Alma Mater Studiorum – Università di Bologna
in cotutela con Université du Luxembourg

DOTTORATO DI RICERCA IN

Law, Science and Technology

Ciclo 35

Settore Concorsuale: Filosofia del diritto 12/H3

Settore Scientifico Disciplinare: IUS/20 Filosofia del diritto

Big Data and AI Applications for Health and Research. Co-Regulation Mechanisms as a Prosed Solution for Data Protection Law Issues

Presentata da: Francesco Vigna

Coordinatore Dottorato

Monica Palmirani

Supervisore

Monica Palmirani

Co-Supervisore

Luca Ratti

Co-Supervisore

Massimo Durante

Esame finale anno 2023

UNIVERSITÉ DU LUXEMBOURG

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

PhD-FDEF-2023-xxx
The Faculty of Law, Economics and Finance

The Department of Legal Studies

DISSERTATION

Defence held on 05-07-2023 in Bologna

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN *DROIT*

AND

DOTTORE DI *RICERCA IN LAW, SCIENCE AND TECHNOLOGY*

by

VIGNA Francesco
Born on 24 July 1991in Moncalieri (TO) (Italy)

BIG DATA AND AI APPLICATIONS FOR HEALTH AND RESEARCH. CO-REGULATION MECHANISMS AS A PROSED SOLUTION FOR DATA PROTECTION LAW ISSUES

Abstract

Big data and AI are paving the way to promising scenarios in clinical practice and clinical research. However, the use of such technologies might clash with some GDPR requirements. Today, two forces are driving the EU policies in this domain. The first is the necessity to protect individuals' safety and fundamental rights. The second is to incentivize the deployment of innovative technologies. The first objective is pursued by legislative acts such as the GDPR or the AI Act proposal, the second is supported by the new data governance stemming from the data strategy recently launched by the European Commission.

Against this background, the thesis analyses the issue of GDPR compliance when big data and AI systems are implemented in the health domain. The thesis focuses on the use of co-regulatory tools for facilitating compliance with the GDPR. This work argues that there are two level of co-regulation in the EU legal system. The first, more general, is the approach pursued by the EU legislator when shaping legislative measures that deal with fast-evolving technologies. The GDPR can be deemed a co-regulatory solution since it mainly introduces general requirements, which implementation shall then by interpretated by the addressee of the law following a risk-based approach. This approach, although useful for fast evolving situations is costly and sometimes burdensome for organisations. The second co-regulatory level is represented by specific co-regulatory tools, such as code of conduct and certification mechanisms. These tools are meant to guide and support the interpretation effort of the addressee of the law.

The thesis argues that the lack of co-regulatory tools which are supposed to implement data protection law in specific situations could be an obstacle to the deployment of innovative solutions in complex scenario such as the health ecosystem. The thesis advances hypothesis on theoretical level about the reasons of such a lack of co-regulatory solutions.

Index

Introduction

1 – Background

Promising scenarios are stemming from the deployment of big data and artificial intelligence systems (hereinafter AI) in health care and scientific research in medicine. [1–3] The advantages provided by AI in health domain is given by the extraction of knowledge from data using inferential and predictive analysis. Such knowledge is then used to guide decision-making process or to improve clinical practice. Several examples of big data and AI potential uses in health domain can be mentioned. For example, predictive analysis can be deployed in order to implement precision and personalised medicine, to improve diagnostic tasks, or to monitor risky patient population. [4–6] The process of drug discovery and clinical research can benefit as well from AI implementation. [7, 8] Even more complex implementations, such as surgical robots or companion robots, are being developed in order to assist human professionals in common clinical practice. [9–11]

The deployment of big data and AI faces obstacles in the General Data Protection Regulation (hereinafter GDPR)[1] principles when it is necessary to process big amount of personal data through fuzzy procedures and for purposes not always defined in advance. These difficulties are essentially due to some incompatible dichotomies between GDPR principles and big data dynamics. [12] However, on the other hand, the EU legislator is trying to boost the use, and especially the re-use, of personal data across EU, for enhancing the market of innovative technologies. [13, 14] In this perspective, the EC has recently launched a new strategy for the enhancement of data sharing and data re-use, with a great focus on electronic health data and their use for scientific research. These policy actions had added a further layer of regulatory provisions that shall be coordinated with the GDPR. [15, 16]

Two different forces are therefore influencing the action of the EU legislator in terms of policy and regulatory strategies. The first one is given by the necessity to develop technologies and systems that respect individual rights and freedoms, but also the safety of human beings. The second one is the desire to improve the sharing and the re-use of high-quality data in order to enhance the use of such technologies. The use of these technologies is motivated by the fact that their deployment in strategic sectors, such as the health domain, is key to enhance performance and results from many points of view, as mentioned before. Nevertheless, the best performance of AI is directly linked to the availability of great amount of data from different and heterogeneous sources. Making such data available for big data and AI technologies is extremely difficult in light of GDPR requirements. [12, 17, 18] The difficulties are not only due to, as mentioned above, some incompatibilities between the GDPR's principles and big data, but also to some practical inefficiencies at the compliance level. In other words, implementing the GDPR's requirements into day-by-day activities is extremely difficult in some domains and context. An example in this sense is the concept of further processing for re-using personal data in biomedical research. [19] This is due to due to some efforts in terms of interpretation that the GDPR requires for having data processing activities being in compliance with data protection law.

[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).

The EU legislator is called to face a challenging task: creating the conditions for the use of big data and AI systems, without waving fundamental rights. Moreover, the goal is to create a regulatory framework that protect individuals without suffocating the deployment of innovative technologies and the potential benefits stemming from them. Compliance costs[2] are indeed among the main regulatory costs impacting organisations. For succeeding in this difficult challenge, a regulatory shift is observable in data protection law. The EU legislator has tried to move from formalistic approach to data protection law towards a risk-based approach in terms of compliance demonstration. This approach shift has been a choice due to different reasons converging in the need to create a regulation able to deal with fast moving technologies. Therefore, most of GDPR's requirements only set some general principles, while large part of their actual implementation and interpretation is left to data controllers and processors. The choice concerning the implementation of the requirements shall be done after a risk assessment, in order to demonstrate that the technical and organisational measures adopted were appropriate to the case. [20] The aim pursued by the EU legislator is removing administrative duties, like ex-ante notifications or ex-post control by data protection authorities, and replace them with a constant (i.e., during the whole data processing life-cycle) risk evaluation. According to the risk level inherent in the data processing, data controllers shall adopt appropriate measures to protect rights and freedoms of individuals. The data controller has, therefore, the duty of being able to demonstrate the proper application of the measures. These duties of evaluation and demonstration stem from the accountability principle introduced by the art. 5(2) GDPR.

In light of the mentioned shift, it is possible to affirm that the GDPR adopts a *co-regulatory approach*. Indeed, in the GDPR the implementation of the law is largely delegated to discretionary risk-assessments carried out by the addressee of the law. This approach, although necessary for facing fast moving technologies, create high compliance burden on data controllers and processors. However, the use of *co-regulatory tools/instruments* – Codes of Conduct (hereinafter also CoC) and Certification Mechanisms (hereinafter also CM) – in the GDPR was meant to smooth the application of such a heavy piece of legislation and tackle the raise of compliance costs.

The role of non-mandatory and quasi-regulatory solutions in data protection law has been enhanced by the GDPR respect the previous Data Protection Directive (DPD)[3]. Indeed, the GDPR has introduced a new co-regulatory instrument: the CMs. On the other hand, the role of CoCs under GDPR has been strengthened, although these instruments were already present under the DPD. As it will be discussed in the thesis, CoCs and CMs in GDPR are supposed to be compliance instruments. Data controllers and processors can use them to reach and demonstrate compliance, in light of the new accountability principle introduced by the GDPR. [21, 22]

[2] Compliance costs can then be divided between 1) administrative burdens; 2) substantive compliance costs for organisations. Administrative burdens are costs stemming from information obligations towards authorities. Substantive compliance costs are all the other costs borne by a group for compliance activities, such as labour costs, implementation costs, or external service costs. [226]

[3] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

2 – Notes about the methodology

Against this background, the thesis investigates data protection legal aspects when big data and AI are implemented in health context. To be more precise, the work focuses on the role played by alternative regulatory solution other than hard law provisions. The instruments here analysed are classified as co-regulatory tools/instruments. Their role is provided by articles 40 and 41 GDPR, as for codes of conduct (CoC), and in article 42 and 43 GDPR, as for certification mechanism (CM). These instruments are supposed to be developed by stakeholders under the control and the final approval of data protection authorities (DPAs). Their adoption is voluntary, but if adopted these instruments can produce some legal effects[4]. However, to date, these solutions do not find large application in the EU data protection law domain.

The aim of the work is to understand to what extent co-regulatory instruments can help to face the complexities of healthcare scenario, taking into account also new regulatory and policy strategies pursued by the EU legislator. Essentially, in light of the burdens in terms of compliance generated by the risk-based approach of the GDPR, this work will try to understand the impact of co-regulatory tools. These instruments should relieve data controllers and processors from part of their burden in reaching and demonstrating compliance. Moreover, such instruments also play other roles, such as trust enhancement in the ecosystem or competitive advantage on the market. [23, 24] All these aspects will be object of the discussion carried out throughout the thesis.

The work adopts a theoretical approach in addressing the issues presented here, starting from the assumption that GDPR compliance is a concrete issue for organizations that process big amount of personal data in the health domain. The literature production cited throughout this work confirms that the implementation of GDPR requirements is a critical aspect, especially for the domain analysed here. As regards the co-regulatory instruments, the literature production is much less abundant. In this regard, the present work tries to fill this gap, encouraging also future empirical research that are needed to assess more precisely cost-related impacts of these instruments on organizations compliance.

3 – Summary of the Content

The thesis is divided into four chapters. Chapter 1 “Big data and AI applications in healthcare” will provide a first introduction of the context. Namely, an overview of the phenomena of big data and AI in health is provided in Section 1. The concept of big data is put in context in Section 1.1, where a contextualised definition and an overview of the big data supply chain in health are presented. Moreover, a summary of the main applications of AI systems in health is provided in Section 1.2. After introducing these concepts in the context of health, the rest of Chapter 1 starts the overview from a legal perspective. Section 2 analyses the legal framework, focusing in the first place on data protection law, in Section 2.1. The thesis also briefly analyses the proposed AI Act in Section 2.2. Section 3, in conclusion, deals with the recent policies adopted by the EC for enhancing the sharing and re-use of data. Section 3.1 in particular presents the Data Governance Act (hereinafter

[4] The legal effects of GDPR co-regulatory instruments are not completely clear and the literature there seem to be different opinions on this point, see [22, 154].

DGA)[5] and the proposal for a regulation on the Common European Health Data Space (hereinafter EHDS)[6]. On the other hand, Section 3.2 describes the concerns that these new policies present in terms of interplay with data protection law legislation.

Having presented the context, both from a technological and legal perspective and the related concerns, Chapter 2 "Co-Regulatory Instruments for Data Processing" moves to the analysis of co-regulatory solutions. Section 1, first of all, provide a definition of the co-regulation phenomenon (Section 1.1) as well as a taxonomy of it (Section 1.2). Section 2 then delve into the discussion of the specific co-regulatory instruments in data protection law domain. Section 2.1 provides a short historical excursus of co-regulation under the DPD. After that, Section 2.2 analyses the specific instruments of CoC under articles 40 and 41 GDPR. On the other hand, Section 2.3 presents the CM instrument under articles 42 and 43 GDPR. Chapter 3 concludes analysing the main impacts and effects that can be generated by the application of co-regulatory measures in data protection law. Namely, Section 3.1 investigates the role played by these instruments in terms of compliance facilitation. Section 3.2 analyses the prospected impacts in terms of legal certainty and reduction of legal fragmentation. Section 3.3 delves into the aspects related to market dynamics. In conclusion, Section 3.4, in light of the forecasted impacts and roles of these instruments try to carry out a first reasoning on the scarce application of these instruments in data protection law in general.

Chapter 3 "Proposal for a Co-regulatory Model of Governance in Health" tries to translate into the context of health the analysis carried out, in general, in Chapter 2. Therefore, Section 1 analyses the two main aspects that might influence the success or the failure of co-regulation solution in health. These points consist in the definition of the material and geographical scope of the co-regulatory measure (Section 1.1) and in the impacts on stakeholders' behaviours (Section 1.2). Section 1.3, recalling in part the discussion at the end of Chapter 2, tries to forecast realistic expectations from the use of these instruments in health while big data and AI solutions are implemented. Section 2 then focuses just on one of the two co-regulatory solutions of the GDPR, i.e., Codes of conduct. The choice to focus just on CoCs has been done for the sake of the discussion, given the relevance of the specific instrument for the domain at issue. Namely, this section tries to speculate some potential content of a CoC for the use and re-use of personal data in health. The following aspects have been analysed: 1) anonymisation and pseudonymisation (Section 2.1); Data re-use for scientific research (Section 2.2); Data subjects' consent and data altruism (Section 2.3); Data controllership (Section 2.4); Legal basis for primary and secondary uses (Section 2.5); Data minimisation, purpose and storage limitation principles (Section 2.6). This is not supposed to be an exhaustive list of arguments that shall be included in such instruments. Otherwise, it should be considered as a list of relevant topics that, among others, generate issues in terms of compliance in health domain. The points mentioned are analysed especially in light of the application of big data and AI solutions as well as in consideration of the new interplay between GDPR and the DGA and the EHDS.

Chapter 4 "Proposal for a Code of Conduct for Health Data Processing" move the analysis on a more practical level. Here, the thesis tries to forecast the fundamental steps for

[5] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).
[6] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, COM(2022) 197 final.

the elaboration of a CoC under article 40 and 41 GDPR for the health domain. First of all, a road map for the elaboration of a CoC is sketched in Section 1. The macro-steps to follow are: 1) the elaboration of a scientific basis for justifying the adoption of the CoC, namely in the form of a white paper (Section 1.1); then the second step should be the definition of the content of the CoC and the objectives to pursue (Section 1.2); the involvement of stakeholders (Section 1.3); and finally the submission to the DPA for obtaining the approval (Section 1.4). After envisaging the road map, the rest of the Chapter deals with the structure of the CoC and the organisation of its content in Section 2. Namely, the content of a CoC could be organised as follow: 1) the essential parts required by the EDPB guidelines (Section 2.1); and the subject matter/the material scope of the CoC (Section 2.2).

