extremely complex, due to the plurality of actors concerned (e.g., scientific community, institutions, students, service providers, etc.), the different levels of regulation involved (i.e., European, international, national, and local interventions) and the regulatory systems that compete (i.e., not only law, but the practices of the scientific community, the market laws pursued by private actors, etc.).

In this scenario, what is the assessment of the EU policies?

Overall, the evaluation is positive. Much has been done in recent years, and much is planned for the future. The European institutions stand out for having intercepted the needs of part of the scientific community, towards a more cooperative, global science, not inward-looking, but open to the world and to the technological innovation. In part, these demands have been met, or at least the foundations have been laid to meet these demands. However, much remains to be done, and three policy considerations need to be taken into account.

First, European projects launched in the field of Open Science must not be distorted. A complex scenario inevitably brings with it a large number of interests pointing in opposite directions. In the field of scientific research funded with public money, checks and balances mechanisms should always strive to safeguard the fundamental and human rights framework, discussed in Chapter 3, which foster the circulation of knowledge, access by all to the scientific results, and democratic participation. This may not always be easy, but it is essential.

However, it is also likely that other interests prevail. The risk, in this case, would be to transform Open Science policies into empty shells. The European Union would have the strength and the potential to promote and make effective and concrete this new and emerging paradigm of science. It remains to be seen what will prevail.

The second policy consideration is related to the risk of overregulation. The pitfall is to convey policies by merely issuing top-down regulations. On one side, the adoption of hard law regulation can be beneficial, especially in the early stages of such a major change as the one generated by the emergence of Open Science. On the other side, the introduction of new practices or rules through hard law should not always be seen as the only solution. As Stefano Rodotà pointed out, "[…] society cannot be changed by mere legislation, but by a widespread culture"[37].

As an illustration, consider the introduction of the "open by default" clause for research data in the ODD. On the one hand, this provision deserved the merit of imposing a national debate on the management of research data (at least in some Member States, such as France), that perhaps only a new piece of legislation coming into force can trigger. On the other hand, the limits of this provision in terms of enforceability have been pointed out in this dissertation: to some extent the final application of this provision is entrusted to the researcher. The field of scientific research has always been characterised by a high level of autonomy, primarily linked to the academic freedom, which must

[37] STEFANO RODOTÀ, *Intervista su privacy e libertà*, PAOLO CONTI (eds.) (Bari: Laterza Editore, 2005). p. 36. [Translation from the Italian original text]. The Author claims the mentioned assumption in a different context, related to the establishment of the privacy and data protection legislations, but it is also effective for the topic under investigation.

always be taken into account when intervening in this area. For this reason, it is equally important to proceed in parallel to the entry into force of new provisions, with an intervention at the cultural level: to support that part of the scientific community still reluctant to adopt the tenets of Open Science. To achieve this result, in the field of research data, for instance, data stewardship could be the real game changer: data stewards can assist researchers in implementing Open Science policies and practices, providing support in both technical and legal terms.

Finally, the third consideration is about the required inclusiveness of the Open Science paradigm. The European Open Science policies will be successful and able to make a new paradigm of science emerge and flourish if and only if they will be as inclusive as possible. This inclusiveness can be considered in two dimensions, i.e., the horizontal and the vertical dimension. Horizontal inclusiveness refers to an equal and effective advancement in all the 27 Member States of the European Union. This is because, in terms of competence between the EU and the Member States, the latter enjoy a great autonomy (e.g., in data protection or in the transposition of directives).

However, inclusiveness should also be vertical, involving the local level, in each Member State: good management of research data, use and promotion of new technologies, open research practices, and global openness should involve all the research organisations, from the large, well-established university with centuries of history, to the small university or research centers, a stronghold of culture on the European landscape. This form of vertical inclusiveness is essential, precisely because of the aforementioned independence of the research domain, which has traditionally enabled even small universities or research centers to flourish. And how can this vertical inclusiveness be fostered? I believe that a key role in this instance is played by coordination at national level. This coordination can start from the research e-infrastructure networks, as the GARR *consortium* is doing in Italy.

GARR is a *consortium* set up in the 1990s, mainly by the concerted effort of the Italian national research organisation (CNR), the conference of the Italian university rectors (CRUI) and other research associations[38], in agreement with the Italian Ministry of Education. While GARR's initial purpose was to coordinate and standardise computer networks for research, its remit has now expanded to become a benchmark for the entire world of research. The GARR *consortium* has set up a working group, the ICDI, i.e., the "Italian Computing and Data Infrastructure", which acts as the Italian representative for European projects on Open Science and digital innovation: "ICDI (Italian Computing and Data Infrastructure) is a *forum* created by representatives of major Italian Research Infrastructures and e-Infrastructures, with the aim of promoting synergies at the national level, and optimising the Italian participation to European and global challenges in this field, including the European Open Science Cloud (EOSC), the European Data Infrastructure (EDI) and HPC"[39].

I believe that much can be done for vertical inclusiveness by associations operating at national level[40]. It should not be forgotten that the entities that will fall behind in the implementation of Open Science will also be the universities and research centers that will lose the digital challenge.

Luciano Floridi, outlining his political ideas for a mature information society, stated that "[…] every society incorporates its own

[38] Initially, the GARR *consortium* emerged by the contribution of the Italian national research organisation (i.e., CNR, *Consiglio Nazionale delle Ricerche*); the conference of Italian university rectors (i.e., CRUI, *Conferenza dei Rettori delle Università Italiane*); the national institute on nuclear physics (i.e., INFN, *Istituto Nazionale per la Fisica Nucleare*); the national agency for the promotion of new technologies, sustainable energy, and economic development (i.e., ENEA, *Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile*); the Italian institute of astrophysics (i.e., INAF, *Istituto Nazioale di Astrofisica*); and the national institute of geophysics and volcanology (i.e., INGV, *Istituto Nazionale di Geofisica e Vulcanologia*). See: https://www.garr.it/it/chi-siamo.

[39] https://www.icdi.it/en/about.

[40] Another worthy example is the National Interuniversity *Consortium* for Informatics (CINI), which strives to coordinate HPC developments in Italy, see: Marco Aldinucci, *et al.*, "The Italian research on HPC key technologies across EuroHPC." *Proceedings of the 18th ACM International Conference on Computing Frontiers* (2021): p. 178, doi: 0.1145/3457388.3458508.

human project, whether assumed only implicitly or pursued explicitly […] Not having a project does not mean choosing to proceed without it, but only opting for a bad project. Therefore, a society without a human project does not exist"[41]. Adapting his thinking and reasoning to the field of scientific research, it is believed that Open Science is the human project for scientific research. From here on, the development will be a matter of design.

7.3 Open Problems for Open Science

Today, Open Science is no longer just an ideal, but a concrete reality. However, as mentioned above, it is a relatively recent phenomenon: it is now facing the most delicate phase, that of actual implementation[42].

In doing so, developments at the European level will certainly have to be monitored, in order to proceed with the investigation in this domain.

From a legal point of view, looking to the future and possible further investigations related to the research presented in this dissertation, it will be worthwhile to make a comparative analysis of the various implementations of the Open Data Directive (ODD) in order to fully understand its impact on scientific research. It is fair to wonder how sharing and reuse of research data can be promoted, given a fragmented legal framework.

The research topic may also benefit from further investigation in relation to the issue of transfer of personal data to third countries or international organisations. It is well-known that universities and research centers use non-European service providers for both research and teaching purposes. If these providers are US-based, the transfer of personal data currently has no legal basis.

[41] LUCIANO FLORIDI, *Il verde e il blu. Idee ingenue per migliorare la politica.* (Milano: Cortina Editore, 2020), p. 104. [Translation from the Italian original text].
[42] On this aspect see, Section 3.4.

By contrast, if these US providers have a European representative and their data centers are located on the EU territory, there would be no transfer of data to third countries, but other concerns arise. In this instance, a strain could be generated by the application of the CLOUD Act. It will therefore be necessary to monitor the situation and see in which direction the European institutions intend to proceed.

Consider that on 22 December 2021, the Austrian national data protection authority, i.e., Datenschutzbehörde or DSB, fined the owner of a website for using Google's services for analysing and tracking users' data, i.e., Google Analytics: the *ratio* of this decision was the infringement of the provisions of the GDPR on the transfer of personal data to third countries[43]. It should be expected that this measure will be followed by further, similar measures, from other national authorities. The field of scientific research is not immune, in fact to a certain extent it is even more vulnerable, which is why it requires greater consideration and awareness.

Further research could be conducted to investigate particularly relevant national initiatives related to this topic. For instance, the CNIL, the French national data protection authority, announced on 18 January 2022 its intention to launch a sandbox devoted to the research and education domain: "[T]he establishment of this sandbox will allow the education sector to better understand the challenges of the GDPR compliance, which are necessary for the development of privacy-friendly offers in the context of public procurement. It will also enable the deployment of trusted school and university solutions for users (teachers, students, parents, institutions, etc.)"[44].

---

[43] This measure resulted from a series of complaints submitted by the Association "noyb – European Center for Digital Rights", founded by Max Schrems, see: noyb European Center for Digital Rights, "Austrian DSB: EU-US data transfer to Google Analytics illegal", (2022), https://noyb.eu/en/austrian-dsb-eu-us-data-transfers-google-analytics-illegal.

[44] CNIL, "La CNIL propose un nouveau « bac à sable » pour accompagner l'innovation numérique dans le domaine de l'éducation", (2022) https://www.cnil.fr/fr/la-cnil-propose-un-nouveau-bac-sable-pour-accompagner-linnovation-numerique-dans-le-domaine-de. [Translation from the French original text].

The choices that the actors involved in scientific research are required to make in this critical phase of implementation of the Open Science must be well balanced and have to take into account the three final policy considerations developed in the previous section, trying to be as close as possible to the original purposes; avoiding overregulation; and ensuring full inclusiveness.

The choices made in the near future by the Open Science actors will be decisive in unlocking the enormous potential of the technological progress for the benefit of society and the public interest: more than in other domain, a strong "joint commitment"[45] will be essential.

[45] Margaret Gilbert, *Joint commitment. How we make the social world.* (Oxford: OUP, 2013), pp. 37-57.

Bibliography

MONOGRAPHIES, SCIENTIFIC PAPERS, ACADEMIC CONTRIBUTIONS

ABRAHA, H. H. (2019) “How compatible is the US ‘CLOUD Act’ with cloud computing? A brief analysis.” *International Data Privacy Law* 9.3, 207-215, doi: 10.1093/idpl/ipz009.

AGARWAL, A., *et al*. (2016) “Bibliometrics: tracking research impact by selecting the appropriate metrics.” *Asian journal of andrology* 18.2, 296-309, doi: 10.4103/1008-682X.171582.

AJANI, G., CAVALLO PERIN, R., GAGLIARDI, B. (2017) “L’Università: un’amministrazione pubblica particolare.” *Federalismi.it* 14, 2-16.

ALBERT, M., LABERGE, S., MCGUIRE, W. (2012) “Criteria for assessing quality in academic research: the views of biomedical scientists, clinical scientists and social scientists.” *Higher Education* 64.5, 661-676, doi: 10.1007/s10734-012-9519-2.

ALDINUCCI, M., *et al*. (2021) “The Italian research on HPC key technologies across EuroHPC.” *Proceedings of the 18th ACM International Conference on Computing Frontiers*, 178-184, doi: 0.1145/3457388.3458508.

ALIPRANDI, S. (2014) *Il fenomeno Open Data: indicazioni e norme per un mondo di dati aperti*. (Milano: Ledizioni).

ALOVISIO, M., BASSI, E. (2013) “Protezione dei dati personali e riutilizzo dell’informazione del settore pubblico”, in RICOLFI, M, SAPPA (eds.) *Quaderni del Dipartimento di Giurisprudenza dell’Università di Torino*, (Napoli: Edizioni Scientifiche Italiane), 193-246.

ALPER, P., BECKER, R., *et al*. (2018) “Provenance-enabled stewardship of human data in the GDPR Era.” *Provenance and Annotation of Data and Processes* (Cham: Springer), 266-269, doi: 10.1007/978-3-319-98379-0_33.

ALSHEIKH-ALI, A. A., *et al*. (2011) “Public availability of published research data in high-impact journals.” *PloS one* 6.9, e24357.

ALSOP, T. (2021) “Revenue from broader high performance computing (HPC) market worldwide from 2015 to 2024, by segment (in million U.S. dollars)”, *Statista*, https://www.statista.com/statistics/724916/worldwide-broader-hpc-revenue-by-segment/.

ANDERSON, C. (2008) "The end of theory: The data deluge makes the scientific method obsolete", *Wired magazine*, https://www.wired.com/2008/06/pb-theory/.

ANDERSON, M. (2018) "How to draft a licence agreement that is fair, reasonable and non-discriminatory: a ten-point plan." *Journal of Intellectual Property Law & Practice*, 13.5, 377-392, doi: 10.1093/jiplp/jpx212.

ASPESI, C., *et al.* (2019) "Landscape Analysis: The Changing Academic Publishing Industry – Implications for Academic Institutions." *SPARC*.