Finally, the thesis attempts to draw some conclusions (Conclusion) on the future of the use of these instruments and on their potential role, as well as on corrections to be done for facilitating their spread especially in the health domain. Moreover, some speculations are made in terms of the suitability of these instruments in facilitating the application of data protection law and on smoothing its interplay with other legislative acts.

Chapter 1 – Big Data and AI Applications in Health

The deployment of AI technologies for medicine-related purposes, is not a novelty, as AI is not a new technology per se. Indeed, AI has roots in studies carried out by several scholars[7] between 1943 and 1955, a period that Russel and Norvig define as "the inception of artificial intelligence". [25] After a first era of great expectations, AI research and investments experienced a slowdown. In general, AI history is characterised by moments of optimism alternated to so called "AI winters". Nevertheless, the big data phenomenon has recently raised again the expectation for a new dawn of AI solutions and, in general, data driven technologies. [25] The health domain is being affected by the these technologies, as well as almost any other sector.

"Health system" is intended here as the system of institutions, public and private, that collaborate to provide assistance to individuals in the medical domain, as well as those which contribute to the advance of clinical research. Therefore, by health system we can mean the complex of entities, and the relationships among them, that make possible the provision of care by healthcare professionals to patients, in order to assess, maintain or restore their state of health[8]. However, it should also include clinical trials, observational studies, drug discovery, and scientific research in general[9].

The deployment of AI in the health domain, despite carrier of outstanding results and improvements in health services provision, presents also concerns, probably more than in other sectors. One concern is due to the sensitivity of the domain at stake, both as regards the information processed and the risks potentially stemming from a wrong functioning of the AI systems. Another issue is the complexity of the sector, which is due to the several types of operators involved (hospitals, medicine professional, pharmaceutical companies, assurances companies, and many others). Finally, economic constraints, especially for public health care providers, should also be taken into account.

The aim of this chapter is to provide a first overview of the phenomena of big data and AI in the health domain and the legal and policy framework surrounding the use these technologies. Data protection law is the focal point for the analysis. The other policies and regulatory actions will be analysed in light of their interplay with data protection law. Therefore, the Chapter is organised as follows. Section 1 will provide an overview of the technical application of big data and AI systems in health. Namely, Section 1.1 will focus on the big data phenomenon at the basis of the AI systems' functioning. In this sense, Section 1.1.1 contextualise the big data phenomenon into the health domain trying also shed some light on the many definitions of big data in health. While, on the other hand, Section 1.1.2 will analyse the steps that constitute the so-called big data supply chain. These steps go from data collection to the interpretation of information through AI system. Afterwards, Section 1.2 defines the concept of AI in general and then put it in the context of health.

[7] Among others, W. McCulloch, W. Pitts, D. Hebb, M. Minsky, D. Edmonds, A. Turing.

[8] Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare.

[9] The concept of scientific research is broad and includes many types of research. In data protection law context, the GDPR provides some indication of the scope of scientific research in recital (159), including, among others, technological development and demonstration, fundamental research, applied research and privately funded research.

Section 1.2.1 indeed focuses on the definitions and the different approaches to AI systems adopted throughout the years. Finally, Section 2.2 is an overview that provides examples of AI applications in health domain.

Moving to Section 2, the thesis analyses the legal framework within which the analysis is conducted. Section 2.1 delve into the domain of data protection law. Section 2.1.1 defines the scope (material and geographical) of the GDPR. Section 2.1.2, on the other hand provides an overview on the principles of the GDPR, although focusing on the principle of accountability. Section 2.1.3 analyses the general risk-based approach adopted by the GDPR for managing compliance activities. Section 2.2 gives a quick overview of the AIA proposal as another piece of legislation strongly connected to data protection law and the implementation of AI systems in health.

Section 3, the last of the chapter, address the set of policies launched by the EC for boosting the use and re-use of personal data in the digital single market. For this, Section 3.1 provides and overview of such strategies and on the legal acts stemming from them. Namely, in Section 3.1.1 the DGA is analysed, while in Section 3.1.2 the EHDS is discussed. Section 3.2, on the other hand, presents the issues of these legislative acts in light of their interplay with the GDPR. In particular, the attention is focused on: 1) definition and terminology issues (Section 3.2.1); 2) the roles and responsibilities of the new actors (Section 3.2.2); 3) the legal basis and data re-use mechanisms (Section 3.2.3). The section and the chapter are closed by some introductory observation on the use of co-regulatory instruments to solve such problems (Section 3.2.4).

1 – Overview of the Big data and AI Phenomena

1.1 – Big Data in General and in Health

The so called "big data" phenomenon has its roots in the advancement of calculators' capacities, as well as in features of the modern society. From a computational point of view, the big data phenomenon is due to the possibility to process bigger amount of data at faster rate than in the past. From a social point of view, the greater availability of data is given by the rise of the information society, which pushes individuals to share personal information in a digitalised format. Through social networks, search engines, or mobile apps, individuals share personal information with public and private companies for disparate purposes. [26]

This data, which often falls in the category of "personal data"[10], are increasingly used to lead the decision-making process of several actors (so called data-driven decision-

[10] The definition of personal data is provided by the GDPR as follow:

"'any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

Further criteria to take into account for the evaluation of the identifiability of a natural person are provided by recital (26) GDPR. Namely, in order to understand if a piece of information can be traced back to a natural person it shall be taken into consideration all the means "reasonably likely to be used" to identify the subject. For example, the analysis should evaluate objective factors such as the costs and time to identify someone in light of the technological state of the art.

making)[11]. The final application of these data elaboration is usually performed through AI technologies. [27–29]

1.1.1 – Contextualisation of Big Data

Probably, the most common definition of big data is the one referring to the three "Vs": "big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation"[12]. However, even more "Vs" have been added to define the big data phenomenon, such as "veracity" or "variability". [30] However, there is no quantitative definition, for instance, in terms of gigabytes (GB), for the classification of data processing as big data. Big data should rather be considered all data operations, starting from collection to exploitation, that could not be handled through traditional software and hardware instruments within an acceptable time frame. [31] These big data operations include, eventually, an automatic extraction of knowledge through advanced data analytics technologies [32], which often fall in the domain of AI.

Despite the definition of big data does not revolves around numbers, some quantitative information can help to understand the phenomenon. Indeed, the growth of data processed in recent years testifies the arise of big data. With 28 zettabytes (ZB)[13] of data processed in 2018, the volume of data grew tenfold respect 2011. Projections see a further rise of data processing by the year 2025 until reaching 175 ZB. [33] The main catalyst of big data is, as mentioned above, the internet. Namely, key is the role played by some platforms which base their business model on data collection from individuals, such as search engines or social networks. To give some examples, every minute 2.3 million of query are performed through Google and 2.7 million of videos are downloaded from YouTube. Facebook, on the other hand, generates 10 PB[14] per month and Alibaba dozens of TB[15] every day[16].

In conclusion, big data is a phenomenon that invests the whole society and that is changing the modus operandi in many sectors, including provision of care and biomedical research. The big data phenomenon is affecting the way data are processed from their collection to their final use. For this reason, the next section will address the big data phenomenon dividing the data life-cycle into different phases.

1.1.2 – The Big Data Supply Chain

Beyond the definition, probably big data could be better understood analysing its life cycle through the lens of the "supply chain" of big data. [34] The supply chain of big data is divided into 1) data collection; 2) data elaboration; 3) data interpretation.

1.1.2.1 – Data Collection

---

[11] The data-driven decision making is opposed to the human-driven or query-driven decision making. The data driven decision making rely on datasets for starting the decision-making process when even the analyst does not know where to look at exactly. [12]

[12] Gartner IT Glossary, definition of big data, https://www.gartner.com/en/information-technology/glossary/big-data (last access: January 2023).

[13] 1 zettabyte = 10^{21} byte.

[14] 1 petabyte = 10^{15} byte.

[15] 1 terabyte = 10^{12} byte.

[16] From https://www.go-globe.com/things-that-happen-every-60-seconds/ (last access: January 2023).

The phase of data collection is further divided into other three sub-phases, which are: a) the generation of data, b) the acquisition of data, and c) data storing. [34, 35] Data generation relates to all the activities performed by individuals that produce information. These activities could be performed through different modalities, from the use of a mobile app or a wearable device, or even through passive activities like the geo-localization of a smartphone from a mobile app in background. [34, 35]

The following phase, the acquisition of data, relates to the gathering of data produced by individuals during the previous phase. The entities that collect the data from individuals, afterwards, will process such data, which have been gathered from different sources. According to the way data is generated, also the standards and the formats of such data would be heterogeneous. In this step, there could be two different scenarios. In the first scenario, data is directly collected from individuals that generated it. In the second case, data is collected through intermediaries, such as data brokers. [26]

Finally, data shall be memorised in systems and kept ready for their elaboration. The storage of big data, of course, requires enough storage capacity. However, the technology development allows nowadays to store big amount of data at limited cost. [34]

1.1.2.2 – Data Elaboration

The second phase of the supply chain consists in the elaboration of data and, namely, in the activities of integration and analysis. Indeed, after having selected data sources and having stored data on a memory, the following step is the aggregation and organization of raw data. The objective is to transform raw data in datasets that can be then analysed. [34]

The analysis of data probably is the most important step of the big data phenomenon. Indeed, this is the process meant to elaborate raw data and extract structured information from them; this information generates new knowledge to be used for practical decisions. The analysis of data is usually performed through the use of algorithms. [34]

AI algorithms are able to extract particularly innovative knowledge from big amount of data using technologies such as machine learning (ML). The use of such technologies for the analysis of data is revolutionizing the old paradigm of decision making, i.e., hypothesis – model – experiment. Indeed, AI technologies do not only use data to verify hypothesis, but also to infer results and connection without explicit human command. On the basis of these inferences new theories are developed autonomously, this is the so-called data-driven decision making[17]. [34, 35]

Key for this innovative way to discover knowledge is the amount of data and the variety of data sources. The more the data is available the better the algorithm can perform the inference task. However, although such an assumption is in general true, also the quality of data is relevant. Data quality in health domain is a broad field which touch different data processing in health domain, from Electronic Health Record (EHR) to clinical data. [36–38] However, since this topic is not in the scope of the present work, it is sufficient to know that also in sectors like health care and research the dictum, popular in the ICT domain, "garbage in, garbage out" is still valid. In other words, inferences based on low-quality data produce not reliable outputs. [34]

1.1.2.3 – Data Interpretation

[17] See above, Section 1.1.1.

The third, and last phase, is the interpretation of big data, which is the final step of the supply-chain. This phase consists in the use of the innovative knowledge extracted from big data for adopting efficient decisions. As already mentioned, AI allows for a decision-making process based on data even without a thorough activity of theory elaboration performed in advance. The theory, in these cases, is directly elaborated by the algorithm thanks to inferences grasped from data, i.e., the data-driven decision-making process. AI algorithms are therefore able to detect correlation among data that humans are not able to see.[29, 39]

In this scenario, data is both used as feedback for the improvements of past decision of the model and input for future decisions. Big data and AI-based decision-making processes have several applications. For instance, they could ameliorate operational decisions within organizations. Moreover, they could provide innovative services that otherwise would not be possible to supply (e.g., traffic information through geo-localization of users). This approach can even be used for personalization of services and products, or to provide better and innovative public services. Even though big platforms such as Google, Facebook, or Amazon are the companies that extract most of the economic value from big data, this is not a phenomenon limited to them. Indeed, also traditional sectors can gain a significant competitive advantage from the exploitation of big data. [34] Nevertheless, it should be borne in mind that the new paradigm of data-driven decision-making cannot be applied as-is in every domain. In the health domain, for example, it shall be coordinated with the traditional medical-scientific method. [26, 40] Therefore, even though several functions and activities can potentially benefit from big data and AI applications in the health domain[18], the implementation of such technologies shall be carefully analysed in light of the context.

1.1.2.4 – Big Data in the Health System: Several Possible Definitions

Proceeding to a contextualisation of the big data phenomenon in health care, it is possible to observe, in a preliminary way, some critical features. Indeed, health care is a sector characterised by a relatively low level of digitalisation. [41] For example, part of the data is still stored in traditional format. Furthermore, even where datasets are digitalised, differences in the data standards hamper the sharing and pooling of big amount of data. [41] This is mainly because investments in digitalisation used to be lower in health care than in other sectors. [41] Nevertheless, recently, actions for improving the sharing and the interoperability of electronic health data have been launched by the EC[19].

Besides the limitations mentioned above about low standardisation and digitalisation, the health system presents other features that have an impact on big data use. These peculiarities concern aspects that make the health domain different to other sectors. Namely, such aspects relate to the heterogeneity of the activities and relationship carried out among individuals and organisations involved. Consequently, also the data processing and the relative purposes will be extremely different to each other. For this reason, many definitions of big data for the health domain have been proposed in order to better shape the phenomenon. In this regard, Mehta and Pandit have conducted a literature review about the topic

[18] For an overview of the areas that can benefit from big data and AI implementation, see [6].

[19] See the proposal for a regulation of the European Health Data Space (EHDS), [110] which will have the aim of promoting the sharing of personal electronic health data for primary and secondary uses. The EHDS will also promote the adoption of interoperability standards.

of big data in medicine. [1] In this regard, it comes up that even though providing a unique definition of big data in the health system has been a common effort shared by many scholars, there is no consistency among them.

Some scholars still try to rely on the paradigm of the "Vs", like Bian et al. or Bates et al, i.e., trying to focus on the characteristic of data. [2, 6] Dinov also develops his theory starting from the characteristic of data, however adding a two additional feature that should define big data in health care. Namely, the "energy" and the "lifespan" of data. Energy refers to the "holistic content included in data", this "energy" is higher in aggregated datasets than in individual ones, making the former more suitable to be exploited than the latter. As regards the lifespan, it should be intended as the ability of the data to keep value throughout the time after the first gathering. [42]

Other authors believe that the deployment of exceptional effort in terms of analytical and management tools for data processing is the feature that characterises big data. He et al. considers as "big" the datasets that conventional methods of data processing cannot handling because of the volume and the complexity of the records. [43] Raghupathi and Raghupathi also rely on a similar assumption, specifying though that the impossibility of elaborating data could either stem from restrictions of traditional software or common data management tools. [44] Scruggs et al., differently, suggests that what differentiate big data from traditional data processing is the opportunity to re-use data over the time, and to accumulate value throughout its life cycle thanks to new layers of knowledge extracted from it. [45] Focusing on the use of data is also Ghani et al., here the authors believe that big data should be potentially used to address several research questions and not a specific one. [46]

Other authors that concentrate on qualitative aspects of data are Auffray et al., which detect the heterogeneity in the types of data as the main peculiarity of big data in health. Indeed, biological, clinical, environmental, lifestyle data can either be collected about single individuals, or cohorts, either once or repeatedly. [7]

Worth of note is the position of Baro et al. which, on the other hand, explicitly seems to endorse a strictly quantitative definition of big data. According to the authors, it should be considered as big data the datasets with Log $(n*p) \geq 7$, where "n" is the number of individuals and "p" is number of variables. [47] However, besides this last contribution, it seems that a quantitative definition of big data is not achievable nor desirable. This conclusion can be argued because the classification of a given volume of data as "high" or "big" is relative to many factors like the time or the advancement in technology. Therefore, such a definition is intended to become obsolete sooner or later.

Interesting is the work of Shilo et al., where the authors do not try to provide a common definition of big data since they assume it a relative concept. Shilo et al. rather explain big data in health as a phenomenon that has a variable geometry. Indeed, in the health domain, big data processing is characterised by several aspects that Shilo et al. call "axes". [48] The axes of big data in health care are a) the "number of participants", b) the "depth of phenotyping", c) the "longitudinal follow-up", d) the "interaction between subjects included in the data", e) the "heterogeneity and diversity of the cohort population", the f) "standardization and harmonization of data", and the g) "linkage between data sources"[20].

[20] See Figure 1, from Shilo et al [48].

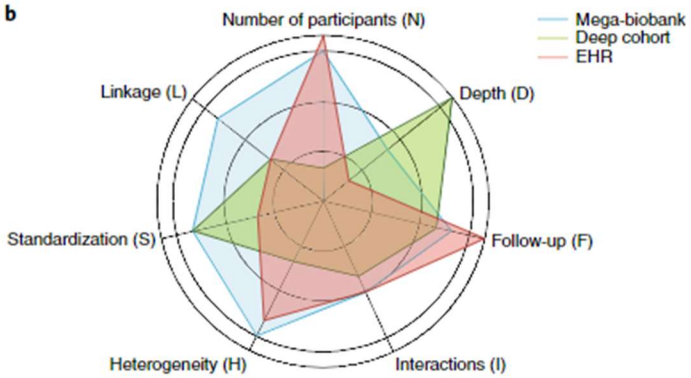

Figure 1 – Shilo S., et al. (2020)

According to how the big data phenomenon is shaped upon these axes, the features will change, forming a different geometry of big data. Indeed, a big data project that maximise all the features in all the axes is neither possible nor desirable. On the contrary, there are trade-offs between the axes. Consider, for example, the axes "number of participants" and "depth of phenotyping". The former indicates the number of individuals whose data are included in dataset, i.e., the size of the cohort. In this regard, reaching enough elevated number of participants is crucial to have "statistical power". On the other hand, the depth of phenotyping indicates how many aspects are observed about each individual. Data could be collected about socioeconomic aspects – e.g., the ethnicity – or anthropometrics features – e.g., weight and height – or about lifestyle habits, or psychological factors, or for example physiological measurements – e.g., blood pressure, heart rhythm. A dataset that measures everything for everybody is not realistic, even using enormous computational power. Therefore, a trade-off shall be reached between the number of subjects involved and the number of features observed for each of them. [48]

For example, EHR will be characterised by high number of participants (even millions), but the aspects observed among them will rarely go beyond few observations; on the other hand, cohorts focused on the studying of a rare disease will deeply observe almost every aspect of a patient, though, in this case, the population observed will be very limited.

Having ascertained that is difficult to find a unique definition of big data in health care, it is better to analyse the big data supply chain through the health care perspective.

1.1.2.5 – Data Collection in Health

As said before, the data collection phase can be divided into three sub-phases: a) data generation; b) data acquisition; and c) data storage.

The "data generation" step in the health system is performed through different ways. The first general distinction can be made between: 1) data generated inside the health care system by the interaction of patients with health care professionals; 2) and data generated by the patients outside the health care system and then brought, only in a following moment, inside the health domain. Within the first category are included data generated by patients visiting hospitals or other health care facilities for several reasons, such as being subjected to surgical operations, routine visits, or monitoring activities. Part of the second category is data initially not generated by patients for the purpose of being deployed in the

health care, but that could however be used for health care provision or biomedical research in a second time. For example, data generated using social networks, mobile app, or wearable device. This second category of data generation is growing in terms of relevance also due to the advancement in IoT technologies[21]. [34, 35]

Another taxonomy of the ways to generate health data is provided by Comandé and Schneider. According to the authors, there are two ways to generate health data: 1) "primary health data sources"; and 2) "secondary health data sources". Primary data are those data that since the beginning of their life have strict connection with the health care, and, therefore, can be directly used for health purposes. On the other hand, secondary data are those that do not have, per se, a strict relation with the health care. Such data could however be used for health-related purposes after aggregation and after a preliminary step of processing. [49] Another distinction is made by the OECD in its 2017 report on New Health Technologies. In this case, primary-use-data in health are considered data collected from a patient and directly used for improving his own health situation. Whereas secondary-use-data is collected from a patient, but then is used for health purposes that do not have a direct impact on the same individual. [50] Creating meaningful categorisation of the ways datasets are generated in the health system is relevant for discussing the model of governance for big data and AI in health, especially from a data protection law perspective. Indeed, one of the tenets enshrined in the GDPR relates to the limitation of purposes pursued by a data processing.

"Data acquisition" is the activity that relates to transfer of data from the entity that generates it to the entity that will process such a data. Even though data is often generated by individuals, the organisation that will process it does not always collect data directly from them. The health system is indeed a sector with extremely heterogeneous data sources. In this sense, Mehta and Pandit classify data sources into four categories. [1] The first is clinical sources, which comprehend sources within the health care system, like hospitals, laboratories, or other kinds of health care providers (such as radiology departments or diagnostic clinics). The second category are the clinical research sources, which are sources of data that still operate within the health care sector, but rather than focusing on common clinical practice, concentrate on research activities. Examples are pharmaceutical companies, research centres, or biobanks. The third category refers to claims' sources, which are composed by organisations that deal with medical reimbursement, such as insurance companies. The fourth category relates to the patient-generated data using social media, wearable devices and IoT sensors. [1]

The "storage" of big amount of data in the health domain is a process that faces many obstacles. First of all, since data sources in health are extremely heterogeneous, as said above, often the standard of the data is also different. Moreover, data, especially in public health care providers datasets, are not fully digitalised. [41, 51, 52] The life span of a dataset will impact the big data dimension of a dataset. Ideally, the more data is stored the more there is opportunity to re-use it for further processing not envisaged at the beginning. For biomedical research is key to re-use data collected for other purposes, especially for

[21] IoT stands for "Internet of Things" and indicates such technologies that revolves around physical objects connected to each other through wireless or wired networks. The health domain is experiencing an increase in the use of IoT technologies in the form of wearable devices, monitoring devices, or telemedicine. [227]

observational study. [53] However, data protection law sets limits to the storage of data for further data processing, even if the purpose is scientific research[22]. [49]

1.1.2.6 – Data Elaboration in Health: Integration and Analysis

"Integration of data" is a fundamental, though critical, process of the big data supply chain in health care. It is fundamental because this is the process that allow organisations to dispose of structured datasets that can be analysed through data analytics techniques. On the other hand, it is critical because it implies connecting and correlating data from different sources. This operation could be critical from two point of view: 1) regulatory constraint could limit the correlation and the aggregation of data from different sources (again, data protection law requirements shall be carefully evaluated in this phase); 2) the correlation of data from different sources could face problems related to different standards in the format of data. [54, 55] Popular solutions to overcome data protection restrictions in clinical research are statistical aggregation or anonymisation. [56] However, despite these processes can lead to non-personal data, i.e., falling out of the material scope of the GDPR, they must be performed following specific data protection guidelines[23].

The "analysis of data" is, as said above, probably the most important phase of big data exploitation as this is the moment when data analytics algorithms are applied to organised datasets. Therefore, this is the process of extraction of structured knowledge from datasets. This knowledge extraction could be performed through traditional algorithms or using innovative algorithms, such as those considered part of the AI domain. Namely, prediction analysis, performed through AI algorithm, is one of the most promising activities. It can reach outcome models from unseen set of inputs, introducing the data-driven decision-making process in health care[24]. Since this is the phase where AI systems are applied it will be further investigated in the next section.

1.1.2.7 – Data Interpretation in Health

In conclusion, the interpretation of health data is when health care professionals adopt decisions on the basis of the insights from the analysis of big data. Since also this final passage strictly relates to the application of AI technologies it will be investigated furthermore in the next section, which is specifically dedicated to AI in health care. However, it is worth noting here that the areas of application of data-driven decision making in health care are abundant. From the improvement of administrative processes, diagnostic, health monitoring of the population, clinical trials, precision medicine, until personalised care, or pharmaceutical domain. It is not the aim of this work to provide an exhaustive list of areas suitable for data analytics applications in the health domain. However, it is possible to find several reviews on this topic in literature. [26, 27]

1.2 – A.I. in General and in Health

1.2.1 – Definitions and Approaches to AI

[22] See the storage limitation principle pursuant to article 5(1)(e).
[23] Several data protection authorities, at national as well at EU level, have provided guidance about how to conduct data anonymisation in compliance with data protection law, see [71, 205–207, 228].
[24] See [1] for a review of data analytics techniques applied in health care.

Likewise in the big data discussion, providing a comprehensive definition of AI could be a difficult task. However, AI might be considered as the science field that tries to build intelligent agents. AI is a multidisciplinary research area, involving among others: philosophy; mathematics; neuroscience; computer science; linguistics; ethics; economics; and statistic. According to Russel and Norvig, AI can be discussed along two dimensions which generate four different approaches. The first dimension relates to the way humans understand the concept of intelligence in AI. In this regard, someone consider intelligence as the process to reach human performance, for others intelligence refers to rationality, i.e., doing the correct thing. The second dimension deals with the object of intelligence. From this perspective, there is who considers the thought process, i.e., the reasoning, as the subject matter of intelligence; otherwise, there is who believes that the object of intelligence is in the behaviour of the agent. [25] On the basis of these dimensions, four different approaches to AI are envisioned: an AI that acts humanly; an AI that thinks humanly; an AI that thinks rationally; and an AI that acts rationally. [25]

The approach that seeks to make the AI act humanly relates to the ability to pass the Turing test, i.e., to deceive an external observer and make him think he is interacting with a human intelligence instead than with an AI system. The thinking rationally approach to AI is the endeavour to make a computer work like a human brain, and therefore to make the input-output sequence correspond to human thinking. This approach is particularly connected with the cognitive science field. The approach that tries to build an AI system that think rationally is the approach based on the laws of thought and on the logic thinking. The logistic strain of AI is based on logic problem solving, using probability theories to fill the void created by conditions that are unobservable in reality. By this way, AI systems create prediction about the future. The last approach is the one that seeks to build an artificial agent able not only to think rationally, but also to act rationally. Therefore, an AI system able to do the right thing, interacting with the environment and adapting itself to achieve its goal, or at least to obtain the best possible outcome. This approach adds a layer to the laws of thought-approach, and this is because the creation of inferences is just one step of acting rationally. However, afterwards, it is necessary that the agent adopt an effective behaviour in new circumstances. Moreover, there are also situations where the creation of inferences, which is the main objective of the think-rationally approach, does not necessarily mean to act rationally. [25]

However, according to Russel and Norvig, the acting rationally approach, despite being considered as the standard model of AI, is not sustainable for the future. The reason is that such model assumes that a fully specified objective is provided by the human programmer. On the contrary, in real world scenario it is almost impossible for the human to completely define an objective to the machine. Therefore, the objective pursuits by the machine might not be perfectly aligned with original goal of the human programmer. This problem is called “value alignment problem”. The negative consequences of this misalignment may be that the machine will try to find alternative ways to reach the goal. These new ways could be unforeseen by the programmer, even violating rules of our society. The risk is having a machine that pursue its own objectives. Russel and Norvig, therefore, suggest to re-shape the model in a way that when the machine finds obstacles on its path to the goal, it is incentivised to act cautiously and try to learn more about the programmer and humans’ preferences. [25]

1.2.2 – Applications of AI in Health

As discussed in the previous Sections, AI systems are usually applied during the phases of data analytics and decision making of the big data value chain, i.e., data analysis and data interpretation. The task of AI systems is to extract knowledge from data and deploy it in a smart way in order to help professionals perform medicine-related tasks in a more efficient way. It should be borne in mind that, since a completely autonomous AI system has not been developed yet, AI can only assist health care professional but not substitute them. [9–11] Moreover, it is debatable how desirable would be to develop a completely autonomous machine, especially in the health care sector.