AURUCCI, P. (2019) "Legal Issues in Regulating Observational Studies: The Impact of the GDPR on Italian Biomedical Research." *European Data Protection Law Review* 5(2), 197-208, doi: 10.21552/edpl/2019/2/9.

AYRIS, P. *et al.* (2016) "Realising the European open science cloud", *Publications Office of the European Union*, 1-24, doi: 10.2777/940154.

BACHELARD, G. (1949) *Le rationalisme appliqué*. (Paris: Presses universitaires de France).

BAKER, M. (2016) "Reproducibility crisis." *Nature* 533.26, 353-366, doi: 10.1038/nature17990.

BARTLING, S., FRIESIKE, S. (2014) *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. (Cham: Springer Nature), http://library.oapen.org/handle/20.500.12657/28008.

BASSI, E. "Dati, Sovranità, Nuovi Modelli di Governance." in PAGALLO, U. DURANTE, M. (eds.), *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società*. (Milano-Udine: Mimesis), 203-220.

BASSI, E. (2011) "PSI, protezione dei dati personali, anonimizzazione." *Informatica e diritto* 37.1-2, 65-83.

BEERKENS, E. (2008) "The Emergence and Institutionalisation of the European Higher Education and Research Area." *European Journal of Education* 43, 407–425, doi: 10.1111/j.1465-3435.2008.00371.x.

BEISBART, C. (2012) "How Can Computer Simulations Produce New Knowledge?", *European Journal for Philosophy of Science*, 2.3, 395-434, doi: 10.1007/s13194-012-0049-7.

BENKLER, Y. (2008) *The wealth of networks: how social production transforms markets and freedom*. (New Haven: Yale University Press).

BERBERICH, F., *et al.* (2019) "European HPC Landscape." *2019 15th International Conference on eScience (eScience)*, 471-478, doi: 10.1109/eScience.2019.00062.

BERNERS-LEE, T. (2000) *Weaving the Web. The original design and ultimate destiny of the World Wide Web*. (New York: HarperBusiness).

BERRA, M., MEO, A. R. (2006) *Libertà di software, hardware e conoscenza. Informatica solidale 2*, (Torino: Bollati Boringhieri).

BERTACCHINI, V., *et al.* (2021) *EOSC Symposium Report*, Zenodo, doi:10.5281/zenodo.5176089.

BERTI SUMAN, A., PIERCE, R. (2018) "Challenges for citizen science and the EU open science agenda under the GDPR." *European Data Protection Law Review* 4, 284-295, doi: 10.21552/edpl/2018/3/7.

BINDA, F., CASO, R. (2020) "Il Diritto Umano alla Scienza Aperta." *Trento Law and Technology Research Group Research Paper Series*, 41, 1-6.

BOBBIO, N. (1987) *The future of democracy: a defence of the rules of the game*. (Minneapolis: University of Minnesota Press).

BOECKHOUT, M., ZIELHUIS, G. A., BREDENOORD, A. L. (2018) "The FAIR guiding principles for data stewardship: fair enough?." *European journal of human genetics* 26.7, 931-936, doi: 10.1038/s41431-018-0160-0.

BONIFIGLIO, F. (2021) *Gaia X Vision & Strategy*, https://gaia-x.eu/news/gaia-x-association-releases-its-vision-and-strategy-document.

BONIOLO, G. (2021) *Il virus dell'idiozia. Sette scrittini su Covid-19, scienza, intellettuali e cittadini*. (Milano-Udine: Mimesis).

BONIOLO, G. (2021) *Molti. Discorso sulle identità plurime*. (Torino: Bollati Boringhieri).

BONIOLO, G. (2018) *Conoscere per vivere: Istruzioni per sopravvivere all'ignoranza*. (Milano-Udine: Mimesis).

BONIOLO, G. (2012) *The art of deliberating: democracy, deliberation and the life sciences between history and theory*. (Berlin, Heidelberg: Springer Science & Business Media), doi: 10.1007/978-3-642-31954-9.

BONIOLO, G. (2011) *Il pulpito e la piazza. Democrazia, deliberazione e scienze della vita*. (Milano: Cortina Editore).

BONIOLO, G. (2003) *Il limite e il ribelle. Etica, naturalismo, darwinismo*. (Milano: Cortina Editore).

BONIOLO, G. (1999) *Metodo e rappresentazione. Per un'altra filosofia della scienza*. (Milano: Mondadori).

BONIOLO, G., CAMPANER, R. (2020) "Life Sciences for Philosophers and Philosophy for Life Scientists: What Should We Teach?" *Biological Theory* 15, 1-11, doi: 0.1007/s13752-019-00333-7.

BONIOLO, G., ONAGA, L. (2021) "Seeing clearly through COVID-19: current and future questions for the history and philosophy of the life sciences." *History and Philosophy of the Life Sciences*, 43, 81-83, doi: 10.1007/s40656-021-00434-2.

BONIOLO, G., VACCARI, T. (2012) "Alarming shift away from sharing results." *Nature* 488.7410, 157.

BONIOLO, G., VIVALDI, P. (2011) *Strumenti per ragionare. Logica e teoria dell'argomentazione*. (Milano: Mondadori).

BONNEY, R., *et al.* (2014) "Next steps for citizen science." *Science* 343.6178, 1436-1437, doi: 10.1126/science.1251554.

BORONOW, K. E., *et al.* (2020) "Privacy risks of sharing data from environmental health studies." *Environmental health perspectives* 128.1.017008,1-12, doi: 10.1289/EHP4817.

BOULTON, G. (2011) "University rankings: Diversity, excellence and the European initiative." *Procedia-Social and Behavioral Sciences* 13, 74-82, doi: 10.1016/j.sbspro.2011.03.006.

BRADFORD, A. (2012) "The Brussels effect." *Northwestern University Law Review* 107, 1-65.

BRAMBILLA PISONI, G., TADDEO, M. (2021) "Apropos Data Sharing: Abandon the Distrust and Embrace the Opportunity.", *DNA and Cell Biology* 41.1, 11-15, doi: 10.1089/dna.2021.0501.

BRIA, F., BAIN, M. (2018) "Manifesto in favour of technological sovereignty and digital rights for cities", version 2.0, https://www.barcelona.cat/digitalstandards/manifesto/0.2/.

BRUCE, A., *et al.* (2004) Interdisciplinary integration in Europe: the case of the Fifth Framework programme." *Futures* 36.4, 457-470, doi: 10.1016/j.futures.2003.10.003.

BUDRONI, P., BURGELMAN, J-C., SCHOUPPE, M. (2019) "Architectures of knowledge: The European Open Science Cloud." *ABI Technik* 39.2, 130-141, doi:10.1515/abitech-2019-2006.

BURGELMAN, J-C. (2021) "Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story)." *Data Intelligence* 3.1, 5-19, doi: 10.1162/dint_a_00069.

BURGELMAN, J-C., *et al.* (2019) "Open science, open data and open scholarship: European policies to make science fit for the 21st century." *Frontiers in Big Data 2*.43, 1-6, doi: 10.3389/fdata.2019.00043.

BURGOS, D. (2020) *Radical Solutions and Open Science: An Open Approach to Boost Higher Education*. (Singapore: Springer Nature), doi:

10.1007/978-981-15-4276-3.
BUTTARELLI, G. (2016) "The EU GDPR as a clarion call for a new global digital gold standard." *International Data Privacy Law*, 6.2, 77-78, doi: 10.1093/idpl/ipw006.
CAPORAEL, L. R., *et al.* (2014) *Developing Scaffolds in Evolution, Culture and Cognition*. (Cambridge: MIT Press).
CARBALLA-SMICHOWSKI, B., DUCH-BROWN, N., MARTENS, B. (2012) "To pool or to pull back? An economic analysis of health data pooling." *JRC Digital Economy Working Paper*, 1-75.
CAROLAN, E. (2016) "The continuing problems with online consent under the EU's emerging data protection principles." *Computer Law & Security Review* 32.3, 462-473, doi: 10.1016/j.clsr.2016.02.004.
CASO, R. (2021) *La società della mercificazione e della sorveglianza: dalla persona ai dati: casi e problemi di diritto civile*. (Milano: Ledizioni).
CASO R. (2019) *La rivoluzione incompiuta: La scienza aperta tra diritto d'autore e proprietà intellettuale*. (Milano: Ledizioni).
CASO, R. (2010) "Alle origini del copyright e del diritto d'autore: spunti in chiave di diritto e tecnologia." in IZZO, U. (ed.), *Alle origini del copyright e del diritto d'autore: tecnologia, interessi e cambiamento giuridico*, (Roma: Carocci), 1-24.
CASO, R. (2009) "Open Access to legal scholarship and copyright rules: a law and technology perspective." *Proceedings law via the Internet: free Access, quality of information, effectiveness of rights*, 97-110.
CASSESE, S. (2021) *Intellettuali*. (Bologna: Il Mulino).
CAVALLO PERIN, R. (2021) "Il Contributo Italiano alla Libertà di Scienza nel Sistema delle Libertà Costituzionali" *Diritto Amministrativo*, 3, 587-620.
CHANG, W. L., GRADY, N. (2015) "NIST big data interoperability framework: Volume 1, big data definitions." *National Institute of Standards and Technology*, 1-43, doi:10.6028/NIST.SP.1500-1r2.
CHAPMAN, A., WYNDHAM, J. (2013) "A human right to science." *Science* 340.6138, 1291, doi: 1126/science.1233319.
CHAITIN, G., J. (1977) "Algorithmic information theory." *IBM journal of research and development* 21.4, 350-359.
CHUBIN, D., E. (1985) "Open science and closed science: Tradeoffs in a democracy." *Science, Technology, & Human Values* 10.2, 73-80.
CIANI, J. (2021) *Il pubblico dominio nella società della conoscenza*. (Torino: Giappichelli Editore).
CIANI, J. (2018) "Governing Data Trade in Intelligent Environments: Taxonomy of Possible Regulatory Regimes between Property

and Access Rights." *Workshop Proceedings of the 14th International Conference on Intelligent Environments, Ambient Intelligence and Smart Environments Series*, 23, 285-297, doi: 10.3233/978-1-61499-874-7-285.

CIESIELSKA, M., JEMIELNIAK, D. (2021) "Fairness in digital sharing legal professional attitudes toward digital piracy and digital commons." *Journal of the Association for Information Science and Technology*, 1-14, doi: 10.1002/asi.24592.

CLEVELAND, W. S. (2001) "Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics." *International statistical review* 69.1, 21-26, doi: 10.2307/1403527.

COLAVIZZA, G., *et al.* (2021) "A scientometric overview of CORD-19." *Plos one* 16.1, 1-18, doi: 10.1371/journal.pone.0244839.

COLLINS, S., *et al.* (2018) "Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data." *Publications Office of the European Union*, 1-76, doi: 10.2777/54599.

CORDINI G. (1995) "Università istituzioni e imprese, aspetti di diritto comparato." *Politico* 60.3.174, 459-486.

CORRETTE PLOEM, M., ESSINK-BOT, M-L., STRONKS, K. (2013) "Proposed EU data protection regulation is a threat to medical research." *Bmj Clinical Research* 346, doi: 10.1136/bmj.f3534.

CROWE, S., *et al.* (2011) "The case study approach." *BMC medical research methodology* 11.1, 1-9, doi: 10.1186/1471-2288-11-100.

CUIJPERS, C., KOOPS, B. (2008) "How fragmentation in European law undermines consumer protection: the case of location-based services." *European Law Review* 33, 880-897.

DASKAL, J. (2018) "Microsoft Ireland, the CLOUD Act, and International Lawmaking 2.0." *Stanford Law Review Online* 71, 9-16.

DAVID, P., A. (2008) "The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution." *Capitalism and Society* 3.2, 1-106, doi: 10.2202/1932-0213.1040.

DE GREGORIO, G. (2020) "The Rise of Digital Constitutionalism in the European Union." *International Journal of Constitutional Law*, 1-28.

DE HERT, P., PAPAKONSTANTINOU, V. (2012) "The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals." *Computer law & security review* 28.2, 130-142, doi: 10.1016/j.clsr.2012.01.011.

DE JOUVENAL, B. (1961) "The Republic of Science", *The Logic of Personal Knowledge.* (London: Routledge and Kegan Paul).

DE MARTIN, J. C. (2020) Opening speech of the first edition of "Biennale Tecnologia",https://www.youtube.com/watch?v=KmZVKhmiGyw.

DE MARTIN, J. C. (2017) *Università futura: tra democrazia e bit*. (Torino: Codice Edizioni).

DEN BROEDER, L., *et al.* (2018) "Citizen Science for public health." *Health promotion international* 33.3, 505-514, doi: 10.1093/heapro/daw086.

DESAI, J., WATSON, D., WANG, V., TADDEO, M., FLORIDI, L. (2022) "The epistemological foundations of data science: a critical analysis." SSRN, doi: 10.2139/ssrn.4008316.

DI DONATO, F. (2009) *La scienza e la rete: l'uso pubblico della ragione nell'età del web*. (Firenze: Firenze University Press).

DIBONA, C., OCKMAN, S. (1999) *Open sources: Voices from the open source revolution*. (Sebastopol, California: O'Reilly Media Inc.).

DIJKERS, M. P. (2019) "A beginner's guide to data stewardship and data sharing." *Spinal Cord* 57, 169-182, doi: 10.1038/s41393-018-0232-6.

DOLDIRINA, C., *et al.* (2018) "Legal approaches for open access to research data." *LawArXiv*.

DOS SANTOS, C., BASSI, E., *et al.* (2012) "Policy Recommendation on Privacy and Personal Data Protection as Regards Re-Use of Public Sector Information (PSI)", *Masaryk University Journal of Law and Technology MUJLT*, 6(3).

DRAXL C., SCHEFFLER, M. (2018) "NOMAD: The FAIR concept for big data-driven materials science." *Mrs Bulletin* 43.9, 676-682, doi: 10.1557/mrs.2018.208.

DUCATO, R. (2020) "Data protection, scientific research, and the role of information." *Computer Law & Security Review* 37, 1-16, doi: 10.1016/j.clsr.2020.105412.

DUCATO, R. (2013) "'Adiós sui generis': A Study of the legal Feasibility of the Sui Generis Right in the Context of Research Biobanks." *Revista de derecho y genoma humano*, 38, 125-146.

DURANTE, M. (2022) "Potere computazionale. Dalle informazioni ai dati." in DURANTE, M., PAGALLO, U. (eds.), *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società*. (Milano-Udine: Mimesis), 59-80.

DURANTE, M. (2021) *Computational Power: The Impact of ICT on Law, Society and Knowledge*. (Abingdon: Routledge).

DURANTE, M. (2019) "Commento all'art. 13 GDPR. Informazioni da fornire qualora i dati personali siano raccolti presso l'interessato." in GABRIELLI, E., (ed.), *Commentario del Codice Civile Utet. Modulo*

delle Persone, Vol. II. (Milano: Utet Giurdica), 218-234.
DURANTE, M. (2017) *Ethics, Law and the Politics of Information: A Guide to the Philosophy of Luciano Floridi*. (Dordrecht: Springer), doi: 10.1007/978-94-024-1150-8.
DURANTE, M. (2017) "An Informational Approach to Politics." in DURANTE, M. (eds.) *Ethics, Law and the Politics of Information. The International Library of Ethics, Law and Technology*, vol. 18 (Dordrecht, Springer), doi: 10.1007/978-94-024-1150-8_9.
DURANTE, M. (2015) "Sicurezza e fiducia nell'età della tecnologia." *Filosofia politica* 29.3, 439-458.
DURANTE, M. (2015) "The democratic governance of information societies. A critique to the theory of stakeholders." *Philosophy & Technology* 28.1, 11-32, doi: 10.1007/s13347-014-0162-y.
DURANTE, M. (2013) "How to cross boundaries in the information society: vulnerability, responsiveness, and accountability.", *Acm Sigcas Computers and Society* 43.1, 9-21.
DURANTE, M. (2013) "Dealing with legal conflicts in the information society. An informational understanding of balancing competing interests." *Philosophy & Technology* 26.4, 437-457, doi: 10.1007/s13347-013-0105-z.
DURANTE, M. (2012) "E-democracy as the Frame of Networked Public Discourse Information, Consensus and Complexity." *25th IVR World Congress: Law, Science and Technology* 20, 1-28.
DURANTE, M. (2010) "Re-designing the Role of Law in the Information Society." *European Journal of Legal Studies* 2.3, 1-18.
DURANTE, M. (2007) *Il futuro del Web: etica, diritto, decentramento. Dalla sussidiarietà digitale all'economia dell'informazione in rete*. (Torino: Giappichelli Editore).
DURANTE, M., SENOR, M. (2019) "Report on the harmonization of Italian Law with the enforcement of the GDPR.", in MCCULLAGH, K., TAMBOU, O., BOURTON, S. (eds.) *National Adaptations of the GDPR*, https://wp.me/p6OBGR-3dP.
EBERS, M., *et al*. (2021) "The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)." in PAGALLO, U., DURANTE, M. (eds.) *Special Issue. The Impact of Artificial Intelligence on Law, J* 4.4, 589-603, doi: 10.3390/j4040043.
ECKHARDT, J., *et al*., "Ecosystems of Co-Creation" *Frontiers in Sociology* 30 (2021), 1-11, doi: 10.3389/fsoc.2021.642289.
ERB, B., *et al*. (2021) "Emerging Privacy Issues in Times of Open Science." *PsyArXiv*, doi: 10.31234/osf.io/u236e.

ERLICH, Y., *et al.* (2018) "Identity inference of genomic data using long-range familial searches." *Science* 362.6415, 690-694, doi: 10.1126/science.aau4832.
ERWAN, L., SHINN, T. (2006) "L'autonomie scientifique face à la mercantilisation. Formes d'engagement entrepreneurial des chercheurs en France." *Actes de la recherche en sciences sociales*, 164,4, 23-50, doi: 10.3917/arss.164.0023.
FABRE, R. (2016) *New challenges for knowledge: Digital dynamics to access and sharing*. (Hoboken, NJ: John Wiley & Sons).
FANELLI, D. (2018) "Opinion: Is science really facing a reproducibility crisis, and do we need it to?" *Proceedings of the National Academy of Sciences* 115.11, 2628-2631, doi: 10.1073/pnas.1708272114.
FEESS, E., SCHEUFEN, M. (2016) "Academic copyright in the publishing game: A contest perspective." *European journal of law and economics* 42.2, 263-294.
FINCK, M., PALLAS, F. (2020) "They who must not be identified—distinguishing personal from non-personal data under the GDPR." *International Data Privacy Law*, 11-36, doi: 10.1093/idpl/ipz026.
FLORIDI, L. (2020) *Il verde e il blu. Idee ingenue per migliorare la politica.* (Milano: Cortina Editore).
FLORIDI, L. (2020) "The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU." *Philosophy & Technology* 33.3, 369-378, doi: 10.1007/s13347-020-00423-6.
FLORIDI, L. (2020) "What the near future of artificial intelligence could be." *The 2019 Yearbook of the Digital Ethics Lab* (Cham: Springer), 1-15, doi: 10.1007/s13347-019-00345-y.
FLORIDI, L. (2017) "Infraethics–on the Conditions of Possibility of Morality." *Philosophy & Technology* 30.4, doi: 10.1007/s13347-017-0291-1.
FLORIDI, L. (2017) "The logic of design as a conceptual logic of information." *Minds and Machines* 27.3, 495-519, doi: 10.1007/s11023-017-9438-1.
FLORIDI, L. (2015) *The onlife manifesto: Being human in a hyperconnected era.* (Cham: Springer Nature).
FLORIDI, L. (2014) *The fourth revolution: How the infosphere is reshaping human reality.* (Oxford: OUP) EPub.
FLORIDI, L. (2012) "Big data and their epistemological challenge." *Philosophy & Technology* 25.4 6, 435-437, doi: 10.1007/s13347-012-0093-4.
FLORIDI, L. (2011) *The philosophy of information.* (Oxford: OUP).
FLORIDI, L. (2011) *Information: A very short introduction.* (Oxford: OUP).

FLORIDI, L. (2010) *Ethics after the information revolution*. (New York: Cambridge University Press).

FLORIDI, L. (2010) "Levels of abstraction and the Turing test." *Kybernetes*, 39.3, 423-440, doi: 10.1108/03684921011036150.

FLORIDI, L. (2009) "The Information Society and Its Philosophy: Introduction to the Special Issue on The Philosophy of Information, Its Nature, and Future Developments." *The Information Society* 25, 153–158. doi: 10.1080/01972240902848583.

FLORIDI, L. (2008) "Artificial intelligence's new frontier: Artificial companions and the fourth revolution." *Metaphilosophy* 39.4-5.

FLORIDI, L. (2008) "The method of levels of abstraction." *Minds and machines* 18.3, 303-329, doi: 10.1007/s11023-008-9113-7.

FRANCESCONI, E., PERUGINELLI, G. (2011) "An open access policy for legal informatics dissemination and sharing." *An Open Access Policy for Legal Informatics Dissemination and Sharing* (Berlin-Heidelberg: Springer), 162-170, doi: 10.1007%252F978-3-642-35731-2.

FYFE, A. (2020) "The production, circulation, consumption and ownership of scientific knowledge: historical perspectives." *CREATe Working Paper* 4, 1-21. doi:10.5281/zenodo.3859492.

GAGLIARDI, B. (2018) *La tutela amministrativa della libertà accademica*. (Milano: CEDAM).

GAJBE, S. B., GOPALJI, A. T., SINGH, R. K. (2021) "Evaluation and analysis of Data Management Plan tools: A parametric approach." *Information Processing & Management*, 58.3, 1-17, doi: 10.106/j.ipm.2020.102480.

GALLINO, L. (2014) *Tecnologia e democrazia. Conoscenze tecniche e scientifiche come beni pubblici*. (Torino: Einaudi) EPub.

GARAVELLI, S., MÄRKÄLÄ, A., LIINAMAA, I. (2021) "EOSC National Structures: an overview of the national EOSC coordination and engagement mechanisms in Europe", *EOSC Secretariat*, doi: 10.5281/zenodo.5602949.

GANZINGER, M., GLAAB, E., *et al.* (2021) "Biomedical and clinical research data management" *SystemsMedicine: Integrative, Qualitative and Computational Approaches*, 3, 532-543, doi: 10.1016/B978-0-12-801238-3.11621-6.

GIANNOPOULOU, A. (2019) "Access and Reuse of Machine-Generated Data for Scientific Research." *Erasmus Law Review* 2, 155-165, doi; 10.5553/ELR.000136.

GIGLIA, E. (2017) "Open Access e Open Science: per una scienza più efficace." *Journal of Biomedical Practitioners* 1.1, 1-24.

GILLOT, M., *et al.* (2021) “Towards Sustainable HPC”, 1-11, http://people.irisa.fr/Francois.Bodin/wp-content/uploads/2021/03/20210330Sustainable_HPC.pdf.

GLORIOSO, A. (2012) “Un nuovo concetto di auto-determinazione informazionale come bussola concettuale per navigare il nuovo mondo digitale.”, in PAGALLO, U., DURANTE, M. (eds.), *Manuale di informatica giuridica e diritto delle nuove tecnologie*. (Milano: Utet Giuridica), 383-394.

GOSS, A. K. (2007) “Codifying a commons: copyright, copyleft, and the Creative Commons project.” *Chicago-Kent Law Review*, 82.2, 963-996.

GOZLAN, C. (2015) “L’autonomie de la recherche scientifique en débats: évaluer l’”impact’ social de la science.” *Sociologie du travail* 57,2, 151–174, doi: 10.1016/j.soctra.2015.02.001.

GRABER-SOUDRY, O. *et al.* (2021) “Legal Interoperability and the FAIR Data Principles.” (1.0) *Zenodo*, 1-77, doi:10.5281/zenodo.4471312.

GREENE, B. (1999) *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. (New York: W.W. Norton).

GUARDA, P. (2021) *Il regime giuridico dei dati della ricerca scientifica*. (Trento: Collana della facoltà di Giurisprudenza dell’Università di Trento), https://iris.unitn.it/handle/11572/315657.

GUARDA, P. (2016) “Research paper on Case Study (iii): Barriers that citizens face regarding their intellectual property rights”, *BEUCitizen, University of Trento*, 1-213.

GUARDA, P. (2013) “Creation of Software Within the Academic Context: Knowledge Transfer, Intellectual Property Rights and Licences.” *IIC-International Review of Intellectual Property and Competition Law* 44.5, 494-523, doi: 10.1007/s40319-013-0078-6.

GUARDA, P. (2011) “Open Access to Legal Scholarship and Open Archives: Towards a Better Future?.” *From Information to Knowledge*. IOS Press, 143-151.

GUARDA, P. (2008) “Privacy e fruizione della conoscenza scientifica”, *Pubblicazioni scientifiche, diritti d'autore ed open access – Atti del Convegno tenuto presso la Facoltà di Giurisprudenza di Trento*, 1-13.

GUÉDON, J-C. (2001) “In Oldenburg’s Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing.” In *Creating the Digital Future : Association of Research Libraries 138th Annual Meeting* (Toronto, Canada), http://eprints.rclis.org/6375/1/ARL_Proceedings_138_In_Oldenburg%27s_Long_Shadow%2C_by_Guedon.htm.

GUIBAULT, L., MARGONI, T. (2015) "Legal aspects of open access to publicly funded research." *Enquiries into Intellectual Property's Economic Impact*, 373-414.

GUIBAULT, L., WIEBE, A. (2013) *Safe to be open: Study on the protection of research data and recommendations for access and usage*. (Göttingen: Universitätsverlag Göttingen).