In 2018, the UK Department of Health and social Care conducted a survey at national level on the deployment of AI in health [57], worth of noting is the subdivision of AI applications in health in this report. The report divides AI applications according to the level of complexity of the system and its integration with other AI systems or other functions. Therefore, a scale of complexity for AI in health domain is created. At the lowest level of complexity there are the single modules of an AI system, such as neural networks, deep learning modules, machine learning algorithms. When the modules are combined with problem-solving elements, for example ML combined with an image processing module, a next level of complexity of AI is reached. Finally, the last step is where several AI systems are combined into a unique more complex system. In this case, the system is meant to support the final decision making or the performance of a clinical task, e.g., surgical robots, diagnostic decision systems, or care companion robots. Unsurprisingly, results of the survey show that most of the AI applications in health care can be classified at the lowest level of complexity, while only few applications of complex AI systems are eventually deployed in health care. [57]

A comprehensive review of the state of the art about AI applications in health care is not the purpose of this work[25]. However, AI promises to ameliorate the management of risk and resources in healthcare through a better activity of prevention. This goal should be achieved thanks to the application of predictive models, for example detecting patients with high risk factors or monitoring the population. [58] Moreover, predictive models could provide clinical decision support to the physician, improving precision medicine, i.e., shaping the best clinical treatment for an individual, thanks to combined analysis of phenotypic and genotypic data. [44, 58] Pharmaceutical processes, from discovery of drugs to monitoring activities, could also benefit from the application of big data analytic and AI technologies. [58] Quality of care may benefit from data analytics and AI through the improvement of organizational process, which could be monitored more efficiently or even in real time through big data analytics. [58] Public health is another area where AI systems should help to manage disaster events such as pandemics or outbreaks through the analysis of data from external sources such as internet query analysis or geo-localization data. [44, 58] In conclusion, of course, AI brings outstanding results in scientific research, for observational studies as well as for drugs research or clinical trials. Promising application of AI, for

[25] Nevertheless, there has been many initiatives in this sense. For example, the European Parliament has conducted a study in 2022 on the application of AI in healthcare. [229] From a general perspective, Stanford University releases a report on AI every year including also a review of the main AI applications for each sector. [59]

example, comes from the exploitation of big pool of data, such as biobanks, in order to conduct wide cohort studies. [44, 58]

To wrap up, the macro-processes most touched by AI in health care should be listed and categorised as follow: first of all, there is the unlocking of additional value from data – which could include precision medicine, research activities, or population monitoring; then there are diagnostic activities – such as early detection of disease, imaging diagnostic or preventive medicine; furthermore, the improvement of organizational processes; and finally the improvement of skills and capabilities – e.g., clinical decision support, support in surgery by robotics. [57, 59]

2 – Legal Framework

2.1 – Data Protection Law

The GDPR is the main piece of legislation that defines the data protection regulatory framework in Europe. The GDPR establishes the general rules and obligations for legal and physical persons that process personal data. Processing personal data is indeed an activity strictly regulated at European Union (EU) level. The GDPR, as it is well known, has been adopted in 2016 and has repealed the previous DPD.

2.1.1 – Scope

The material scope of the GDPR includes all the processing of personal data carried out by automated or partially automated means, or personal data not processed by automated means which though are part of a filing system or are intended to be part of it[26]. Crucial for the application of data protection law is the definition of personal data. The GDPR defines personal data as "any information relating to an identified or identifiable natural person"[27]. A natural person shall be deemed identifiable when it is possible to identify it, either directly or indirectly through the use of an identifier[28]. There are different types of information which could play the role of identifiers. Some of them are considered "direct identifiers", such as the name, the I.D. number, the driver license number. These types of identifiers alone can be used to directly identify a natural person. However, are also considered personal identifiers those data that only if combined with other information can contribute to the identification of a natural person. [60]

Upon the possibility to identify a natural person using indirect identifiers works the concept of anonymous and anonymised data. Indeed, to classify data as personal, recital 26 GDPR recommends considering all the means reasonably likely to be used by either the controller, or a third party, to link an information to a subject. "Means reasonably likely to be used" shall be considered, for instance, the costs and the time required to identify a natural person, in consideration of the technological state of the art.

These evaluations define the (blurry) boundaries between personal and non-personal data. The difficulty lies in evaluating objectively whether third parties may have the possibility to link some information to a natural person. For example, once some allegedly anonymised personal data are shared to an indefinite number of third parties, it could become

[26] Article 2(1) GDPR.
[27] Article 4(1) GDPR.
[28] Recital 26, GDPR and article 4(1) GDPR.

difficult to carry out such evaluations. In this perspective, a data controller should evaluate whether it is possible for other data controllers to re-identify individuals using personal data that the controller itself has disclosed. However, there is often a lack of knowledge about what additional information third parties possess, which make the reidentification risk impossible to be evaluated in absolute and objective terms. [61, 62]

Anonymisation is a relevant subject in health care and clinical research domains. Because it might allow a data controller to fall out of the material scope of the GDPR. Therefore, anonymisation may allow data controllers to process data, especially for research and statistical purposes, without the burden of GDPR requirements. However, anonymisation finds its limits in the trade-off between demonstrating the impossibility to re-identify data subject from anonymised dataset and the necessity to maintain the value of the dataset. Indeed, if too much information is stripped away in the endeavour of anonymising the dataset in absolute way, the data could become useless in terms of scientific value. [63]

Scholars have discussed the concept of anonymisation and the risk of re-identification [64–66]. In particular, some empirical studies have shown the weakness of some anonymisation techniques in a big data context. [67–70] The Working Party Article 29 (WP29) [71] has issued an Opinion about anonymisation techniques which constitute the main reference today about how to perform anonymisation in compliance with data protection law requirements. The opinion at issue, however, if read in conjunction with the disposition of the GDPR, support the interpretation that the risk of re-identification from anonymised data does not have to be zero. Therefore, an absolute approach is not possible to be enforced. This interpretation is the mainstream opinion also among scholars, though zero-risk approaches interpretations have been taken into consideration in some works. [64]

Pseudonymisation is another widespread technique in health and medical domain. In this case, data never falls out of the material scope of the GDPR. The aim is to separate data from identifiers but not in an irreversible way. The goal of pseudonymisation is not being a technique for bringing data processing out of the material scope of the GDPR, but rather to protect data as a security measure. Both pseudonymisation and anonymisation aim at protecting data subject's rights and freedoms, though in different ways. On the one hand, anonymisation might be attractive for data controllers because it brings data outside the GDPR scope, on the other hand it conceals risks for data subjects as well as for the controller itself. Indeed, third parties could re-identify data subjects with detrimental effects for individuals. For the same reason, data controllers could face fines for not anonymising data in a proper way. Moreover, as said before, stripping away an excessive amount of information from data could make anonymisation not efficient for data controllers. [72]

Concerning the geographical scope of the GDPR, the EU legislator has tried to enlarge the application of the data protection law even outside the EU. All data controllers or processors who process personal data in the "context of the activities of an establishment" in EU are subject to the GDPR, even if the data processing does not take place physically in the EU[29]. Moreover, the GDPR applies even when the processing is not in the context of activities of an establishment placed in the EU, but the controller provides services or goods to data subjects who are in the EU. The GDPR applies also if a data controller or processor monitors the behaviour of individuals who are located in one of the MS[30].

[29] Article 3(2) GDPR.
[30] Article 4(2) GDPR.

2.1.2 – Principles: Focus on Accountability

The principles of EU data protection law are listed in article 5 GDPR[31]. The general rationale is that data controllers shall process the minimum amount of personal data for reaching purposes that are defined in advance in a transparent manner and supported by a proper legal basis. Moreover, data is supposed to be collected and used in a secure and accurate way and stored as far as they are useful for reaching the purpose identified. Afterwards, personal data shall be deleted or anonymised.

The last principle of article 5 GDPR is the accountability principle. Data controllers and data processors, pursuant to this tenet, are supposed to demonstrate, using appropriate technical and organisational measures, the compliance with all the other principles. This principle defines the governance approach of the regulatory measure. The accountability principle is crucial when discussing the governance and regulatory choice that shapes the GDPR's functioning. Indeed, it could be argued that through this principle (and the risk-based approach) the EU legislator has enshrined, to some extent, a co-regulatory approach into data protection law. [73] The GDPR defines general principles and objectives, but the ways to reach such objectives is a choice left to the addressee of the law, i.e., data controllers or data processors. Plus, in order to reach compliance, a set of tools are provided into the regulation, such as, codes of conduct, certification, risk analysis, DPIA. Some of these tools can be defined as co-regulatory tools because are elaborated by private stakeholders but shall be approved by a DPA. They are CoCs and CMs and are the object of the discussion in this work.

Apart from accountability, the other principles that guide data protection law approach are not a novelty. Indeed, the GDPR has not changed the basic principles already enshrined into the DPD. For example, the purpose limitation principle or the data minimization principle, as well as the storage limitation principle, where already part of the DPD. Such principles have been reasoned and developed when the big data phenomenon was still unknown. The DPD was the result of discussions developed decades before the adoption of the DPD itself in the 1995. Back then, the paradigm of data processing was indeed a "small data" one. [18] The GDPR, on the other hand, is expected to cope with the issues brought by big data, which seems to be even more problematic in the specific field of digital health. [74] However, as said above, the GDPR has not abandoned most of the principles of the DPD. Criticisms have been raised in literature for an allegedly lack of long-term vision in the decision of maintaining many principles that do not fit well in big data context. [12, 75, 76] In this sense, especially the principles of purpose and storage limitation and data minimization are deemed to limit the spread of big data solutions.

Nevertheless, it is probably wrong to believe that the GDPR has brought no innovation in the field of data protection law. On the contrary, the GDPR tries to enable big data processing, while preserving data subjects' rights and freedoms. [77, 78] The goal is to make big data an opportunity for the EU internal market, as it is clear looking at the effort of the EC in building up a digital single market. [13] The GDPR seems to open the door to big data, somehow, in the wording of some principles, if compared to their previous version as stated into the DPD. [18] Moreover, the GDPR, introducing the accountability principle

[31] Article 5 GDPR includes the following principles: "lawfulness, fairness and transparency"; "purpose limitation"; "data minimisation"; "accuracy"; "storage limitation"; "integrity and confidentiality"; "accountability".

and the risk-based approach, is trying to move on from the consent and notice-based approach [79] typical of the DPD. [18] The choice is due to the fact that the complexity of data processing and their collective dimension do not allow individuals to adopt conscious decisions about their personal data [20].

The central role in the GDPR is played by the accountability principle which has the purpose to relieve data subjects from the burden of understanding and evaluating the risks for their rights and freedoms. On the other hand, there is a duty upon organisations to evaluate the risks for data subjects' rights and freedom and act in compliance with such evaluation. This approach is not only meant to avoid unconscious actions by the data subjects concerning the sharing of their data, but also to allow a wider and easier use of data by companies, as long as they are in compliance with the accountability principle. In this sense, the GDPR has taken a step towards to a regulatory framework that embrace big data use and re-use, especially in light of AI technologies applications. [18] Indeed, it is possible to argue that organisations have now more room to process data without notice and authorization duties toward DPAs and rigid and formalistic compliance activities. However, the accountability principle is not free from side effects. The cost that an organisation faces for compliance are inevitably raised up by such an approach. Moreover, legal uncertainty affects many general provisions the application of which is not straightforward and that can be interpreted in different ways.

In addition to what have been said, it seems that the GDPR lacks a real harmonization function among MSs, this is due to the room left to national legislators in defining aspects of the GDPR's principles implementation. [80] The risk is to have a fragmented implementation of GDPR among MSs. It shall be noted that the aspects that are especially prone to be object of additional implementation rules from MSs are indeed those that deal with big data processing for statistical and research purposes. The health sector is rich of these data processing, because of the great opportunities brought by big data and AI in this domain, as it has been discussed in the previous Section 1.1 of this chapter.

In conclusion, it is possible to argue that the GDPR has not perfectly balanced the protection of data subjects' rights and freedom with some big data and AI needs. For sure it leaves many grey areas of difficult interpretation, as well as principles that are difficult to implement by organisations. Nevertheless, the GDPR has for sure modernised data protection law in EU, creating a regulatory model that today is looked at as a model by regulators around the world[32]. What it is missing for a proper implementation of data protection law, especially in critical areas, such as big data and AI in health domain, is a good activity in shaping meta-level rules. Such meta-level rules should support the interpretation costs for organisations stemming from the accountability and risk-based approach. These meta-rules shall either be provided by DPAs, during their interpretation activities, or by the development of co-regulatory compliance tools. It shall be reminded that the EDPB and the EDPS are called to provide interpretation on aspects of difficult application issuing opinions and guidelines. These solutions operate, however, on a very high level of abstraction. On a more practical level, the GDPR provides the opportunity to develop tools of compliance such as CoCs and CMs. These tools shall be used in order to fill the gap between abstract rules and technical and organisational measures.

[32] See, for example, [230, 231].

2.1.3 – A Risk-Based Approach

Another important element of the data protection legal framework is the role played by the concept risk. The GDPR is said to be a risk-based regulation, i.e., the requirements are not always the same for every situation. They change according to inherent risks of the specific data processing operations at stake. For this reason, compliance can vary, in terms of heaviness of requirements, according to the risk level enshrined into the specific data processing at stake. [81–85] This approach is in the first place meant to adjust the burden of GDPR requirements to the specific situation and context. On the one hand, this is done in order to avoid excessive regulatory duty for situations that are not carrier of high risk. Moreover, the risk-based approach is complementary to the accountability principle. In this sense, data controllers are asked to demonstrate that their assessments of the risk and the respective mitigation measures adopted were appropriate. [77, 78] However, the notion of risk as enshrined into the GDPR, and its assessment, are not straightforward. The risk-based approach of the GDPR is indeed complemented by the more traditional right-based approach. Data protection is indeed a fundamental right recognised by article 8 of the EU Charter of Fundamental Rights (ECFR).

It can be argued that the GDPR has included the risk evaluation as an element of data protection in response to the complexities brought by recent technological evolutions. More specifically, protection of individuals' rights and freedoms seem to be ensured no more through "tick boxes" compliance operations, as it was often the case under the DPD. On the contrary, the GDPR has introduced the concept of scalability in compliance activities. Such scalability means that, given a minimum amount of protection ensured to each data processing, the requirements of data protection law increase at the increasing of the risk at stake. However, the WP29 states that the risk-based approach in data protection law cannot be an excuse for lowering the protection to personal data. On the other hand, the right to protection of personal data remains a fundamental right in EU legal framework. [86] To wrap up, the GDPR continues to embrace the right based approach, though adopting a risk-based approach when compliance matters come up.