GUZIK, T. J., AHLUWALIA, A. (2019) "Plan S: in Service or Disservice to Society? The controversial plan for scientific research publications to be published in compliant Open Access Journals or on compliant Open Access Platforms, is discussed.", *European Heart Journal*, 40.12, 949-952, doi: 10.1093/eurheartj/ehz065.

HABERMAS, J. (2008) *Europe. The faltering project*. (Cambridge: Polity Press).

HABERMAS, J. (1996) *Between facts and norms. Contributions to a discourse theory of law and democracy*. (Cambridge, Massachusetts: The MIT Press).

HABERMAS, J. (1972) *Knowledge and human interest*. (Boston: Beacon Press).

HACKER, P. (2021) "A legal framework for AI training data—from first principles to the Artificial Intelligence Act." *Law, Innovation and Technology* 13.2, 257-301, doi: 10.1080/17579961.2021.1977219.

HALLINAN, D. (2020) "Broad consent under the GDPR: an optimistic perspective on a bright future." *Life Sciences, Society and Policy* 16.1, 1-18, doi: 10.1186/s40504-019-0096-3.

HANSSON, S. O. (2021) "Science and Pseudo-Science", *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/pseudo-science/.

HARARI, Y. N. (2016) *Homo Deus: A brief history of tomorrow*. (New York: Harper Collins) EPub.

HARRIS, K. (2020) "Kamala Harris' full victory speech", CNN, Youtube, https://www.youtube.com/watch?v=ExPm_hJQYpQ.

HEPBURN, B., ANDERSEN, H. (2021) "Scientific method", *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/scientific-method/.

HEY, A. JG., TANSLEY, S., TOLLE, K. M. (2009) *The fourth paradigm: data-intensive scientific discovery*. (Redmond, WA: Microsoft research).

HICKS, D., *et al*. (2015) "Bibliometrics: the Leiden Manifesto for research metrics." *Nature News* 520.7548, 429-431, doi: 10.1038/520429a.

HO, C-H. (2017) "Challenges of the EU' general data protection regulation' for biobanking and scientific research." *Journal of Law, Information and Science* 25.1, 84-103.

HOTTENROTT, H., LAWSON, C. (2014) "Research grants, sources of ideas and the effects on academic research." *Economics of Innovation and New Technology* 23.2, 109-133, doi: 10.1080/10438599.2013.814425.

HUBERMAN, B. A. (2012) "Sociology of science: Big data deserve a big audience." *Nature*, 482, 308.

HUI, J. S., GERBER, E. (2015) "Crowdfunding science: sharing research with an extended audience." *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, New York*, 31-43, doi: 10.1145/2675133.2675188.

HUTSON, M. (2018) "Artificial intelligence faces reproducibility crisis." *Science* 359.6377, 725-726, doi: 10.1126/science.359.6377.725.

IENCA, M., *et al.* (2019) "How the general data protection regulation changes the rules for scientific research." *European Parliamentary Research Service (EPRS), Scientific Foresight Unit (STOA)*, doi: 10.2861/17421.

IGARASHI, Y., *et al.* (2016) "Three levels of data-driven science." *Journal of Physics: Conference Series*, 699.1, 1-13, doi: 10.1088/1742-6596/699/1/012001.

IRWIN, A. (1995) *Citizen Science: A Study of People, Expertise and Sustainable Development*. (Abingdon: Routledge).

JONES, S., ABRAMATIC, J-F. (2019) "European Open Science Cloud (EOSC) Strategic Implementation Plan." *Publications Office of the European Union*, doi: 10.2777/202370.

JOSEPH, E. C., *et al.* (2010) "A Strategic Agenda for European Leadership in Supercomputing: HPC 2020," *IDC Final Report of the HPC Study for the DG Information on Society of the European Commission*.

KALBE, G. (2019) "The European Approach to the Exascale Challenge." *Computing in Science & Engineering*, 21.1, 42-47, doi: 10.1109/MCSE.2018.2884139.

KALLINIKOS, J. (2009) "The making of ephemeria: on the shortening life spans of information." *International Journal of Interdisciplinary Social Sciences* 4.3, 227-236.

KAMARA, I., DE HERT, P. (2018) "Understanding the balancing act behind the legitimate interest of the controller ground: A pragmatic approach." In SELINGER,E., POLONETSKY, J., TENE O. (eds.), *The Cambridge Handbook of Consumer Privacy*. (Cambridge: Cambridge University Press), 321-352, doi:

10.1017/9781316831960.019.
KAMOCKI, P., MAPELLI, V., CHOUKRI, K. (2018) "Data management plan (DMP) for language data under the new general data protection regulation (GDPR)." *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, 135-139.
KITCHIN, R., MCARDLE, G. (2016) "What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets", *Big Data & Society*, 3.1, 1-10, doi: 10.1177/2053951716631130.
KOŠČÍK, M., MYŠKA, M. (2018) "Data protection and codes of conduct in collaborative research." *International Review of Law, Computers & Technology* 32.1, 141-154, doi: 10.1080/13600869.2018.1423888.
KOŠČÍK, M., MYŠKA, M. (2017) "Database authorship and ownership of sui generis database rights in data-driven research." *International Review of Law, Computers & Technology* 31.1, 43-67, doi: 10.1080/13600869.2017.1275119.
KOTSEV, A., *et al.* (2020) "From spatial data infrastructures to data spaces—A technological perspective on the evolution of European SDIs." *ISPRS International Journal of Geo-Information,* 9.3, doi: 10.3390/ijgi9030176.
KULLENBERG, C., KASPEROWSKI, D. (2016) "What is citizen science?–A scientometric meta-analysis." *PloS one* 11.1, e0147152, doi: 10.1371/journal.pone.0147152.
KUHN, T. (1970) *The structure of scientific revolutions*. (Chicago: UPC).
KUNER, C. (2021) "Territorial Scope and Data Transfer Rules in the GDPR: Realising the EU's Ambition of Borderless Data Protection." *University of Cambridge Faculty of Law Research Paper,* 20, 1-35.
KUNER, C., DOCKSEY, C., BYGRAVE, L. (2020) *The EU General Data Protection Regulation: A Commentary* (Oxford: OUP).
LABASTIDA, I., MARGONI, T. (2020) "Licensing FAIR data for reuse." *Data Intelligence* 2.1-2, 199-207, doi: 10.1162/dint_a_00042.
LAKATOS, I. (2015) *Proofs and refutations: The logic of mathematical discovery*. (Cambridge: Cambridge University Press).
LEDYAYEV, R., RICHTER, H. (2014) "High performance computing in a cloud using openstack." *Cloud Computing*, 108-113.
LEE-MAKIYAMA H., GOPALAKRISHNAN, B. N. (2020) "Economic cost of ex ante Regulations", *ECPIE*, https://ecipe.org/publications/ex-ante/.
LEONELLI, S. (2020) "Scientific Research and Big Data", *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/science-big-data/.

LEONELLI, S. (2018) *La ricerca scientifica nell'era dei big data. Cinque modi in cui i Big Data danneggiano la scienza, e come salvarla.* (Milano: Meltemi Editore).

LEONELLI, S. (2017) "Global Data Quality Assessment and the Situated Nature of 'Best' Research Practices in Biology." *Data Science Journal*, 16.32, 1-11, doi: 10.5334/dsj-2017-032.

LEONELLI, S. (2014) "What difference does quantity make? On the epistemology of Big Data in biology." *Big data & society* 1.1, 1-11, doi: 10.1177/2053951714534395.

LESLIE, D. (2020) "Tackling COVID-19 through responsible AI innovation: Five steps in the right direction." *Harvard Data Science Review*, 1-58, doi: 10.1162/99608f92.4bb9d7a7.

LESSIG, L. (2011) "The architecture of access to scientific knowledge: just how badly we have messed this up", *CERN Colloquium and Library Science Talk,* 18 April 2011, video, http://cdsweb.cern.ch/record/1345337.

LESSIG, L. (2004) "The Creative Commons." *Montana Law Review* 65.1.

LESSIG, L. (1999) *Code and other laws of cyberspace.* (New York: Basic Books).

LEV-ARETZ, Y. (2018) "Data philanthropy." *Hastings Law Journal* 70, 1491-1546.

LICHTENTHALER, U. (2011) "Open innovation: Past research, current debates, and future directions." *Academy of management perspectives* 25.1, 75-93, doi: 10.5465/amp.25.1.75.

LONGO, E. (2019) "The European Citizens' initiative: too much democracy for EU polity?" *German Law Journal* 20, 181-200, doi: 10.1017/glj.2019.12.

MAGNUSSON SJÖBERG, C. (2017) "Scientific research and academic e-learning in light of the EU's legal framework for data protection." in CORRALE, M., FENWICK, M., FORGÒ, N. (eds.), *New Technology, Big Data and the Law.* (Singapore, Springer), 43-63, doi: 10.1007/978-981-10-5038-1_3.

MANIS, M. L. (2017) "The processing of personal data in the context of scientific research. The new regime under the EU-GDPR." *BioLaw Journal* 11.3, 325-354, doi:10.15168/2284-4503-259.

MARGONI, T. (2016) "The harmonisation of EU copyright law: the originality standard." *Global governance of intellectual property in the 21st century* (Cham: Springer), 85-105, doi:10.1007/978-3-319-31177-7_6.

MARTONE, M. (2014) *Data Citation Synthesis Group of FORCE11. Joint Declaration of Data Citation Principles.* (San Diego, CA: FORCE11), doi: 10.25490/a97f-egyk.

MAYER-SCHÖNBERGER, V., CUKIER, K. (2013) *Big data: A revolution that will transform how we live, work, and think.* (Boston, Massachusetts: Houghton Mifflin Harcourt).

MAZZOCCHI, F. (2015) "Could Big Data be the end of theory in science? A few remarks on the epistemology of data-driven science." *EMBO reports,* 16.10, 1250-1255, doi: 10.15252/embr.201541001.

MEDE, N. G, *et al*. (2021) "The 'replication crisis' in the public eye: Germans' awareness and perceptions of the (ir) reproducibility of scientific research." *Public Understanding of Science* 30.1, 91-102, doi: 10.1177/0963662520954370.

MELL, P., GRANCE, T. (2011) "The NIST definition of cloud computing." *Special Publication* 800-145, 1-3, https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.

MERTON, R. K. (1973) *The sociology of science. Theoretical and empirical investigations*. (Chicago: UCP).

MERTON, R. K. (1957) "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22.6, 635-659.

MERTON, R. K. (1938) "Science and the Social Order." *Philosophy of Science* 5.3, 321–337.

MÉSZÁROS, J., HO, C-H. (2021) "AI research and data protection: Can the same rules apply for commercial and academic research under the GDPR?." *Computer Law & Security Review* 41, 105532, 1-10, doi: 10.1016/j.clsr.2021.105532.

MÉSZÁROS, J., HO, C-H. (2018) "Big data and scientific research: the secondary use of personal data under the research exemption in the GDPR." *Hungarian Journal of Legal Studies* 59.4, doi: 10.1556/2052.2018.59.4.5.

MILLS, P. (2016) "Ethical reuse of data from health care: Data, persons and interests." in MITTELSTADT, B. D., FLORIDI, F. (eds.) *The ethics of biomedical big data.* (Cham: Springer), 429-444, doi: 10.1007/978-3-319-33525-4_18. p. 430.

MINARI, J., YOSHIZAWA,G., SHINOMIYA, N. (2020) "COVID-19 and the boundaries of open science and innovation: Lessons of traceability from genomic data sharing and biosecurity." *EMBO reports* 21.11, doi: 10.15252/embr.202051773.

MITTELSTADT, B. D., FLORIDI, L. (2016) "The ethics of big data: current and foreseeable issues in biomedical contexts." *Science and Engineering Ethics*, 445-480, doi: 10.1007/s11948-015-9652-2.

MIYAKAWA, T. (2020) "No raw data, no science: another possible source of the reproducibility crisis." *Molecular Brain*, 13, 13-24, doi: 10.1186/s13041-020-0552-2.

MONDSCHEIN, C. F., MONDA,C. (2019) "The EU's General Data Protection Regulation (GDPR) in a research context." in KUBBEN, P., DUMONTIER, M., DEKKER, A. (eds.), *Fundamentals of clinical data science.* (Cham: Springer Nature), 55-71, doi: 10.1007/978-3-319-99713-1.

MONS, B. (2018) *Data stewardship for open science: Implementing FAIR principles*. (Boca Raton, Florida: CRC Press).

MONS, B., *et al.* (2017) "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud." *Information Services & Use* 37.1, doi: 10.3233/ISU-170824.

MOOR, J. H. (1985) "What is computer ethics?." *Metaphilosophy* 16.4, 266-275.

MORANDO, F. (2013) "Legal Interoperability: Making Open (Government) Data Compatible with Businesses and Communities.", *Italian Journal of Library and Information Science*, 4.1, 441-452, doi:10.4403/jlis.it-5461.

MORLEY, J., FLORIDI, F., KINSEY, L., ELHALAL, A. (2020) "From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices." *Science and engineering ethics* 26.4, 2141-2168, doi: 10.1007/s11948-019-00165-5.