The relationship between these two approaches in the GDPR can be explained as the endeavour of the EU legislator to adapt the data protection law principles to modern data processing. Indeed, as also argued in the present work, DPD was failing in ensuring protection to individuals in big data processing because some data protection principles clash with the features of big data phenomena. [12, 49] Nonetheless, the EU legislator has chosen to confirm most of the principles of the DPD into GDPR, though modifying the way data controllers have to comply with them. The real innovation in GDPR approach is indeed about compliance and its management through the risk analysis. Compliance is no longer a matter of "yes or no", it is rather a matter of "how much". [82] The "scalability" of data protection principle is confirmed by the WP29, which states that "all controllers must act in compliance with the law, though this can be done in a scalable manner". [86] In other words, the GDPR has introduced a new mechanism for reaching data protection compliance, but the principles of data protection law have not remarkably changed. Nevertheless, the GDPR is not clear as regards the practical functioning of the risk-based approach.

The concept of risk and risk analysis have a long history with wide application in many domains. [87, 88] Risk is essentially "a technique for creating knowledge and certainty about future events that are uncertain by definition". [88] On the other hand, risk analysis

is the methodology that is supposed to give concrete meaning to the concept of risk helping organisations to adopt their decisions. [83] At international level, definitions and concepts about risk and risk analysis can be found in the international standard ISO 31000:2018[33]. [89] Risk analysis is usually divided into two different steps: 1) risk assessment; 2) risk management. The former is meant to evaluate, in terms of severity and likelihood, whether there is a risk at all; the latter, on the other hand, is supposed to determine if the risk at stake is too high to be undertaken. More specifically, risk assessment is composed by the risk criteria and by the risk identification. [82] Risk criteria indicate the types of feared damages and the methodology to evaluate the risk-level. Risk identification is the process that should identify, recognize and describe the risks. [82]

Scholars have discussed how this concept of risk analysis operate in the risk-based approach of GDPR. In particular, the discussion can be carried out under two different perspectives: 1) how the risk-based approach is embedded into the GDPR; 2) how risk analysis shall be conducted in the framework of the GDPR. Under the first perspective, it is possible to envisage a parallel between the risk assessment step and the requirements in article 6(1) GDPR. This article requires data controller to detect a proper legal basis to legitimate data processing, and therefore understand whether there will be a data processing or not. In the same vein, the risk assessment step is meant to understand whether a risk exists or not. On the other hand, art. 5 GDPR, which enshrines data protection principles that protect data subject's rights and freedom, constitute the risk management procedure. Indeed, as data management is meant to reduce the harms stemming from risk through mitigation measures, article 5 establishes the measures that should prevent data processing from harming data subjects. [82] However, it shall be noted that the "mitigation measures" provided by article 5 GDPR, i.e., the data protection principles, are always the same for each data processing. On the other hand, in risk analysis, the mitigation measures vary on a case-by-case evaluation. This aspect could appear in contrast with the traditional approach of risk analysis, nevertheless it shall be borne in mind that risk in GDPR is a compliance risk. Therefore, what is subject to a risk-based approach is how to reach compliance with the principles, i.e., in GDPR risk is a matter of compliance. It means that whereas the data protection principles are always the same (because of the right-based nature of GDPR) the technical and organisational measures used by data controllers to fulfil the principles can vary in nature and intensity. [81]

Moving to discussing how a risk analysis shall be conducted in a more practical way, articles 32 and 35 GDPR are key to understand the risk analysis procedures in GDPR. The GDPR envisages a unique double-level risk analysis, the first one is enshrined in articles 32 and 24 GDPR. Such articles require data controller to evaluate the level of risk in each data processing and on the basis of such evaluation (risk assessment) to adopt adequate technical and organisational measures (risk management). The second level is, on the other hand, established in article 35 GDPR. Article 35 imposes to carry out a Data Protection Impact Assessment (DPIA) on data processing that present a high level of risk. [20] The difference between the two assessments, is that the first one is a general risk assessment that shall be performed on every data processing the data controller decides to carry out.

[33] In ISO 31000:2018 risk is defined as "effect of uncertainty on objectives". Moreover, it is specified that an "effect is a deviation from the expected. It can be positive, negative or both, and can address, create or result in opportunities and threats". Therefore, the traditional theory classifies a risk either as a positive or a negative event.

The second one is a more detailed and strict procedure concerning only some data processing. The assessment in articles 24 and 32 is meant to identify the level of risk that a data processing could lead to three specific events: disclosure of personal data to unauthorized third parties; unauthorized modification of personal data; or unauthorized cancellation of personal data. After such analysis, an overall level of risk is to be assigned to each data processing. In case the overall risk mentioned above happens to be high, a DPIA shall be also carried out. The DPIA is a broader analysis concerning other elements of data processing, besides the three risks factors of article 32. It shall also be noted that art. 35.2 identifies itself some situations where the risk of data processing is by default considered high. Therefore, in conclusion, there are two ways that can lead a data processing to necessitate a DPIA: the first one is if after the risk analysis according to articles 32 and 24 the data controller realises that there is a high risk; the second one is when the data processing falls in one of the situations considered as high risk by default by article 35[34].

DPIA is an extended analysis of the risks enshrined into a data processing. DPIA procedure stems from previous experiences either in data protection law or in other legal domains. Indeed, the WP29, even before GDPR adoption, had transposed impact assessments from other industry experiences into some specific data processing context, such as RFID and smart grid. [90, 91] Moreover, the use of impact assessments and risk-based approach found application already in other sectors, such environmental regulation. [85] The choice to transpose this approach into data protection law has not been done just by the GDPR, but it is a process started already under the DPD. However, it is under the GDPR that the risk-based approach has found its wider application and functionality.

As already noted, the implementation of the risk-based approach is not always straightforward. Doubts are mainly due to the relationship between the concept of risk in GDPR and risk analysis as traditionally intended, as well as between the risk-based approach and the right based approach. Indeed, as noted by Van Dijk et al, the relationship between the concept of risk and legal rights changes according to the context and the political and social pressure behind the sector at stake. Therefore, sometimes the risk/right relationship is envisaged as a trade-off, where the higher the risk the more the rights can be restricted (risk vs rights perspective). This is the case, for example of risks for public or national security that authorise the restriction of individual rights to privacy and data protection. [85] In other occasions, the rights are seen as a risk for organisations, i.e., rights are sources of risk. [85] When an organisation adopts a decision at corporate level, it faces risk to violate rights of third parties. In the domain of personal data processing, data controllers might violate data protection requirements; such a violation could lead to sanctions by DPAs or to court cases. [92] Under this perspective, risks are seen in light of potential detrimental effects upon companies' assets. This approach, which seems to be partially followed in some DPIA tools (e.g., CNIL, ENISA), [93–95] is more quantitative than qualitative. It focuses on the likelihood that risk consequences concretise, rather than on qualitative analysis of the nature

[34] Situations considered by default as high risk are listed by article 35(2) GDPR and include data processing that consist of
"(a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person; (b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or (c) a systematic monitoring of a publicly accessible area on a large scale.".

of the risk. In other words, the legal analysis is secondary to the endeavour of making the data protection risk as something that is mathematically quantifiable. [85]

The risk-based approach in GDPR represents the endeavour to build a legal framework that anticipate the harm rather than to remediate to it. The traditional legal practice is on the exact opposite though. Indeed, the legal reasoning has always elaborated the legal solutions on the basis of past experiences and alleged legal breaches. On the other hand, risks analysis are practices through which organisations try to forecast future consequences of their behaviours and the possible impacts. [89] Merging the concepts of risk and rights leads to a result that do not fit neither in the traditional legal thinking, nor in the classic risk analysis. Critics have been raised about merging risk analysis techniques with legal reasoning. For example, it is not sure to what extent is possible to analyse legal issues with the objectivistic approach of risk analysis. Such an approach entails also a further doubt, i.e., whether it is possible to use quantitative analysis to classify legal events and their consequences. Indeed, risk assessment procedures suggested by the main DPIA schemes (e.g., WP29, CNIL, ICO, ENISA) address the risk for rights and freedoms of individuals adopting a "probability x severity" calculation. [93, 94] The evaluation of the impact that a feared event can have on data subject's rights and freedoms requires a legal analysis. In this case, the consequences on data subjects' legal sphere stemming from the infringement of data protection law requirements shall be evaluated. These risk/legal analyses are usually performed by a team of experts without involving the individuals whose rights would be harmed. Even though the GDPR recommend involving data subject's representative when performing DPIA[35], this is not a strong requirement, it seems to be rather conditional to an appropriateness evaluation upon the data controller itself.

As a concluding remark, it is possible to say that data protection law is a unicum in the landscape of regulation that follow a co-regulatory approach. The GDPR combines precise requirements with broad principles and objectives. As said above, such requirements and principles shall be complemented by technical and organisational measures chosen by the data controller or processor following a risk-based approach. In this perspective, data controller and processor are supposed to demonstrate their compliance with GDPR requirements, according to the accountability principle. However, the controversial point is that the GDPR does not provide a solid framework of compliance mechanisms to data controllers or processor for reaching compliance in an easy way. On the contrary, other types of regulations, such as the regulations under the New Legislative Framework (NLF)[36] revolves around a sound system that manufacturers and providers of products can use for reaching and demonstrating compliance. The NLF, indeed, relies on the harmonised standards, conformity assessment and notified body as instruments to ensure the safety and the conformity of risky products with the EU regulations. On the other hand, the two compliance tools under the GDPR, i.e., CoCs and CMs, are not enough developed yet. For this reason, it is difficult to create a solid framework for compliance assessment. Additionally,

[35] Article 35(9) GDPR.

[36] The NLF is package of regulatory measures adopted at EU to set a legal framework for placing on the market some products ensuring their safety and strong internal market conditions. Namely the NLF set a regulatory framework for market surveillance of risky products, transparent accreditation of conformity assessment body, quality of products' conformity assessment, the use the CE marking. See https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en (last access: January 2023).

the co-regulatory instruments, as written in the GDPR, probably present some flaws that could hinder the spread of such instruments. These aspects will be further discussed in Chapter 2.

2.2 – Other Legal Frameworks Involved

Other legal frameworks are involved when deploying technologies that works on big data analytics. Data protection law is not the only piece of legislation that regulate the deployment of big data and AI in health care. The framework is wider and organisations willing to deploy such systems shall take into account also other requirements. Some are included in regulatory measures already in force, such as the Medical Device Regulation (MDR)[37], others are part of forthcoming legislative acts like the Artificial Intelligence Act proposal (AIA proposal or AI Act proposal)[38]. These measures contribute to the creation of a complex regulatory framework which pursue different aims. In the rest of this section, it will be shortly analysed the AIA proposal. The AIA will be part of the NLF, which is legal framework including also rules for certifying compliance with a set of legal requirements[39]. Moreover, the instrument foresees a strong interplay also with the new measures meant to incentivise the sharing of data among EU, i.e., DGA and the EHDS.

2.2.1 – AIA Proposal and its Interplay with Data Protection Law

The EU Commission has issued a white paper on AI in 2020 [96] setting the ground for the adoption of measures to regulate the deployment of AI technologies to be placed on the EU market. Following the adoption of the white paper, a High-Level Expert Group on Artificial Intelligence (AI HLEG) has been nominated for establishing ethical and legal principle AI systems should adhere with. Such principles are indeed supposed to help developers elaborate AI systems that respect fundamental rights and humankind in general. [97, 98] After several years of discussion, the EC has eventually issued a proposal for the adoption of a regulation for the harmonisation of rules that deal with AI, i.e., the so-called AIA proposal or AI Act proposal. [99, 100]

The AI act will be a horizontal legislative measure, which classifies AI systems according to risks brought by the AI system itself. The AI act distinguishes between prohibited AI systems, which are in general banned from the EU market, and high-risk AI systems. This latter category shall respect a set of requirements during their development enshrined in Title III, Chapter 2 of the AI act proposal. Moreover, the AI act establishes some transparency requirements for certain types of AI systems that could be classified as "low-risk" AI systems, such as chatbots or deep fakes. Finally, the AI act also aims to incentivise the application of high-risk requirements to AI systems that do not fall into such a category, especially through the use of codes of conduct[40].

---

[37] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.
[38] 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.
[39] See above, Section 2.1.3.
[40] Article 69 AI Act proposal.

The high-risk category of AI system is the most interesting, as it will include the majority of AI systems deployed in industry, including the medical field. [101] Namely, there are two ways an AI system falls into the category of high-risk. The first one is if the AI system is used for one of the purposes indicated in Annex III of the AI Act. For example, according to Annex III, are considered high risk AI systems used for "biometric identification and categorisation of natural persons" or for recruiting purposes of possible candidates for a job position, as well as those for assigning promotion and for taking decision related to work relationships. The second way to be included into the high-risk category is that the AI system is a safety component of a product, or a product itself, that is regulated under the NLF, and that this product is subject to a third-party conformity assessment. This will be the case of many AI systems included in medical devices. [101]

The official objectives of the AI act are ensuring safety of AI systems and respect existing law on fundamental rights and Union values, ensure legal certainty to facilitate investment and innovation in AI, and therefore to facilitate the development of the single market[41]. [102] It is clear that the AI act is a market-oriented provision, with the main aim to foster the competitivity of EU companies in the sector of AI development. This purpose is also clear when looking at legal basis of the regulation, which is article 114 TFEU, though paired with art 16 TFEU.

The inclusion of AI based products and services into the NLF is a clear indicator of the intention of the EC. The intention is, first and foremost, the strengthening of the EU market for this products and services. [103] Nevertheless, the use of harmonised standards [104], which are at the core of the NLF functioning, as well as the notified bodies and the conformity assessment procedure, could have impact also on other aspects other than market conformity. [105] Indeed, the explanatory memorandum highlights how harmonised standards are instrumental in reaching high level of compliance with requirements of high-quality data, accuracy, human oversight, and all the other rules of Title III, Chapter 2 of the AIA proposal. Having a high level of compliance in this respect permits the protection of fundamental rights and safety of individuals that interact with the AI system[42].