MORO, L., *et al.* (2021) "Quantum compiling by deep reinforcement learning", *Communication Physics*, 4.178, 1-8, doi: 10.1038/s42005-021-00684-3.

MOUTON, S. (2005) *Les fondementes consitututionnels de la liberté de la recherché.* (Toulouse: Presses de l'Université de Toulouse), 93-111.

MUNAFÒ, M.R., NOSEK, B. A., BISHOP, D., *et al.* (2017) "A manifesto for reproducible science." *Nature Human Behaviour* 1, 0021, 1-9, doi: 10.1038/s41562-016-0021.

MURILLO, A. P. (2020) "An examination of scientific data repositories, data reusability, and the incorporation of FAIR." *Proceedings of the Association for Information Science and Technology* 57.1, 1-3, doi: 10.1002/pra2.386.

NASCIMBENI, F. (2020) *Open Education.* (Milano: Franco Angeli Editore).

NATURE EDITORIAL (2020) "Calling all coronavirus researchers: keep sharing, stay open" *Nature*, 576, 7, doi: 10.1038/d41586-020-00307-x.

NATURE NEUROSCIENCE EDITORIAL (2022) "Nature Neuroscience offers

open access publishing." *Nature Neuroscience* 25, 1, doi: 10.1038/s41593-021-00995-2.
NELSON, R. R. (2004) "The market economy, and the scientific commons." *Research policy* 33.3, 455-471, doi: 10.1016/j.respol.2003.09.008.
NETTO M. A., *et. al.* (2018) "HPC Cloud for Scientific and Business Applications: Taxonomy, Vision, and Research Challenges." *ACM Computing Surveys*, 51.1, 8.2, doi: 10.1145/3150224.
NICHOLLS, S. G., LANGAN, S. M., BENCHIMOL, E. I. (2016) "Reporting and transparency in big data: the nexus of ethics and methodology." in MITTELSTADT, B. D., FLORIDI, L. (eds.) *The ethics of biomedical big data.* (Cham: Springer), 339-365, doi: 10.1007/978-3-319-33525-4_15.
NIELSEN, M. (2012) *Reinventing discovery. The new era of networked science.* (Princeton: Princeton University Press).
NIINILUOTO, I. (2019) "Scientific Progress", in ZALTA, E. N. (ed.) *The Stanford Encyclopedia of Philosophy,* https://plato.stanford.edu/archives/win2019/entries/scientific-progress/.
NOSEK, B. A., *et al.* (2022) "Replicability, robustness, and reproducibility in psychological science." *Annual Review of Psychology*, 73, 1-94, doi: 10.1146/annurev-psych-020821-114157.
OROFINO, M. (2018) "Diritto alla protezione dei dati personali e sicurezza: osservazioni critiche su una presunta contrapposizione", *MediaLaws-Rivista di diritto dei media* 2, 1-23.
ORFINO, M. (2018) "Minori e diritto alla protezione dei dati personali." in ORFINO, M., PIZZETTI, F. (eds.), *Privacy, Minori e Cyberbullismo.* (Torino: Giappichelli Editore), 1-30.
OROFINO, M. (2017) "Trattamento dei dati personali e libertà di espressione e informazione." in CALIFANO, L., COLAPIETRO, C., *Innovazione tecnologica e valore della persona.* (Napoli: Editoriale Scientifica), 509-538.
OROFINO, M. (2015) "The Web 2.0 and its impact on relations between citizens and political representatives." *Democratic Governance and Active Citizenship in the European Union, Saarbrucken: Lambert Academic Publishing*, 81-100.
OROFINO, M. (2008) *Profili costituzionali delle comunicazioni elettroniche nell'ordinamento multilivello.* (Milano: Giuffrè Editore).
PADUA, D. (2011), *Encyclopedia of Parallel Computing.* (Heidelberg, Springer), doi: 10.1007/978-0-09766-4.
PAGALLO, G., HAUSSLER, D. (1990) "Boolean feature discovery in empirical learning." *Machine learning* 5.1, 71-99.
PAGALLO, U. (2022) "La grande trasformazione" in DURANTE, M., PAGALLO,

U. (eds.), *La politica dei dati. Il governo delle nuove tecnologie tra diritto, economia e società*. (Milano-Udine: Mimesis), 123-140.

PAGALLO, U. (2022) *Il dovere alla salute. Sul rischio di sottoutilizzo dell'intelligenza artificiale in ambito sanitario*. (Milano-Udine: Mimesis).

PAGALLO, U. (2021) "On the principle of privacy by design and its limits: Technology, ethics and the rule of law." in CHIODO, S., SCHIAFFONATI, V., (eds.), *Italian Philosophy of Technology*. (Cham: Springer), 111-127, doi: 10.1007/978-3-030-54522-2_8.

PAGALLO, U. (2020) "Sovereigns, Viruses, and the Law." *Law in Context. A Socio-legal Journal* 37.1, 11-23.

PAGALLO, U. (2018) "Algo-rhythms and the beat of the legal drum." *Philosophy & Technology* 31.4, 507-524, doi: 10.1007/s13347-017-0277-z.

PAGALLO, U. (2017) "From automation to autonomous systems: A legal phenomenology with problems of accountability." *26th International Joint Conference on Artificial Intelligence, IJCAI 2017*, 17-23.

PAGALLO, U. (2017) "The legal challenges of big data: putting secondary rules first in the field of EU data protection." *European Data Protection Law Review* 3, 36-46, doi: 10.21552/edpl/2017/1/7.

PAGALLO, U. (2017) "LegalAIze: Tackling the Normative Challenges of Artificial Intelligence and Robotics Through the Secondary Rules of Law." in CORRALES, M., *et al.* (eds), *New Technology, Big Data and the Law. Perspectives in Law, Business and Innovation*. (Singapore: Springer), 281-300, doi: 10.1007/978-981-10-5038-1_11.

PAGALLO, U. (2017) "Good Onlife Governance: On Law, Spontaneous Orders, and Design." in FLORIDI, L. (ed.), *The Onlife Manifesto: Being Human in a Hyperconnected Era*. (Dordrecht: Springer), 161-177, doi: 10.1007/978-3-319-04093-6_18.

PAGALLO, U. (2017) "The broken promises of democracy in the information era." in PRINS, C., *et al.* (eds), *Digital democracy in a globalized world*. (Cheltenham, UK: Edward Elgar Publishing), 77-99, doi: 10.4337/9781785363962.

PAGALLO, U. (2014) *Il diritto nell'età dell'informazione: il riposizionamento tecnologico degli ordinamenti giuridici tra complessità sociale, lotta per il potere e tutela dei diritti*. (Torino: Giappichelli Editore).

PAGALLO, U. (2008) *La tutela della privacy negli Stati Uniti d'America e in Europa: modelli giuridici a confronto*. (Milano: Giuffrè Editore).

PAGALLO, U., BASSI, E. (2013) "Open Data Protection: Challenges, Perspectives, and Tools for the Reuse of PSI", in HILDEBRANDT, M., *et al.* (eds), *Digital Enlightenment Yearbook 2013.* (Amsterdam: IOS Press), doi:10.3233/978-1-61499-295-0-179.

PAGALLO, U., BASSI, E. (2011) "The future of EU working parties' "the future of privacy" and the principle of privacy by design." *An Information Law for the 21st Century*, 286-309.

PAGALLO, U., DURANTE, M. (2016) "The philosophy of law in an information society." in FLORIDI, L., *The Routledge Handbook of Philosophy of Information*. (Abingdon: Routledge), 396-407.

PAGALLO, U., DURANTE, M. (2012) *Manuale di informatica giuridica e diritto delle nuove tecnologie*. (Milano: Utet Giuridica).

PAGALLO, U., DURANTE, M., MONTELEONE, S. (2017) "What Is New with the Internet of Things in Privacy and Data Protection? Four Legal Challenges on Sharing and Control in IoT." in LEENES, R., *et. al.* (eds), *Data Protection and Privacy: (In)visibilities and Infrastructures. Law, Governance and Technology Series*. 36, 59-78, doi: 10.1007/978-3-319-50796-5_3.

PALMIRANI, M. (2020) "Big Data e Conoscenza." *Rivista di filosofia del diritto*, 1, 73-91, doi: 10.4477/97021.

PARLAND-VON ESSEN, J., *et al.* (2018) "Supporting FAIR data: categorization of research data as a tool in data management." *Informaatiotutkimus* 37.4, 128-144, doi: 10.23978/inf.77419.

PASERI, L. (2021) "COVID-19 Pandemic and GDPR: When Scientific Research becomes a Component of Public Deliberation", in HALLINAN, D. LEENES, R., DE HERT, P. (eds.), *Data Protection and Privacy: Enforcing Rights in a Changing World.* (London: Hart Publishing), 157-185.

PASERI, L., VARRETTE, S., BOUVRY, P. (2021) "Protection of Personal Data in High Performance Computing Platform for Scientific Research Purposes." *Annual Privacy Forum* (Cham: Springer), 123-142, doi: 10.1007/978-3-030-76663-4_7.

PASERI, L. (2019) "Crowdfunding of Science and Open Data: Opportunities, Challenges, and Policies." in KŐ, A., *et al.* (eds.), *Electronic Government and the Information Systems Perspective: 8th International Conference.* 11709 (Cham: Springer), 3-15, doi: 10.1007/978-3-030-27523-5_1.

PASQUALE, V., LAZZERI, E., GIGLIA, E. (2021) "Data steward per i dati FAIR." *Conferenza GARR 2021 "Sostenibile/Digitale"*, 77-80, doi: 10.26314/GARR-Conf21-proceedings-16.

PELOQUIN, D., DIMAIO, M., BIERER, B., BARNES, M. (2020) "Disruptive and avoidable: GDPR challenges to secondary research uses of data."

European Journal of Human Genetics, 28, 701-702, doi: 0.1038/s41431-020-0596-x.
PENZIAS, A. (1989) *Ideas and information: Managing in a high-tech world.* (New York: WW Norton & Co., Inc.).
PERRONE, G., PERDUCA, M. (2021) *Così San Tuttə. Diritto alla scienza: istruzioni per l'uso*. (Roma: Fandango libri).
PERUGINELLI, G., FARO, S. (2018) "Research quality evaluation: The case of legal studies." *The evaluation of research in social sciences and humanities.* (Cham: Springer), 103-129, doi: 10.1007/978-3-319-68554-0_5.
PERUGINELLI, G. (2009) *Multilinguismo e sistemi di accesso all'informazione giuridica* (Milano: Giuffrè Editore).
PETERS, M. A. (2010) "On the philosophy of open science." *Review of Contemporary Philosophy* 9, 105-142.
PETRIDIS, P., *et. al*. (2017) "The role of science in sustainability transitions: citizen science, transformative research, and experiences from Samothraki Island Greece." *ISJ* 12.1, (2017): 115–134, doi: 10.24043/isj.8.
PFEIFER, H., OPITZ, C., GEIS, A. (2020) "Deliberating Foreign Policy: Perceptions and Effects of Citizen Participation in Germany." *German Politics* 0, 1–18. doi: 10.1080/09644008.2020.1786058.
PHILLIPS, M., KNOPPERS, B.M. (2019) "Whose Commons? Data Protection as a Legal Limit of Open Science." *The Journal of Law, Medicine & Ethics* 47.1, 106-111.
PIEVATOLO, M. C. (2019) "Open science: human emancipation or bureaucratic serfdom?". *SCIRES-it*, 35-52, doi: 10.2423/i22394303v10Sp35.
PIEVATOLO, M. C. (2017) "La bilancia e la spada: scienza di stato e valutazione della ricerca." *Bollettino Telematico di Filosofia Politica*, https://btfp.sp.unipi.it/it/2017/05/libric/.
PIEVATOLO, M. C. (2012) "L'università e le sue crisi: una riflessione storica". *Bollettino telematico di filosofia politica.*
PIEVATOLO, M. C. (2012) "Open Access/Accesso Aperto." *Le parole dell'innovazione*, 1-5, https://archiviomarini.sp.unipi.it/437/.
PISTOR, K. (2019) *The code of capital. How the law creates wealth and inequality*. (Princeton: Princeton University Press).
POCAR, F. (2001) *Commentario Breve ai Trattati dell'Unione europea*, (Padova: Cedam).
POLANYI, M. (1983) *The tacit dimension*. (Gloucester, Mass: Peter Smith ed.).
POLANYI, M. (1958) *Personal knowledge: towards a post-critical philosophy*. (London: Routledge & Kegan Paul).

POLANYI, M., ZIMAN, J., FULLER, S. (2000) “The republic of science: Its political and economic theory” *Minerva*, 1(1) (1962), 54-73, in *Minerva* 38, 1–32.

POLLICINO, O., BASSINI, M. (2017) “Bridge is down, data truck can’t get through... a critical view of the Schrems Judgment in the context of European constitutionalism.” *A Critical View of the Schrems Judgment in The Context of European Constitutionalism*, 1-28.

PORSDAM, H., PORSDAM MANN S. (2021) *The Right to Science. Then and Now*. (Cambridge: CUP).

PRICE, W. J, BASS, L. W. (1969) “Scientific Research and the Innovative Process: The dialogue between science and technology plays an important, but usually nonlinear, role in innovation.” *Science* 164.3881, doi: 10.1126/science.164.3881.802.

PRICTOR, M., *et al.* (2019) “Consent for data processing under the General Data Protection Regulation: Could ‘dynamic consent’ be a useful tool for researchers?.” *Journal of Data Protection & Privacy* 3.1, 93-112.

PULIDO, C. M., *et al.* (2020) “COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information.” *International Sociology* 35, 377–392. doi: 10.1177/0268580920914755.

QUESSADA, D. (2010) “De la Sousveillance.” *Multitudes*, 1.40, 54-59.

QUINN, P. (2021) “Research under the GDPR – a level playing field for public and private sector research?” *Life Sciences, Society and Policy* 17.4, 1-33, doi: 10.1186/s40504-021-00111-z.

RAGNI, C. (2020) *Scienza, diritto e giustizia internazionale*. (Milano: Giuffrè).

REED, C. (2012) *Making laws for cyberspace*. (Oxford: OUP).

RESTA, G. (2011) “The new frontiers of personality rights and the problem of commodification: European and comparative perspectives.” *Tulane European and Civil Law Forum*, 26, 33-66.

RHOEN, M. (2016) “Beyond consent: improving data protection through consumer protection law.” *Internet Policy Review*, 5.1, 1-15, doi: 10.14763/2016.1.404.

RICCIO, G. M., SCORZA, G., BELISARIO, E. (2018) *GDPR e normativa privacy* (Milano: Wolters Kluwer).

RICOLFI, M. (2002) “Intellectual Property Rights and Legal Order”. *Global Jurist* 2.1, 1-18, doi: 10.2202/1535-1661.1052.

RIDGWAY, S. L., CLARK, T., BAGHERI, B. (2021) *Scientific parallel computing*. (Princeton: Princeton University Press).

RIFKIN, J. (2000) *The age of Access: The new culture of hypercapitalism, where all of life is a paid-for experience.* (New York: Putnam Publishing).

RITSON, S., CAMILLERI, K. (2015) "Contested boundaries: The string theory debates and ideologies of science." *Perspectives on Science* 23.2, 192-227, doi:10.1162/POSC_a_00168.

ROBINSON, K. R. D., SIMONEC, A., MAZZONETTO, M. (2021) "RRI legacies: co-creation for responsible, equitable and fair innovation in Horizon Europe." *Journal of Responsible Innovation* 8.2, 209-216, doi: 10.1080/23299460.2020.1842633.

RODOTÀ, S. (2012) *Il diritto di avere diritti.* (Roma-Bari: Laterza Editore).

RODOTÀ, S. (2005) *Intervista su privacy e libertà*, CONTI, P. (eds.) (Roma-Bari: Laterza Editore).

RODOTÀ, S. (1995) *Tecnologie e diritti.* (Bologna: Il Mulino).

ROJSZCZAK, M. (2020) "CLOUD act agreements from an EU perspective." *Computer Law & Security Review* 38, 1-14, doi: 10.1016/j.clsr.2020.105442.

ROSS-HELLAUER, T. (2017) "What is open peer review? A systematic review." *F1000Research* 6, 1-37, doi: 10.12688/f1000research.11369.2.

ROSSI, P. (1999) *La nascita della scienza moderna in Europa.* (Roma-Bari: Laterza Editore).

RUTHERFORD, M. (2019) "The CLOUD Act: Creating executive branch monopoly over cross-border data access." *Berkeley Technology Law Review*, 34, 1186-1193, doi: 10.15779/Z387940V34.

SAMARA, K. (2020) "Selling Personal Data: The Legal Framework and Nature of Personal Data Selling Transactions Under GDPR." *Personal Data Protection and Legal Developments in the European Union, IGI Global*, 34-59, doi: 10.4018/978-1-5225-9489-5.ch003.

SCHERMER, B. W., CUSTERS, B., VAN DER HOF, S. (2014) "The crisis of consent: How stronger legal protection may lead to weaker consent in data protection." *Ethics and Information Technology* 16.2, 171-182, doi: 10.1007/s10676-014-9343-8.

SCHIERMEIER, Q. (2021) "How Europe's€ 100-billion science fund will shape 7 years of research." *Nature* 591.7848, 20-21, doi: 10.1038/d41586-021-00496-z.

SCHNEIDER, G. (2019) "Disentangling health data networks: a critical analysis of Articles 9(2) and 89 GDPR." *International Data Privacy Law*, 9.4, 253-271, doi: 10.1093/idpl/ipz015.

SCHWARTZ, P. M., SOLOVE, D., J. (2014) "Reconciling personal information in the United States and European Union." *California Law Review,* 102, 877-916.

SGANGA, C., *et al.* (2022) "Recommendations for legal and policy harmonization of open and FAIR science in the EU", *EOSC-Pillar Project*, 1-8, doi: 10.5281/zenodo.6327691.

SHAVER, L. (2010) "The right to science and culture." *Wisconsin Law Review*, 121-184.

SIEMS, M. M. (2009) "The taxonomy of interdisciplinary legal research: Finding the way out of the desert." *Journal of Commonwealth Law and Legal Education* 7, 5–17. doi: 10.1080/14760400903195090.

STAUNTON, C., SLOKENBERGA, S., MASCALZONI, D. (2019) "The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks." *European Journal of Human Genetics* 27, 1159-1167, doi: 10.1038/s41431-019-0386-5.

STALLA-BOURDILLON, S., KNIGHT, A. (2016) "Anonymous data v. personal data-false debate: an EU perspective on anonymization, pseudonymization and personal data." *Wisconsin International Law Journal*, 34, 284-322.

STEINER G. (2015), *The idea of Europe. An essay*. (New York, Abrams).

SUBER, P. (2012) *Open Access*. (Cambridge, Massachusetts: The MIT Press).

TADDEO, M. (2020) "The Ethical Governance of the Digital During and After the COVID-19 Pandemic." *Minds and Machines* 30, 171-176, doi: 10.1007/s11023-020-09528-5.

TADDEO, M. (2020) "The Civic Role of OSPs in Mature Information Societies.", SSRN, doi: 10.2139/ssrn.3584187.

TADDEO, M. (2020) "Virus e tecnologia: perché sono importanti scelte etiche (Virus and Technology: Why Ethical Decisions Matter).", *Aspen Institute Italia, Aspenia* 89, 1-6, https://ssrn.com/abstract=3632977.

TADDEO, M. (2017) "Trusting Digital Technologies Correctly." *Minds & Machines* 27, 565-568, doi: 10.1007/s11023-017-9450-5.

TADDEO, M. (2017) "Data philanthropy and individual rights." *Minds and Machines* 27.1, 1-5, doi: 10.1007/s11023-017-9429-2.

TADDEO, M. (2016) "Philosophy and computing in Information Societies." *Minds & Machines* 26, 203-204, doi: 10.1007/s11023-016-9400-7.

TADDEO, M. (2016) "Data philanthropy and the design of the infraethics for information societies." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083, 1-12, doi: 1.1098/rsta.2016.0113.

TADDEO, M. (2015) "The Struggle Between Liberties and Authorities in the Information Age.", *Science and Engineering Ethics volume* 21, 1125-1138., doi: 10.1007/s11948-014-9586-0.

TADDEO, M., FLORIDI, L. (2017) *The Moral Responsibilities of Online Service Providers*. (Cham: Springer), doi: 10.1007/978-3-319-47852-4_2.

TADDEO, M., FLORIDI, L. (2016) "The debate on the moral responsibility of online service providers." *Science and Engineering Ethics* 22, 1575-1603, doi: 10.1007/s11948-015-9734-1.

TADDEO, M., MILLER, K. (2020) "Ethics and Information Technologies: History and Themes of a Research Field.", in TADDEO, M., MILLER, K. (eds.), *The Ethics of Information Technologies* (Abingdon: Routledge).

TENNANT, J. P., *et al.* (2019) "Ten hot topics around scholarly publishing." *Publications* 7.2, 1-24, doi: 10.3390/publications7020034.

TENNANT, J. P. (2018), Twitter Account, https://twitter.com/protohedgehog/status/951413580167110656.

THAM, B., FINDLAY, J. M. (2020) "COVID-19 Vaccine Research, Development, Regulation and Access." *SMU Centre for AI & Data Governance Research*, 3.

THOMPSON KLEIN, J. (2018) "A Conceptual Vocabulary of Interdisciplinary Science." *Practising interdisciplinarity* (Toronto: University of Toronto Press).

TIPALDO, G. (2019) *La società della pseudoscienza*. (Bologna: Il Mulino).

TRACOL, X. (2020) " 'Schrems II': The return of the Privacy Shield." *Computer Law & Security Review* 39, 1-11, doi: 10.1016/j.clsr.2020.105484.

TREVISAN, M., *et al.* (2019) "4 Years of EU Cookie Law: Results and Lessons Learned." *Proceedings on Privacy Enhancing Technologies* 2019.2, 126-145, doi: 10.2478/popets-2019-0023.

TROSOW, S. E. (2004) "Sui generis database legislation: A critical analysis." *Yale Journal of Law and Technology* 534.7, 534-642.

TURILLAZZI, A., TADDEO, M., FLORIDI, L., CASOLARI, F. (2022) "The Digital Services Act: An Analysis of Its Ethical, Legal, and Social Implications." *Legal, and Social Implications*, 1-22.

UHLIR, P., CLEMENT, G. (2016) "Legal Interoperability of Research Data: Principles and Implementation Guideline." *RDA-CODATA Legal Interoperability Interest Group*.

UHLIR, P., SCHRÖDER, P. (2007) "Open data for global science." *Data Science Journal* 6, OD36-OD53.

URSIC, H., CUSTERS, B. (2016) "Legal Barriers and Enablers to Big Data Reuse." *European Data Protection Law Review*, 2.

VACCARI, L., *et al.* (2012) "Integrative research: the EuroGEOSS experience." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5.6, 1603-1611, doi: 10.1109/JSTARS.2012.2190382.

VAN CALSTER, B., *et al*. (2019) "Predictive analytics in health care: how can we know it works?" *Journal of the American Medical Informatics Association*, 26.12, 1651–1654, doi: 10.1093/jamia/ocz130.

VAN DER MAREL, E., BAUER, M., LEE-MAKIYAMA, H., VERSCHELDE, B. (2016) "A methodology to estimate the costs of data regulations." *International Economics* 146, 12–39. doi: 10.1016/j.inteco.2015.11.001.

VAN EECHOUD, M. (2021) "A Serpent Eating Its Tail: The Database Directive Meets the Open Data Directive." *IIC - International Review of Intellectual Property and Competition Law volume* 52, 375-378, doi:10.1007/s40319-021-01049-7.

VAN KLINK, B., TAEKEMA, S. (2012) "Limits and Possibilities of Interdisciplinary Research into Law.", in KIRSTE, S., *et al*. (eds.) *Interdisciplinary research in jurisprudence and constitutionalism*. (Stuttgart: Steiner).

VARGAS, V-M. (2019) "The new economic good: Your own personal data. An integrative analysis of the Dark Web." *Proceedings of the International Conference on Business Excellence*. 13.1, 1216-1226, doi: 10.2478/picbe-2019-0107.

VARRETTE, S., BOUVRY, P., PLUGARU, V., *et al*. (2018) "Overview and Challenges of the UL HPC Facility at the EuroHPC Horizon," *University of Luxembourg, 7th High Performance Computing School*, Luxembourg.

VASSILAKOPOULOU, P., SKORVE, E., AANESTAD, M. (2016) "Premises for clinical genetics data governance: Grappling with diverse value logics." in MITTELSTADT, B. D., FLORIDI, L. (eds.) *The ethics of biomedical big data.* (Cham: Springer), 239-256, doi: 10.1007/978-3-319-33525-4_11.

VAYENA, E., TASIOULAS, J. (2016) "The dynamics of big data and human rights: the case of scientific research." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083, doi: 10.1098/rsta.2016.0129.

VAYENA, E., TASIOULAS, J. (2015) "We the Scientists: A Human Right to Citizen Science." *Philosophy & Technology* 28, 479-485, doi: 10.1007/s13347-015-0204-0.

VEALE, M., ZUIDERVEEN BORGESIUS, F. (2021) "Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach." *Computer Law Review International* 22.4, 97-112, doi: 10.9785/cri-2021-220402.

VICENTE-SÁEZ, R., MARTÍNEZ-FUENTES, C. (2018) "Open Science now: A systematic literature review for an integrated definition." *Journal of business research* 88, 428-436, doi: 10.1016/j.jbusres.2017.12.043.

VINER, N., POWELL, P., GREEN, R. (2004) "Institutionalized biases in the award of research grants: a preliminary analysis revisiting the principle of accumulative advantage." *Research Policy* 33.3, 443-454, doi: 10.1016/j.respol.2003.09.005.