However, the AI act, in pursuing its objectives, inevitably will deal with personal data. For this reason, a significant interplay with the GDPR is envisaged. This perspective is clear looking at the requirements that high-risk AI system will have to comply with, according to Title III, Chapter 2 of the AI act proposal. Especially, article 10 of the AIA proposal will set the conditions for data-related requirements for deploying high-risk AI systems on the EU market. Such an article, named "data and data governance", will indeed set conditions in terms of data collection and elaboration activities in AI systems. The aim is to reach high level of data quality within the training, validation and testing datasets. Particularly interesting is the possibility provided by article 10(5) to process personal data, even special categories under article 9(1), GDPR for de-biasing activities. The AI act proposal constitutes a legal basis for processing such categories of personal data, the processing of which is in general prohibited by the GDPR unless a proper legal basis is detected. In this sense the AI Act would introduce an exemption to the GDPR. [102]

[41] See explanatory memorandum of the AI act proposal.
[42] *Idem*, page 7.

3 – Policies for Big Data Sharing

At the European level, several strategies and governance initiatives have been undertaken in recent years concerning big data and AI, some of them specifically focusing on healthcare and health data. The objective common to all these strategies is to enhance the sharing of personal and non-personal data within Europe to foster the use and deployment of innovative technologies. Indeed, Europe is behind other countries for the exploitation of big data and the development of AI technologies[43]. In the following Sections the main policies and governance initiatives adopted by the EU for big data sharing will be presented.

3.1 – The EU Strategies on Big Data, AI and Health Data

For several years, the EU has been delivering strategies to better manage and exploit big data and AI technologies. [106] The distinctive factors of the EU approach, compared to other countries' strategies, has always been the protection of individuals' rights while exploiting big data solutions, especially those of privacy and data protection. Therefore, at least in theory, the EU has never waived the protection of individuals' rights to get economic goals in the field of big data analytics and AI technologies. This pattern is in line with the approach of the Council of Europe (CoE), i.e., an approach that considers data protection as a fundamental right. Every initiative and strategy promoted at the EU level about big data and AI, therefore, states that data protection shall be at the centre of the EU's vision and approach towards such subjects.

It has been said that data protection law in EU has its main piece of legislation in the GDPR, which sets the general rules for the processing of personal data. Additionally, sector-specific directives and regulations rule particular areas of data protection law, for example data processing carried out by EU institutions[44], or data processing carried out for law enforcement purposes[45]. Moreover, some legislative acts discipline complementary fields, such as the E-privacy directive[46], which set up rules for data protection in the domain of telecommunications.

Other fields of EU law, though not setting data protection rules, are strictly related to data protection law. For example, the free flow data regulation[47] or the Open data

[43] For example, according to the AI Index Report 2022, elaborated by Stanford University, [59] the US is the leading country in terms of investments in AI private companies, followed at large distance by China.

[44] Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC.

[45] 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.

[46] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

[47] Regulation (Eu) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union.

directive[48]. All these provisions stem from a process started after the Treaty of Lisbon, which seeks to set out policies and strategies in order to improve the digital economy in Europe while, at the same time, protecting digital rights of EU citizens. The first policy enacted by the EU Commission (EC) from this point of view was the first digital agenda for the decade 2010-2020. This programmatic document identified for the first time a key role for ICT services in the economic position that the EU is to cover in the next years. The goal of the first digital agenda was to set up rules to enable EU citizens and business to get access to digital services easier and at lower prices. Moreover, the digital agenda was meant to develop a stronger set of rules to protect citizens rights threatened by the misuse of technologies. [13] From the first digital agenda, it stems the process to renew the data protection legal framework, back then regulated by the DPD. The result of this process has been the adoption in 2016 of the GDPR[49].

In 2020, a new the document called "shaping Europe digital future" and the communication "2030 Digital Compass: The European way for the Digital Decade" created the second digital agenda, which sets a series of objectives for the period 2020-2030. The Communication "shaping Europe's digital future" pursues (at least) three key objectives. Such objectives are: a) to regulate some innovative technologies, in order to make them work for people; b) to develop a fairer and more competitive digital and data economy that benefits the whole society; and, finally, c) to use data technologies to build an open, democratic society. [107] In order to achieve the first objective, i.e., to make technologies work for people, the white paper on AI has been adopted. The white paper on AI was the first step for the adoption of the current EC AIA proposal[50]. The key actions adopted to pursue the second objective – i.e., a fairer and more competitive economy of data technology – are two follow-up strategies: the European data strategy; and the digital services act package.

The European data strategy [14] has led to the adoption of two important proposals by the EC. Namely, the DGA and the Data Act (hereinafter DA)[51]. On the other hand, the revision of the digital services act package has led to the adoption of two new legislative measures: the Digital Services Act (hereinafter DSA)[52] and the Digital Markets Act (hereinafter DMA)[53]. [106] The analysis of the DA, DSA and DMA does not fall within the scope of this work although.

The European data strategy is based on two pillars: 1) personal data protection and the free circulation of data; and 2) the exploitation of such data through data analytics techniques. These two objectives are not a novelty in the EU strategies for datafication of the European single market. Indeed, the EU was already pursuing them through the GDPR, the FFD, the Cybersecurity act and the open data Directive, for example. The specific goal of the policy is however to set up a European Union data space, to grant the flow of data across

[48] 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 (recast).
[49] See above Section 2.1.
[50] See above Section 2.2.1.
[51] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM(2022) 68 final.
[52] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act).
[53] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

sectors. However, this objective shall be pursued respecting EU values, such as data protection, consumer protection and fair competition for companies. In order to do that, the European strategy for data introduces a governance mechanism for a clear and fair access and use of data. By means of the data strategy the EU Commission tries to fill the gap between EU and other countries in terms of AI and big data analytics deployment. The goal is to set up an ecosystem where organizations, both the public and private, as well as individuals can get access to and share data. The creation of such conditions should boost the use of innovative technologies.

Nevertheless, the EU Commission is aware that every sector has its own characteristics. For this reason, a set of "European data spaces" are envisaged in the data strategy to adapt the new general data governance framework into specific domains. This should get to the building of sectorial data pools. For each sector specific regulatory initiatives are to be adopted. The first European data space to be implemented will probably be the European Health data space, for which a regulation proposal has already been issued by the EU Commission[54].

The EC strategy for data moves from the acknowledgement that there are some issues that is necessary to overcome in order to actually develop a European digital single market and compete at global level. The first of these issues is the legal fragmentation among MS that risks jeopardizing the functioning of the digital single market and of the European data spaces. From here, the necessity to develop a common set of rules at the EU level for data governance. Moreover, there are specific issues due to the lack of data availability (both in G2B, B2B, B2G and G2G relationships). This shortage leads to the impossibility to use and re-use data to exploit big data analytics but can also lead to imbalances in the market power. The new strategy for data will then try to tackle the lack of data interoperability or data quality, as well as difficulties for individuals to exercise their rights, lack of skills and data literacy, and cybersecurity issues. [14]

Having in mind these issues, the data strategy, which is developed according to the better regulation principles [108], aims at setting up a cross-sectoral governance framework for enhancing the access and use of data. Building upon this general data governance framework, a system of common European data spaces should then set sectorial rules. To make the system work, then a set of investments in EU capabilities and infrastructures for hosting, processing and using data, as well as in individuals and SMEs' skills, are planned. [14]

The cross-sectorial rules are supposed to create an overarching legal framework for the sharing and re-use of data, establishing common rules among MSs and sectors. This legislative framework will regulate the governance of data in the common data spaces, ruling in which situations data can be lawfully used. The cross-border data use should be facilitated through for example setting standards and interoperability requirements. Ensuring a high quality of data, especially when data is from the public sector, and making them available for re-use is another objective of the strategy. However, also the sharing of data across different sectors shall be pursued. [109]

As said earlier, besides a cross-sectorial set of rules, a set of common European data spaces will be detected as strategic sectors where the re-use and share of data shall be given priority in order to incentive AI and big data technologies. Therefore, the creation of sectorial pools of data with ad hoc rules (though still based on the general data governance

[54] See below Section 3.1.2.

mechanism) shall boost the data agile economy in these domains. The domains of relevant public interest will be (among others): the manufacturing; the green deal data space; the mobility data space; the health data space; the financial data space; the energy data space; the agriculture data space; the public administration data space; the skills data space. [14] In the data strategy was already possible to find a sketch about the common European health data space. With regard to the building of these data spaces, it seems that the EU Commission wants to rely on both legislative and non-legislative actions. For example, it suggests relying also on CoC pursuant to article 40 GDPR for smoothing the application of data protection law. [14] The European health data space will improve the opportunities for citizens to access health data and data portability, it will then enhance the cross-border provision of health data services and products. The infrastructure and data computing capacities built through the common European data space would then support national EHR systems interoperability among MS favouring in general the cross-border health data sharing. Moreover, big data projects would be supported thanks to repositories that will collect different types of health-related data, such as EHRs, genomic data, or digital images. [110]

Concerning the AI, as discussed before, a proposal for a regulation for ensuring the safety of AI systems and harmonizing the deployment of AI systems on the EU market has already been issued.

The interplay of the new data governance mechanism as well as the AI proposal with the GDPR are remarkable, since sharing and re-using personal data is key in order to develop efficient AI systems. However, as noted above respect the AI Act, the interplay between the new data governance framework and the GDPR is not always straightforward, and some lack of consistency comes up when analysing the new measures. [16] If these instruments are not properly aligned with the GDPR requirements, companies might struggle in reaching compliance with an already complex legal framework. These aspects will be object of Section 3.2. Otherwise, in the next section will be quickly presented the two instruments that are expected to generate the most interplay with data protection law, i.e., the DGA and the EHDS.

3.1.1 – The Data Governance Act (DGA)

The Data Governance Act[55] has been recently adopted as a regulation of the Council and of the Parliament. The DGA is the first piece of legislation stemming from the data strategy. The DGA aims to increase the availability of data for stakeholders, especially in G2B data flows. A new data sharing mechanism and new data intermediaries are therefore introduced by the DGA. The overall objective pursued by the DGA is to create a secure environment where different types of data can be shared. Within these secure spaces, data can be re-used for different purposes, including scientific research. [111]

Therefore, the rationale behind the DGA is making public data available for re-use. Data that fall into the DGA material scope is data subject to right of others, namely data protection rights, IP rights, trade secrets, and commercial sensitive information[56]. Organisations that operate as data intermediaries should, on the one hand, enhance trust in individuals about such data re-use, while on the other hand help individuals exercise their rights

[55] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).
[56] Article 3 DGA.

according to GDPR. A data altruism mechanism, moreover, is introduced for enhancing the sharing of data for altruistic purposes, i.e., without remunerative goals.

Worth of note is that the aim of the legislative measure is not introducing a new right to data re-use. Otherwise, the goal is just harmonising the conditions under which data may be re-used. GDPR principles therefore are still to be respected when the re-use of data take place, this generates significant interplay between the DGA and data protection law. Finally, it can be noted that the DGA has been envisaged taking inspiration from FAIR data principles[57]. [112]

Although the DGA will be focusing on the re-use of protected public sector data, the Chapter III also aims at increasing the sharing of data in B2B and C2B relationship. This aspect should be pursued by lowering transaction costs and the creation of a notification regime. [109]

The data altruism mechanism is regulated by Chapter IV of the DGA. The mechanism, despite not new in the research domain, has been for the first time enshrined into a hard law measure. [111] The objective is to allow natural and legal persons to share their personal and non-personal data for altruistic purposes, without receiving a compensation for that. In this case, the organisation that provides a data intermediation service will be called "data altruism organisations" and will be subject to ad-hoc rules under Chapter IV of the DGA. In this situation, the purposes for which data could be shared will be limited to those that could create certain amount of benefit for the society. The DGA has introduced also some new players, such as the data sharing intermediaries, which are supposed to support the sharing of data as well as the exercise of individual rights on personal data.

The DGA will work in conjunction with sectorial rules for the use and re-use of data. These laws will further implement DGA's framework and functions in sectors considered of particular interest. The planned Common European Data Spaces, this is the name for the future sectorial regulations, are: 1) automotive; 2) Payment service providers; 3) smart metering information; 4) electricity network data; 5) intelligent transport systems; 6) environmental information; 7) spatial information; 8) health sector. It shall be noted that the European Health Data Space is already a proposal for a regulation that has been submitted by the EC[58]. Therefore, it is presumable that this is going to be the first European Common Data Space. The endeavour to regulate the use and re-use of data held by PSBs, throughout a regulation, is due to the risk of fragmentation brought by leaving room to national legislators in this domain. However, as already said, the DGA set horizontal rules that are then flexible in leaving space for sectorial rules.

The DGA is divided into nine chapter. Chapter I defines the subject matter of the regulation and provides the definitions. Namely, the article 1 defines the subject matter and the scope: the DGA lays down conditions for the re-use of certain categories of data held by PSBs, it set up a notification system and a supervision framework for data intermediaries, moreover the DGA set down the general rules for the data altruism mechanism, and finally it introduces the role of the European Data Innovation Board (EDIB). Article 1(3) of the DGA states that the new data governance framework is without prejudice to data protection law in EU. In case of contrast between DGA and data protection law, the latter shall prevail.

[57] FAIR is an acronym that stands for Findability, Accessibility, Interoperability, and Reuse. These are the principles research dataset should follow in order to enhance re-usability of datasets.
[58] See below Section 3.1.2.