VOSS, G. V. (2020) "Cross-Border Data Flows, the GDPR, and Data Governance." *Washington International Law Journal*, 29.3, 485-532.

WADE BISHOP, B., *et al.* (2019) "Scientists' data discovery and reuse behavior: (Meta) data fitness for use and the FAIR data principles." *Proceedings of the Association for Information Science and Technology* 56.1, 21-31, doi: 10.1002/pra2.00004.

WANG, L. L., *et al.* (2020) Cord-19: The covid-19 open research dataset." *ArXiv.*

WEBER, R. H. (2014) "Legal interoperability as a tool for combatting fragmentation." *Global Commission on Internet Governance Paper Series n. 4.*

WELLER, M. (2014) The battle for openness. (London: Ubiquity Press).

WIEBE, A., DIETRICH, N. (2017) *Open Data Protection-Study on legal barriers to open data sharing-Data Protection and PSI* (Göttingen: Universitätsverlag Göttingen).

WILCOX, D. (2018) "Supporting FAIR data principles with Fedora." *LIBER Quarterly* 28.1, 1-8, doi: 10.18352/lq.10247.

WILKINSON, M. D., *et al.* (2016) "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1, 1-9, doi:10.1038/sdata.2016.18.

WILSON, M. C. (2017) "Universities spend millions of accessing results of publicly funded research." *The Conversation*, https://theconversation.com/universities-spend-millions-on-accessing-results-of-publicly-funded-research-88392.

WISE, J., *et al.* (2019) "Implementation and relevance of FAIR data principles in biopharmaceutical R&D." *Drug discovery today* 24.4, 933-938, doi: 10.1016/j.drudis.2019.01.008.

WOLF, M., *et al.* (2021) "Reusability first: toward FAIR workflows." *2021 IEEE International Conference on Cluster Computing CLUSTER*, 444-455, doi: 10.1109/Cluster48925.2021.00053.

WYNDHAM, J. M., WEIGERS VITULLO, M. (2018) "Define the human right to science." *Science* 362.6418, 975-975, doi: 10.1126/science.aaw1467.

YIN, R. K. (2009) *Case study research. Design and methods*, 4 eds. (London: Sage Publication).
ZAMBRANO, V. (2020) "Il «diritto umano alla scienza» e l'emergenza da CoViD-19." *BioLaw Journal-Rivista di BioDiritto* 20.1, 259-267, doi: 10.15168/2284-4503-584.
ZUBOFF, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs).

LEGISLATIVE ACTS, REPORTS OF AUTHORITIES AND DOCUMENTS OF ENTITIES OR ASSOCIATIONS

AMERICAN ASSOCIATION OF UNIVERSITY PROFESSORS (1915) Declaration of Principles on Academic Freedom and Academic Tenure, https://www.aaup.org/sites/default/files/Wilson_1.pdf.
BERLIN DECLARATION ON OPEN ACCESS TO KNOWLEDGE IN SCIENCES AND HUMANITIES (2003), https://openaccess.mpg.de/Berlin-Declaration.
BETHESDA STATEMENT ON OPEN ACCESS PUBLISHING (2003), https://osc.universityofcalifornia.edu/2003/04/bethesda-statement-on-open-access-publishing/.
BIDEN, J. R. (2020) "Memorandum on Restoring Trust in Government Through Scientific Integrity and Evidence-Based Policymaking", *The White House*, https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/memorandum-on-restoring-trust-in-government-through-scientific-integrity-and-evidence-based-policymaking/.
BUDAPEST OPEN ACCESS DECLARATION (2002), https://www.budapestopenaccessinitiative.org/read.
CNIL (2022) "La CNIL propose un nouveau « bac à sable » pour accompagner l'innovation numérique dans le domaine de l'éducation, https://www.cnil.fr/fr/la-cnil-propose-un-nouveau-bac-sable-pour-accompagner-linnovation-numerique-dans-le-domaine-de.
CNIL (2021) "La CNIL appelle à des évolutions dans l'utilisation des outils collaboratifs états-uniens pour l'enseignement supérieur et la recherche", https://www.cnil.fr/fr/la-cnil-appelle-evolutions-dans-utilisation-outils-collaboratifs-etatsuniens-enseignement-superieur-recherche.

COALITION S (2018) "Plan S. Making full and immediate Open Access a reality", https://www.coalition-s.org/.

CREDIT SUISSE (2020) "Global Wealth Report", https://www.credit-suisse.com/about-us/en/reports-research/global-wealth-report.html.

DEUTSCHE FORSCHUNGSGEMEINSCHAFT. DFG (2021) "Data tracking in research: aggregation and use or sale of usage data by academic publishers", *German Research Foundation*, https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_en.pdf.

EDPB (2021) *Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research*, https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en.

EDPB (2020) *Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak*, https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-032020-processing-data-concerning-health-purpose_en.

EDPB (2019) *Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (art.70.1.b)*, https://edpb.europa.eu/our-work-tools/our-documents/opinion-art-70/opinion-32019-concerning-questions-and-answers_en.

EDPS (2021) *Preliminary Opinion of the European Data Protection Supervisor on the European Health Data Space*, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021XX0323(02).

EDPS (2020) *Opinion 3/2020 European Strategy for Data*, https://edps.europa.eu/data-protection/our-work/publications/opinions/european-strategy-data_en.

EDPS (2020) *A Preliminary Opinion on data protection and scientific research*, available at: https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en.

EOSC Executive Board (2021) "Strategic Research and Innovation Agenda (SRIA)", v. 1.0, https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf.

EOSC Executive Board Working Groups FAIR and Architecture (2021) "EOSC interoperability framework." *Publications Office of the European Union*, doi: 10.2777/620649.

EOSC Executive Board Working Group (WG) Landscape (2020) *Landscape of EOSC-Related Infrastructures and Initiatives*, v. 2 (Luxembourg: Publications Office of the European Union), doi: 10.2777/132181.

EOSC Pillar Project (2021) "Legal and Policy Framework and Federation Blueprint.", v. 2.2, doi: 10.5281/zenodo.4486610.

EOSC Portal (2021) "EC President Ursula von der Leyen talks EOSC in Davos.", www.eosc-portal.eu/news/ec-president-ursula-von-der-leyen-talks-eosc-davos.

EOSC Symposium (2020) "Widening to the public and private sectors - Oct 21 #EOSCSymposium2020", EOSC Portal, Youtube, https://www.youtube.com/watch?v=dX_lPdiC24M.

ESFRI Long-Term Sustainability Working Group (2017) "Long-term sustainability of research infrastructures", *ESFRI Scripta Volume 2*.

EUROPEAN COMMISSION (2021) "Monitoring the open access policy of Horizon 2020: final report." *Publications Office*, 1-122, doi: 10.2777/268348.

EUROPEAN COMMISSION (2021) *Commission work programme 2022. Making Europe stronger together*, COM(2021) 645 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2021:645:FIN.

EUROPEAN COMMISSION (2021) "Infringement procedures against 19 Member States for not having transposed the ODD", https://data.europa.eu/en/news/commission-urges-19-member-states-comply-open-data-directive.

EUROPEAN COMMISSION (2021) *Horizon Europe: Budget*, DG Research and Innovation, doi: 10.2777/714209.

EUROPEAN COMMISSION (2021) Proposal for a Regulation of the European Parliament and of the Council, *laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM/2021/206 final, 21.04.2021, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206.

EUROPEAN COMMISSION (2020) *European Innovation Scoreboard 2020*, https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_1150.

EUROPEAN COMMISSION (2020) *on the protection of confidential information by national courts in proceedings for the private enforcement of EU competition law*, C/2020/4829, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020XC0722(01).

EUROPEAN COMMISSION (2020) Proposal for a Regulation of the European

Parliament and of the Council, *on contestable and fair markets in the digital sector (Digital Markets Act)*, COM/2020/842 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:842:FIN.

EUROPEAN COMMISSION (2020) Proposal for a Regulation of the European Parliament and of the Council, *on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC*, COM/2020/825 final, 15.12.2020, ELI: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2020:825:FIN.

EUROPEAN COMMISSION (2020) Proposal for a Regulation of the European Parliament and of the Council, *on European data governance (Data Governance Act)*, COM/2020/767, 25.11.2020, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767.

EUROPEAN COMMISSION (2020) *A new ERA for Research and Innovation*, COM/2020/628 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:628:FIN.

EUROPEAN COMMISSION (2020) *A European strategy for data*, COM/2020/66 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066.

EUROPEAN COMMISSION (2019), European Research and Innovation Days, 24-26 September 2019, Brussels, https://www.eoscsecretariat.eu/events/european-research-and-innovation-days.

EUROPEAN COMMISSION (2019) “Open Science”, https://ec.europa.eu/info/sites/default/files/research_and_innovation/knowledge_publications_tools_and_data/documents/ec_rtd_factsheet-open-science_2019.pdf.

EUROPEAN COMMISSION (2018) “How is consent for processing in scientific research obtained?”, https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/grounds-processing/how-consent-processing-scientific-research-obtained_en.

EUROPEAN COMMISSION (2018), *Proposal for a Regulation of the European Parliament and of the Council, establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination*, COM/2018/435 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018PC0435.

EUROPEAN COMMISSION (2018), *Towards a common European Data Space,* COM 2018/232, ELI: https://eur-lex.europa.eu/legal-

content/en/TXT/?uri=CELEX:52018DC0232.
EUROPEAN COMMISSION (2018), Recommendation (EU) 2018/790, *On access to and preservation of scientific information*, of 25 April 2018, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018H0790.
EUROPEAN COMMISSION (2017) *The Mid-Term Review on the implementation of the Digital Single Market Strategy. A Connected Digital Single Market for all*, COM/2017/228 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017DC0228.
EUROPEAN COMMISSION (2017) *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*, version 3.2, https://webgate.ec.europa.eu/funding-tenders-opportunities/display/OM/Online+Manual.
EUROPEAN COMMISSION (2016) *Guidelines on FAIR Data Management in Horizon 2020*, version 3.0, Directorate-General for Research & Innovation, doi: 10.25607/OBP-774.
EUROPEAN COMMISSION (2016) *European Cloud Initiative – Building a competitive data and knowledge economy in Europe*, COM/2016/178 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016DC0178.
EUROPEAN COMMISSION (2015) *A Digital Single Market Strategy for Europe*, COM/2015/192 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015DC0192.
EUROPEAN COMMISSION (2012) *High-Performance Computing: Europe's place in a Global Race*, COM/2012/45 final, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012DC0045.
EUROPEAN COMMISSION (2012) *A Reinforced European Research Area Partnership for Excellence and Growth*, COM/2012/392 final, ELI:https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012DC0392.
EUROPEAN COMMISSION (2012) Recommendation (EU) 2012/417, *On access to and preservation of scientific information*, of 17 July 2012, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012H0417.
EUROPEAN COMMISSION (2010) *A digital agenda for Europe*, COM/2010/245 final, ELI: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52010DC0245.
EUROPEAN COMMISSION (2002) Proposal for a Directive of the European Parliament and of the Council, *on the re-use and commercial exploitation of public sector documents*, COM/2002/0207, ELI:

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52002PC0207.

EUROPEAN COMMISSION (2000) *Towards a European research area*, COM/2000/0006 final, ELI: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:52000DC0006.

EUROPEAN MEDICINES AGENCY (2021), Vaccine Authorisation Press Release, https://www.ema.europa.eu/en/news/ema-recommends-first-covid-19-vaccine-authorisation-eu.

EUROPEAN UNION (2021) State of transposition and the implementation of the ODD in Member States, https://eur-lex.europa.eu/legal-content/EN/NIM/?uri=uriserv%3AOJ.L_.2019.172.01.0056.01.ENG.

EUROPEAN UNION (2021) Council Regulation (EU) 2021/1173 of 13 July 202, *on establishing the European High Performance Computing Joint Undertaking and repealing Regulation (EU) 2018/1488*, OJ L 256, 19.7.2021, ELI: http://data.europa.eu/eli/reg/2021/1173/oj.

EUROPEAN UNION (2021) Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021, *establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013*, OJ L 170, 12.5.2021, p. 1–68, ELI: http://data.europa.eu/eli/reg/2021/695/oj.

EUROPEAN UNION (2020) ERAvCORONA Action Plan, https://ec.europa.eu/info/sites/info/files/research_and_innovation/research_by_area/documents/ec_rtd_era-vs-corona.pdf.

EUROPEAN UNION (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019, *on open data and the re-use of public sector information*, OJ L 172, 26.6.2019, 56–83, ELI: http://data.europa.eu/eli/dir/2019/1024/oj.

EUROPEAN UNION (2019) Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019, *on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*, OJ L 130, 17.5.2019, 92–125, ELI: http://data.europa.eu/eli/dir/2019/790/oj.

EUROPEAN UNION (2019) *Agreement with the United States on mutual legal assistance*, https://eur-lex.europa.eu/summary/EN/jl0052.

EUROPEAN UNION (2018) Council Regulation (EU) 2018/1488 establishing the European High Performance Computing Joint Undertaking (2018) OJ L 252, 2, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R1488.