Moreover, article 1(3) explicitly states that it does not create a new legal basis for the processing of personal data, nor does it affect any of the rights and obligations set out in the data protection law legal framework. A set of definitions is then provided by article 2 DGA. Among these definitions it shall be noted that a definition of data is provided, plus also a definition of data holder, data user and data intermediation service. Some of these definitions have been object of discussion at academic level, as well as among data protection authorities, for a possible lack of consistency with data protection law[59]. [16]

Chapter II of the DGA establishes the re-use mechanism for data held by PSBs that are protected by rights of others. However, as said above, this chapter does not create a right to re-use such data. Nevertheless, the DGA introduces some horizontal rules that define the conditions under which data re-use can be allowed. It means that PSBs that decides to permit the re-use of data shall be technically equipped to comply with the requirements of the chapter at stake.

The categories of data held by PSBs that fall within the material scope of the regulation are data protected on the grounds of: (a)commercial confidentiality, including business, professional and company secrets; (b) statistical confidentiality; (c) intellectual property rights of third parties; or (d) protection of personal data, as long as such data fall outside the scope of Directive (EU) 2019/1024[60].

Article 4 DGA sets a prohibition to set up exclusive arrangements or agreements pertaining the re-use of data held by public sector bodies referred to in article 3(1). Therefore, it is prohibited to PSBs to conclude agreements that aims at reducing the availability of data to third parties. However, a set of exceptions are provided by paragraph 2 to 6. Article 5, on the other hand, defines the conditions for data re-use. In this perspective, PSBs are competent to grant or refuse access for the re-use of the data categories of article 3(1). In this activity of granting and refusing access PSBs may be assisted by the some "competent bodies" identified in article 7 of the DGA. It is important that the PSBs do not set conditions for data re-use that are discriminatory. [109]

According to article 5(3) DGA, the PSBs shall ensure the protection of rights concerning the data to be re-used, e.g., data protection rights. In doing so, PSBs may require the respect of some conditions for the re-use of data. In the first place, to grant access to data only after it has been ensured data has been anonymised or "modified, aggregated or treated by any other method of disclosure control". PSBs, then, may require that data shall be accessed and re-used remotely only within a secure environment provided or controlled by the PSB itself. Finally, the PSBs may also set the condition that the access and the re-use of data take place within physical premises in which the secure environment is located if the remote access might jeopardise rights and interests of third parties. In case the PSB chooses to allow data re-users to access data only through a remote connection to the secure environment or within physical premises, it is a task of the PSBs to ensure the integrity of the technical system of the secure processing environment. On the data re-users' side, confidentiality obligations apply to datasets accessed. Therefore, it shall be prohibited to disclose any information that could jeopardise rights and freedoms of others. Moreover, data re-users shall not try to re-identify data subjects from anonymised datasets and shall adopt measures to prevent re-identification from third parties. [109]

[59] See below Section 3.2.
[60] Article 3(1) DGA.

However, if the conditions according to article 5(3) cannot be applied, because that would hinder the data re-use, article 5(6) left open another opportunity. Indeed, data could still be re-used under a legal basis of the GDPR if any of them is applicable. In case the consent of data subjects is the legal basis chosen, PSBs and the competent body shall provide assistance to data re-users in seeking consent of the data subjects, as long as this is not a disproportionate burden for the PSBs. Article 6 of DGA then allows PSBs to charge data re-users with fees for providing the access to data and other services related to it, such as anonymisation or the maintenance of the secure environments.

Each MS is asked to designate one or more competent bodies, which will probably be competent for a particular sector (such as the Health Data Access Bodies under the EHDS) to assist the PSBs in their activity to grant or refuse access to data. Such competent bodies may even be empowered to directly grant access for the re-use of data, by Union or national law. The competent bodies designated shall support the PSBs, especially from a technical point of view. For example, they shall support PSBs in ensuring a secure data processing environment, in the pseudonymisation and anonymisation of personal data, as well as in implementing PETs. They shall then support PSBs in helping da re-users to collect consent from data subjects.

Chapter III is dedicated to the trust mechanism for the sharing of data and to decreasing the costs linked to B2B and C2B data sharing. For this purpose, a notification regime for data sharing providers is introduced. Namely providers are asked to remain neutral about the data shared, i.e., data cannot be used for further purposes by data sharing providers. A competent authority designed by MS will be in charge of monitoring compliance of data sharing intermediaries with the requirements in this chapter. According to article 10 DGA, some data intermediation services shall be regulated. These services are basically the intermediation between data holders or data subjects and potential data users. Data intermediaries may be organisations that provide technical means for the sharing of such data, like the creation of platforms or databases, as well as means to data subjects for exercising their data protection rights. Data services intermediaries can also be in the form of data cooperatives.

A notification system is provided by article 11 DGA for such data intermediation services. Namely, data intermediaries service providers shall notify the competent authority and indicate what kind of data sharing services they plan to provide. On the other hand, article 12 sets down the conditions according to which data sharing intermediaries can provide their services. In terms of data protection, data intermediation services cannot use the data at their disposal for purposes other than to put them at the disposal of data users. The same tenet applies to metadata related to data intermediaries' activities, such as the date, time and geolocation. These data shall be made at the disposal of the data holders upon request. Another condition is that data intermediaries shall, if necessary, transform the format of the data to make it more interoperable and facilitate the exchange of it. The format should be in compliance with European data standards. Data intermediaries could then provide additional services and tools if requested by data holders or data subjects. These services include storage of data, curation, conversion, or even anonymisation or pseudonymisation. [109]

Each MSs is supposed to identify a competent authority that shall be in charge of the notification task for data intermediation services. The powers of these competent authorities are without prejudice to the powers of data protection authorities. Such authorities have

the task to monitor the compliance of data intermediaries with this chapter. The competent authorities ruled in this chapter shall not be competent to monitor data altruism organisations or other not-for-profit entities, as long as these organisations do not establish commercial relationships between data users and data holders or data subjects.

Chapter IV sets out the data altruism mechanism, this is a mechanism that will allow individuals, as well as companies, to share data on an altruistic basis. The chapter provides the opportunity to organisations to be registered as "Data Altruism Organisations". These organisations should therefore increase the trust in individuals or companies to share their data. Also, a new data altruism consent will be created. This type of consent allows data subjects to share their personal data as said before, but it will not replace the consent as a legal basis under GDPR[61]. Therefore, an overlapping between this consent and GDPR's one is quite predictable. [16, 111] Data altruism organisations, in order to be recognised as such, shall operate on a not-for-profit basis. Moreover, they must be registered in the national registry of recognised data altruism organisations. The requirements and the procedure to be registered are defined in articles 18 and 19 DGA. A competent authority will then evaluate if the application satisfies all the requirements and, on the basis of that, whether to accept or not the organisation as data altruism organisation. The chapter at issue then establishes some specific requirements to protect rights and freedoms of data subjects and data holders that decide to share their data with data altruism organisations. In this sense, information duties apply to data altruism organisations. They shall inform data subjects and data holders about the objective of "general interest" and, if the case, about the explicit and legitimate purpose for which personal data is to be processed.

The data altruism organisations are prohibited from processing data collected for altruistic activities for objectives other than those of general interest. Moreover, data altruism organisations are supposed to provide the tools for collecting data subjects' consent for altruistic purposes as well as the permit to process data from data holders. Data altruism organisations are then responsible for the security of data processing. [109] Article 25, the last provision of the chapter, introduces the possibility for the EC to elaborate, through implementing act, a European data altruism consent form. Such a form shall be elaborated involving also the EDPB, the EDIB and relevant stakeholders. The form should facilitate the collection of consent for data altruism purposes. Indeed, it should follow a modular approach allowing also customisation for specific sectors. However, it shall be borne in mind that where the consent concerns also personal data, it shall still be in compliance with the consent requirements of the GDPR[62].

Chapter V sets out requirements for the functioning of competent authorities designated to monitor and implement the notification framework for data-sharing and entities engaging in data altruism. Chapter VI creates the European Data Innovation Board. This body is going to be a formal expert group in charge of the task to facilitate the emergence of best practices by MS authorities. Moreover, the EDIB will support and advise the EC on governance of cross-sectorial standardisation and cross-sector standardisation requests. Article 29 DGA establishes that the EDIB shall be constituted by representatives of the competent authorities for data intermediation services and the component of authorities for the registration of data altruism organisations of MS. The EDPB, the EDPS, ENISA and the EC

[61] See below Section 3.2.
[62] See below Section 3.2.

will also be part of the EDIB. The EDIB will then be organised in three different sub-groups indicated in article 29(2) DGA.

Article 30 of the DGA defines the tasks of the EDIB. First of all, the EDIB has the task to advise and assist the EC from many points of view. For example, these advisory tasks consist in supporting the development of practices for data altruism across the EU in a consistent way. Otherwise, the advice of the EDIB could be required for the development of guidelines of cybersecurity aspects. Also, consistent practices for public sector bodies for handling requests for the re-use of data could be object of EDIB activities in assisting the EC. The EDIB will be key for developing guidelines for the forthcoming European data spaces. The EIDB shall support the creation of a framework and common standards for enhancing the sharing of data. The final aim should be to develop new products and services, boost scientific research or civil society initiatives. The EDIB also assists the EC in developing data altruism consent.

Chapter VII concerns the international access to and transfer of data. It shall be noted that the data holder, the data user, the data intermediation service provider, or the recognised data altruism organisation shall adopt technical, legal and organisational measure to prevent the international transfer or the governmental access to non-personal data outside the EU. This provision applies as long as the access is in contrast with the EU or MS laws. Articles 31(2) and 31(3) then set some exceptions to this prohibition. Chapter VIII and Chapter IX contains, respectively the exercise of delegative power by the EC and the final and transitional provision of the legal act.

3.1.2 – The Common European Health Data Space Regulation Proposal

The proposal for a Regulation on European Health Data Space (EHDS)[63] is a legislative measure setting out sectorial requirements for a data governance mechanism that should both support primary and secondary use of data in health domain. Primary use of data in health stands for the use of data to deliver health care services directly to the person the data belongs, such as surgical operations, care activities, monitoring or personalised medicine. On the other hand, the secondary use of data in health domain relates to data processing to support research activities, policy making or any other purpose without immediate effects on the individual. [50]

Therefore, the EHDS will enable individuals' control on their EHRs, as well as facilitate the use and re-use of such data for organisations. This second objective is meant to allow researchers, innovators and policy makers to use such data in secure way also protecting individuals' privacy [110]. As regards the individuals, they should become able to exercise their rights more effectively than now, such as access to data and the transmission of it, even in transborder situations. From a third point of view, the proposed EHDS should remove the barriers also to the placement on the market of products and services in health domain. Indeed, differences of standards and the lack of interoperability between health products pose barriers to producers operating in different MSs. [110] Therefore, the new data governance framework, that the EHDS should help to implement in EU, would impact electronic health data sharing in order to reach three objectives. In the first place, it should lead to better diagnosis and treatments for EU citizens. Moreover, it should allow

[63] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, COM(2022) 197 final.

policymakers to develop better policies. [113]. Finally, a single market goal should be realised thanks to the harmonisation of rules for digital health products and services. The exchange of electronic health data will concern EHRs, genomic data, patient registry, and other kinds of data. In general, the EC's objective is to improve healthcare delivery, research and innovation, policymaking, personalised medicine, and in general the secondary use of health-related data.

The EHDS, along with the DGA will also introduce a data altruism mechanism. This mechanism is supposed to be the voluntary authorisation from data holders or data subjects to share their data (even personal data) for some specific secondary uses. The data altruism mechanism will substantially be a kind of consent for the re-use of data. It shall be noted that such a data altruism mechanism will have to be co-ordinated with the rules on consent under the GDPR, where personal data are involved. The interplay between consent under the GDPR and data altruism consent has been criticised by many scholars for a lack of consistency between the two legal acts[64]. [114–116] The interplay will be even more problematic in the context of biomedical research for the well-known issues related to the use of consent as legal basis in this context.

As regards the data protection context, the EHDS builds upon the possibilities provided by the GDPR, i.e., the legal basis, for processing health data for medical diagnosis, health care treatment, or the management of health care systems and services[65]. Moreover, the use of personal data in the EHDS context should also be supported by the legal basis of the GDPR that permits the use of health data for scientific or historical research, and statistical purposes[66], as well as the public interest in the area of public health[67]. However, also from this angle, scholars, the EDPB, and the EDPS have underlined as there is no full consistency between GDPR and EHDS. [15, 16]

The EHDS, moreover, builds upon the MDR and VDR and the AI Act. According to the EC, the EHDS should support the development of medical devices and in vitro medical devices, for example harmonising requirements for EHR systems, used to store electronic health data. This action should support interoperability and portability of such systems. For example, in case a manufacturer of medical devices and high-risk AI systems declare interoperability with HER systems, they will be subject to essential requirements posed by the EHDS.

The EHDS builds upon the DGA and the DA, as long as it set up rules for the secondary use of electronic health data. We can say that EHDS is to some extent a verticalization, in the health sector, of DGA's rules for the re-use of data. The DGA, as said above, provides a horizontal framework for the re-use of data held by PSBs, without creating a right to reuse such data though. The DA, on the other hand, aims at fostering the portability of user-generated data, which can include also health data. Therefore, the EHDS integrates horizontal rules of DGA and DA, completing the legal framework in the health sector. The provisions set forth by the EHDS will regulate the exchange of electronic health data. Such rules will impact the providers of health data sharing services, imposing formats that facilitate data portability. Moreover, data sharing services will be encouraged to take part of the data altruism mechanism in the health sector. [110]

[64] See below Section 3.2.
[65] See article 9(2)(h) GDPR.
[66] See article 9(2)(j) GDPR.
[67] See article 9(2)(i) GDPR.

The twofold purpose of enhancing fundamental rights and making the EU internal market more competitive is also clear when looking at the legal basis of the EHDS proposal. The EHDS is indeed, supported by both article 114 TFEU and article 16 TFEU. EHDS should remove some obstacles from the internal market, such as those brought by legal fragmentation and lack of technical standards. Indeed, the majority of the provision in the EHDS are meant to improve the internal market functioning and the free movement of goods and services. Namely, article 114(3) TFEU sets the ground for EU actions that are supposed to ensure a high level of protection for human health while achieving harmonisation in the EU internal market. Therefore, this legal basis appropriate for regulatory initiatives in the domain of public health protection. [15] Article 16 TFEU, on the other hand, is used to support the adoption of regulations that are meant to complement the GDPR. Indeed, as noted by the EC itself, some GDPR rights cannot be implemented in practice, because of lack of interoperability, limited harmonisation, and technical standard fragmentation. In this sense, article 16 TFEU is meant to support EHDS provisions with the aim to integrate some data protection rights. For example, the right to portability is strongly limited in the health sector. Therefore, the EHDS set up some requirements that will be instrumental to foster the portability of electronic health data[68].