EUROPEAN UNION (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, *on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, OJ L 119, 4.5.2016, 1–88, ELI: http://data.europa.eu/eli/reg/2016/679/oj.

EUROPEAN UNION (2016) Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016, *on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA*, OJ L 119, 4.5.2016, p. 89–131, ELI: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016L0680.

EUROPEAN UNION (2013) Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC, o*n the re-use of public sector information*, OJ L 175, 27.6.2013, p. 1–8, ELI: http://data.europa.eu/eli/dir/2013/37/oj.

EUROPEAN UNION (2012) Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012, *on certain permitted uses of orphan works*, OJ L 299, 27.10.2012, 5–12, ELI: http://data.europa.eu/eli/dir/2012/28/oj.

EUROPEAN UNION (2012) Consolidated version of the Treaty on the Functioning of the European Union (TFEU), OJ C 326, 26.10.2012,47–390, http://data.europa.eu/eli/treaty/tfeu_2012/oj.

EUROPEAN UNION (2009) Conclusions of the Council on the definition of a 2020 vision for the European research area OJ C 25, 2009/C 25/01, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52009XG0131(01).

EUROPEAN UNION (2009) Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009, *on the legal protection of computer programs*, OJ L 111, 5.5.2009, 16–22, ELI: http://data.europa.eu/eli/dir/2009/24/oj.

EUROPEAN UNION (2007) Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007, *establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)*, OJ L 108, 25.4.2007, ELI: http://data.europa.eu/eli/dir/2007/2/oj.

EUROPEAN UNION (2004) Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004, *on the enforcement of intellectual property rights*, OJ L 157, 30.4.2004, 45–86, ELI:

http://data.europa.eu/eli/dir/2004/48/oj.

EUROPEAN UNION (2003) Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003, *on public access to environmental information and repealing Council Directive 90/313/EEC*, OJ L 108, 25.4.2007, p. 1–14, ELI: http://data.europa.eu/eli/dir/2003/4/oj.

EUROPEAN UNION (2003) Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003, *on the re-use of public sector information*, OJ L 345, 31.12.2003, 90-96, ELI: http://data.europa.eu/eli/dir/2003/98/oj.

EUROPEAN UNION (2001) Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001, *on the harmonisation of certain aspects of copyright and related rights in the information society*, OJ L 167, 22.6.2001, 10–19, ELI: http://data.europa.eu/eli/dir/2001/29/oj.

EUROPEAN UNION (1996) Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996, *on the legal protection of databases*, OJ L 77, 27.3.1996, 20–28, ELI: http://data.europa.eu/eli/dir/1996/9/oj.

EUROSTAT (2021) Research and development statistics at regional level, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Research_and_development_statistics_at_regional_level#Research_and_development_expenditure.

FOOD AND DRUG ADMINISTRATION (2020), Vaccine Authorisation Press Release, https://www.fda.gov/news-events/press-announcements/fda-takes-key-action-fight-against-covid-19-issuing-emergency-use-authorization-first-covid-19.

FOSTER PROJECT (2018), Open Science Training Handbook, https://book.fosteropenscience.eu/en/02OpenScienceBasics/01OpenConceptsAndPrinciples.html.

FRENCH PARLIAMENT (1978) Loi n. 78-17 du 6 janvier 1978, *relative à l'informatique, aux fichiers et aux libertés*, ELI: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/2021-12-11/.

GAIA-X EUROPEAN ASSOCIATION FOR DATA AND CLOUD AISBL (2021) *Gaia-X Architecture Document*, https://www.gaia-x.eu/publications.

GOVERNMENT OF CANADA, "Canada's 2018-2020 National Action Plan on Open Government," https://open.canada.ca/en/content/canadas-2018-2020-national-action-plan-open-government#toc8.

ICO (2018) *Principle (e): Storage limitation*, https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/storage-

limitation/#retention_periods.
LINDAU NOBEL LAUREATE MEETINGS (2021), *Lindau Annual Report* https://www.lindau-nobel.org/wp-content/uploads/2021/11/2021_Annual-Report_Web.pdf.
LORENTZ CENTER (2014) "Jointly designing a DATA FAIRPORT", Conference Report Workshop at Lorentz Center, http://www.lorentzcenter.nl/lc/web/2014/602/info.php3?wsid=602.
LUXEMBOURG PARLIAMENT, Loi du 1er août 2018, *portant organisation de la Commission nationale pour la protection des données et du régime général sur la protection des données* (Luxembourg Data Protection Law) (2018), ELI: https://data.legilux.public.lu/eli/etat/leg/loi/2018/08/01/a686/jo.
MINISTRE DE L'ENSEIGNMENT SUPÉRIEUR DE LA RECHERCHE ET DE L'INNOVATION FRANÇAIS (2021) *Deuxième Plan national pour la science ouverte*, https://www.enseignementsup-recherche.gouv.fr/cid159131/plan-national-pour-la-science-ouverte-2021-2024-vers-une-generalisation-des-pratiques-de-science-ouverte-en-france.html.
NIST Working Draft (2016) *An Action Plan for High Performance Computing Security*, https://www.nist.gov/system/files/documents/2018/03/15/working_draft_actionplanhpc.pdf.
NOYB, EUROPEAN CENTER FOR DIGITAL RIGHTS (2022) Austrian DSB: EU-US data transfer to Google Analytics illegal, https://noyb.eu/en/austrian-dsb-eu-us-data-transfers-google-analytics-illegal.
OECD (2022) *Inequalities in household wealth and financial insecurity of households*, https://www.oecd.org/social/income-distribution-database.htm.
OECD (2021) *Open Data in action. Initiatives during the initial stage of the COVID-19 pandemic*, https://www.oecd.org/gov/digital-government/use-of-open-government-data-to-address-covid19-outbreak.htm.
OECD (2020) *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation* (Paris: OECD Publishing), doi: 10.1787/947717bc-en.
OECD (2015) "Making Open Science a Reality", O*ECD Science, Technology and Industry Policy Papers*, n. 25 (Paris: OECD Publishing), doi: 10.1787/5jrs2f963zs1-en.
OECD (2007) *OECD Principles and Guidelines for Access to Research Data from Public Funding* (Paris: OECD Publishing), doi:

10.1787/9789264034020-en-fr.
PARLAMENTO ITALIANO, Decreto Legislativo 30 giugno 2003, n. 196 "*Codice in materia di protezione dei dati personali, recante disposizioni per l'adeguamento dell'ordinamento nazionale al regolamento (UE) n. 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE*", https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2003-06-30;196!vig.
PWC EU SERVICES (2018) *Cost of not having FAIR research data. Cost-Benefit analysis for FAIR research data*, Luxembourg: Publications Office of the European Union, doi: 10.2777/02999.
ROYAL SOCIETY (2015) *UK research and the European Union. The role of the EU in funding UK research*. (London: The Royal Society).
SYNERGY RESEARCH GROUP (2021) "European Cloud Providers Double in Size but Lose Market Share", https://www.srgresearch.com/articles/european-cloud-providers-double-in-size-but-lose-market-share.
TASK FORCE FAIR METRICS AND DATA QUALITY (2021) *EOSC Task Force FAIR Metrics and Data Quality Charter*, v. 1, https://www.eosc.eu/sites/default/files/tfcharters/eosca_tffairmetricsanddataquality_draftcharter_20210614.pdf.
UK PARLIAMENT (2018) UK Data Protection Act, https://www.legislation.gov.uk/ukpga/2018/12/contents/enacted.
UNESCO (2021) "UNESCO Recommendation on Open Science", https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en.
UNESCO (2020) "#ThinkBeforeSharing - Stop the spread of conspiracy theories",https://en.unesco.org/themes/gced/thinkbeforesharing.
UNESCO (2017) "UNESCO Recommendation on Science and Scientific Researchers",https://unesdoc.unesco.org/ark:/48223/pf0000260889.page=116.
UNITED NATION (2020) General comment No. 25 to the ICESCR, "On science and economic, social and cultural rights," (Article 15 (1) (b), (2), (3) and (4) of the International Covenant on Economic, Social and Cultural Rights), https://undocs.org/Home/Mobile?FinalSymbol=E%2FC.12%2FGC%2F25&Language=E&DeviceType=Desktop.
UNITED NATION (1994) General Assembly resolution, *Human rights and scientific and technological progress*, UN Doc. A/RES/48/140, 7 March 1994, https://www.aaas.org/resources/article-15/resources-international-human-rights-documents.

UNITED NATION (1966) International Covenant on Economic, Social and Cultural Rights, 1966, https://www.ohchr.org/en/professionalinterest/pages/cescr.aspx.

UNITED NATION (1948) Universal Declaration of Human Rights, https://www.un.org/en/about-us/universal-declaration-of-human-rights.

UNIVERSITY OF NORTH CAROLINA (2019) Hpc cloud facilities, http://herald.web.unc.edu/2019/11/how-cloud-computing-supports-green-technology/.

UNIVERSITY OF VIENNA (2018) The Vienna Declaration on the European Open Science Cloud, https://eosc-launch.eu/declaration/.

US CONGRESS (2018) Clarifying the Lawful Use of Overseas Data Act, CLOUD Act, Pub. L. No. 115–141, 132 Stat. 348 (codified as amended in separate sections of 18 U.S.C.) https://cli.re/BwPk5Q.

WORLD ECONOMIC FORUM (2020) "Ursula von der Leyen speaks at the World Economic Forum", Euronews, Youtube, 2020, https://www.youtube.com/watch?v=_A7Q514z_dw&feature=youtu.be&t=649.

WORLD HEALTH ORGANISATION (2019) Public statement for collaboration on COVID-19 vaccine development, https://www.who.int/news/item/13-04-2020-public-statement-for-collaboration-on-covid-19-vaccine-development.

WORLD HEALTH ORGANISATION, WHO Coronavirus (COVID-19) Dashboard, https://covid19.who.int/.

WP29 (2014) *Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC*, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/index_en.htm.

JUDICIAL RULINGS

BUNDESVERFASSUNGSGERICHT (1983) 1 BvR 209/83, Judgment of 15 December 1983, ECLI:DE:BVerfG:1983:rs19831215.1bvr020983.

CONSEIL CONSTITUTIONNEL FRANÇAIS (1984) Decision n. 83-165, *sur le libertés universitaires*, 20 Janvier 1984, https://www.conseil-constitutionnel.fr/decision/1984/83165DC.htm.

CONSEIL CONSTITUTIONNEL FRANÇAIS (1977) Decision n. 77-87, *sur la liberté d'enseignement*, 23 Novembre 1977, https://www.conseil-constitutionnel.fr/decision/1977/7787DC.htm.

EUROPEAN COURT OF HUMAN RIGHTS (2014), *Mustafa Erdoğan and Others v. Turkey,* Case 346/04 ECLI:CE:ECHR:2014:0527JUD000034604.
EUROPEAN COURT OF JUSTICE (2020) Case C-311/18, *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems*, ECLI:EU:C:2020:559.
EUROPEAN COURT OF JUSTICE (2020) *European Commission v Hungary* Case C-66/18, ECLI: EU:C:2020:792.
EUROPEAN COURT OF JUSTICE (2018) *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg,* Case C-304/07, ECLI:EU:C:2008:552.
EUROPEAN COURT OF JUSTICE (2017) Case C-434/16, *Peter Nowak v Data Protection Commissioner*, ecli:EU:C:2017:994.
EUROPEAN COURT OF JUSTICE (2015) *Evonik Degussa* v *Commission*, Case C-162/15 P, EU:C:2017:205.
EUROPEAN COURT OF JUSTICE (2015) Case C-362/14 *Maximillian Schrems v Data Protection Commissioner*, ECLI:EU:C:2015:650.
EUROPEAN COURT OF JUSTICE (2014) Case C-131/12, *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González*, ECLI:EU:C:2014:317.
EUROPEAN COURT OF JUSTICE (2012) *Akzo Nobel and Others* v *Commission*, Case T-345/12, EU:T:2015:50.
EUROPEAN COURT OF JUSTICE (2009) *Idromacchine* v *Commission*, Case T-88/09, EU:T:2011:641.
EUROPEAN COURT OF JUSTICE (2009) *Infopaq International A/S v Danske Dagblades Forening*. C-5/08 - Infopaq International, ECLI:EU:C:2009:465.
EUROPEAN COURT OF JUSTICE (2003) *Bank Austria* v *Commission,* CaseT-198/03, EU:T:2006:136.
TRIBUNALE AMMINISTRATIVO REGIONALE LOMBARDIA (2020) n. 201900007, Sezione Terza, 20/01/2020, ECLI:IT:TARMI:2020:115SENT.
TRIBUNALE AMMINISTRATIVO REGIONALE LAZIO (2020) n. 201900108, Sezione Terza, 22/07/2020, ECLI:IT:TARLAZ:2020:8579SENT.
US SUPREME COURT (2018) *Microsoft Ireland v United States*, 583 U.S., https://supreme.justia.com/cases/federal/us/584/17-2/.