The interplay between EHDS and GDPR is also testified by the EC statements, in the explanatory memorandum, that the EHDS builds upon the possibility to process health data for some purposes under the GDPR. [15] Namely, the EHDS revolves around the possibility to process personal data for medical diagnosis, provision of healthcare or treatment or management of health care system and services. Moreover, the use and re-use of personal health data, under the GDPR is allowed for public interest in the area of public health, for facing cross-border threats to health, as well as for ensuring high standards of quality and safety of health and medical products. Finally, the EHDS works in the space that is left by the GDPR for scientific and historical and statistical research purposes. [15]

The proposal does not aim at harmonising the way MSs provide health services and national healthcare in general. However, the EC noted that differences at national or regional level, as well as the excessive use of specification at national level of GDPR provisions, hamper the exercise of GDPR rights from individuals. [80] In the same vein, such differences and standard fragmentation create an obstacle to the placing on the market of digital health products and services. Today, the cross-border sharing of electronic health data is still very limited, with serious limitations for researchers and policy makers. The harmonisation of data flows in healthcare should help such actors in carrying out their activities. It shall be noted that the EHDS builds upon and try to continue the work of the CBHC Directive[69]. Such a Directive, in article 14 established "the eHealth Network" with the aim to support the application of cross-border applications of patients' rights. This was indeed the first reference to eHealth in EU. The nature of the eHealth provisions in CBHC Directive were facultative in nature though. For this reason, the results stemming from the application of CBCH Directive have been considered as not satisfying. Nevertheless, this directive for the first time introduced legal provisions in support of harmonisation, introducing a common EU approach to use electronic health data. [117, 118]

[68] See in this sense recitals (11), (12), (19), (35), as well as article 10(2)(m) of the proposed EHDS.
[69] Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare.

In the study "Assessment of EU MS' rules on health data in the light of GDPR", the stakeholder vision highlighted how different legal basis under the GDPR make it difficult sharing data cross-border. Moreover, especially for the secondary uses, stakeholders suggested, already in the study on "Health data, digital health and AI in healthcare" [80], the necessity to build up a legal and governance framework for health data sharing. This framework has been suggested to revolve around the role played by data access bodies.

In terms of fundamental rights, the EHDS will have a strong interplay with GDPR each time personal data will be processed. Indeed, the sharing of and the access to personal data, that the EHDS aims to foster, will have to be in compliance with GDPR. In particular, when personal data will be used for secondary purpose, all the security measures to protect fundamental rights under the GDPR shall be applied. The GDPR provides the opportunity to re-use personal data for scientific, historical or statistical purposes. The legal basis under article 9(2)(h), (i), and (j) GDPR has been deemed suitable for data processing under the EHDS. Therefore, the EHDS itself constitutes an EU law provision that provides suitable and specific measures for protecting individuals' rights. Indeed, the abovementioned articles of the GDPR authorise personal data processing carried out on the basis of an EU or MSs law[70].

Under article 9(2)(h), the purposes of the data processing shall be "preventive or occupational medicine", or "the assessment of the working capacity of the employee", or "medical diagnosis, the provision of social care or treatment or management of health or social care". Under article 9(2)(i), the data processing shall be "necessary for reasons of public interest in the area of public health"[71]. This legal basis requires also that the EU or MS laws at stake shall provide "suitable and specific measures to safeguard rights and freedoms of data subject". Finally, article 9(2)(j) provides a legal basis for data processing "necessary for archiving purposes in the public interest", or "scientific or historical research", as well as "statistical purposes". The use of such a legal basis is subordinate to respecting the safeguards under article 89(1). Moreover, the law at EU or State level shall "be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguards the fundamental rights and the interests of the data subject". The discussion about which legal basis shall support personal data processing within the EHDS new governance framework has been object of broad debate in literature[72].

Moving to a brief analysis of the specific content of the EHDS proposal, we can see that this is an ambitious regulation with more than seventy-two articles organised throughout IX Chapters.

Chapter I defines the subject matter and the scope of the regulation, which is the provision of rules for building common standards and practices, as well as an infrastructure and a governance framework for primary and secondary uses of electronic health data. As said above, therefore, the EHDS aims at strengthening natural persons rights on their electronic health data, especially those of availability and control over personal data. The EHDS then lays down rules for placing on the market or putting into service EHR systems in the Union.

[70] Or, under article 9(2)(h) only, pursuant to a contract with a health professional.
[71] For instance, when data processing is necessary to protect individuals against cross-border threat to health. The quality and safety of health care of medical products or medical devices also fall under this legal basis.
[72] See below Section 3.2.

Rules and mechanism to foster the secondary use of electronic health data are object of the regulation as well. The building of a mandatory cross-border infrastructure for enabling the primary and secondary use electronic health data is finally provided by article 1(2)(e) the EHDS proposal.

The scope of the regulation includes, first of all, the manufacturers and suppliers of EHR systems and wellness applications to be placed on the market or put into service and the users of these products. Moreover, the regulation applies to all data controllers and processors established in the Union who process electronic health data on Union citizens, as well as data of third-country citizens who are though residents of a MS. It also applies to data controllers or processors established in third countries who are connected to or are interoperable with the MyHealth@EU, pursuant to article 12(5). Finally, fall into the scope of the EHDS all data users to whom electronic health data are made available by data holders in the UE.

Article 2 provides the definitions that apply to the regulation. In this regard, it shall be noted that EHDS transposes the definitions of the GDPR, as well as some definitions included in other Union laws. Namely, from the Directive 2011/24/EU the EHDS transpose some definition concerning the healthcare domain, such as the those of "healthcare provider" or "health professionals" or "healthcare" itself. The EHDS relies then on the definition of the DGA in relation to functioning of the data governance framework, such as for the definitions of "data altruism", or "public sector body". Moreover, some definitions of Regulation (EU) 2019/1020, Regulation (EU) 2017/745 and Regulation (EU) 910/2014 also apply to the EHDS.

Chapter II of the proposal is dedicated to the primary use of personal data. The chapter introduces new rights and mechanisms that will complement the GDPR in relation to electronic health data. Obligations, on the other hand, will be imposed to health professionals still regarding electronic health data. On the MSs side, national legislators will have to set up a "digital health authority" responsible for monitoring the new rights and mechanisms. Moreover, MSs will also have to designate a national contact point in order to enforce this chapter. In conclusion, the MyHealth@EU is identified as the infrastructure for the cross-border data sharing of electronic health data.

Chapter III contains the obligations for several economic operators, including the essential requirements (specified in the Annexes) that is necessary to comply with in order to obtain the CE marking. The conformity is self-assessed by the economic operator itself. The EU legislator has chosen not to impose a third-party assessment for EHR systems' conformity assessment. The EC has the possibility to adopt common specifications to specify the essential requirements, if needed, and facilitate the compliance procedures. A market surveillance authority shall be designated by MSs, which will be responsible for the implementation of the Chapter at stake. In conclusion Chapter III provide a system of voluntary labels for wellness applications that are interoperable with EHR systems.

Chapter IV is dedicated to the secondary use of electronic health data. In the first place, this chapter set down a set of categories of data that data holders shall make available for secondary uses, as long as the condition in the present chapter are respected. These categories include but are not limited to: EHRs; genomic data; data generated by medical devices or data generated by wellness applications; health administrative data. These data categories are indeed held by entities and bodies, either public or private, that operate in the healthcare sector, or perform research activity in this domain, as well as Union institutions,

bodies, agencies and offices. The chapter at issue, moreover, clarifies the purposes for which the electronic health data abovementioned can be re-used. The purposes explicitly indicated in the proposal are: 1) public interest in the area of health; 2) support EU bodies and institutions to carry out their task in the area of health; 3) for statistical purposes at national or EU level; 4) education or teaching activities in the healthcare sector; 5) scientific research in the health or care domains; 6) innovation and research of products and services in the health sector, or ensuring high levels of quality and safety of medicinal products and medical devices; 7) the training or testing of algorithms, including AI systems and medical devices; 8) the provision of personalised healthcare, based on health data of other persons.

It shall be noted that there are also some purposes that are explicitly prohibited under the EHDS. Such prohibited purposes include, first of all, decision making based on health data that are detrimental to the data subject itself, in the sense that such decision produces legal or comparable effects on the individual. It is also prohibited the re-use of data meant to exclude individuals or groups from insurance services. Moreover, it should not be permitted the re-use of electronic health data for targeting health professionals or health care providers for proposing marketing solutions (i.e., advertising activities). It shall be forbidden to share electronic health data with third parties not indicated in the data permit. Finally, data re-users shall not use such data for developing certain types of products that could potentially harm individuals or the society at large. Examples, in this sense are illicit drugs, alcoholic or tobacco products.

It has been said, in the previous part of this Section, that the EHDS proposal aims at building a governance framework for the re-use of certain electronic health data building up the DGA. Chapter IV of the proposal set down the rules from this perspective. At the centre of the governance framework for data re-use there will be the "health data access bodies". Each MS will have to designate one or more of these bodies, either relying on existing bodies or establishing a new one. Health data access bodies will have a number of tasks. They will first of all decide on data access applications, authorising the access to electronic health data for secondary uses, in accordance with the provisions of the Chapter at stake and with Chapter II of the DGA. They will collect, combine, prepare and disclose the data for secondary uses, on the basis of the data permit. They will have to prepare a secure processing environment where the data are put at the disposal of the data users. Health data access bodies shall then contribute to the data altruism mechanism. They shall also support the development of AI systems, and the development of harmonised standards under the AI regulation proposal. Moreover, such bodies are also supposed to collaborate with data holders in order to ensure data quality and utility label set out in the EHDS itself. Other tasks of cooperation with national and Union level institutions, as well as of facilitation, are assigned to the health data access bodies.

It shall be noted that, under the EHDS health data access bodies also have some obligations towards data subjects. Indeed, such bodies are supposed to provide information about the data processing concerning the re-use of data subjects' data. The legal basis upon which the re-use is based shall be explicitly made available to the public. The same obligation concerns the technical and organisational measures to protect data subjects' rights and freedoms. Moreover, also rights that natural persons can exercise about data re-use shall be explicitly indicated, as well as the modalities for such an exercise. Finally, the outcome of the projects for which electronic data have been re-used shall be made publicly available.

The EHDS specifies the provisions about data altruism already set up by the DGA. Namely, when a data altruism organisation process personal electronic health data shall do it within a secure processing environment in compliance with the relevant provisions of the EHDS.

About the duties of the data holders, i.e., public sector bodies that held data falling into the scope of EHDS and obliged to make them available for re-use, the following aspects shall be noted. Data holders shall cooperate with the health data access body communicating a general description of the datasets, including also data quality aspects, and providing the required data within a certain amount of time. A system of fees is included in the EHDS as a compensation for the health data access body and data holders' activities. Health data access bodies shall monitor the compliance with requirements of EHDS from data holders and data users. Health data access bodies can also impose penalties when they find data holders or data users not compliant.

Another section of the chapter at issue provides some rules about the conditions under which the data shall be made available to data users. In the first place there is a data minimisation requirement. In other words, the health data access body shall ensure that the data users have access only to the information necessary to pursue the aims indicated in the data application and authorised in the data permit. Moreover, the data shall be provided in an anonymised format, as long as the purposes can be pursued with anonymised data. If the purpose cannot be reached out using anonymised data, then the data can be provided in a pseudonymised form. Anyway, the data access body shall retain the additional information necessary to re-identify data subjects and not share it with anyone else. On the data users there is a duty to not try to re-identify data subjects from pseudonymised datasets.

It is interesting to note that if the data user wants to obtain pseudonymised data instead anonymised data it shall indicate in the application addressed to the health data access body the legal basis for data processing under article 6(1) GDPR. Otherwise, another option for data users is to present a "data request" instead a "data access" to the health data access body. A data request does not entail a real access to electronic health data, but it constitutes in a submission of a specific request for having anonymised statistical results from data. If a data user needs access to data held by just one data holder in a single MS, he can choose to directly address the request to the single data holder. In this case, the data holder and the data user shall be deemed joint data controllers. In all the other cases, i.e., when the data users get access to data through the health data access body, the data user and the health data access body should be deemed as joint data controllers, as explicitly provided by article 51 of the proposed EHDS.

In order to foster the secondary use of electronic health data in a cross-border context, each MS is asked to designate a national contact point. Such a contact point will therefore be responsible for making electronic health data available for secondary use. Reasonably the national contact point will be the health data access body, or the coordinator health data access body where there are more than one. The national contact point shall then be a participant of the cross-border infrastructure for secondary use of electronic health data, i.e., HealthData@EU. Participating in HealthData@EU shall also be EU institutions, bodies, offices and agencies involved in research, health policy or analysis, as well as research infrastructures. Even third countries or international organisations may become part of the infrastructure, as long as they comply with the rules set down in this chapter and provide access to data users located in the Union, imposing the same conditions of health data

access body. The EC on the other hand shall develop, deploy and operate a platform for HealthData@EU by providing also information technology services.

Finally, Chapter IV of the EHDS proposal provide some requirements of data quality and data utility, although using a voluntary label. Namely, data holders when making datasets available to health data access bodies may also provide such a label. The label indicates compliance with some data quality and utility aspects. For example, the label should indicate the technical quality, the completeness, the accuracy, the validity of data. The labels are then supposed to demonstrate compliance also with other aspects, such as: the standards used; data quality management processes, including biases; the representativity of the population sampled; time between the collection and the addition of data to the dataset; information about eventual enrichments on the datasets.

Chapter V of EHDS provides measures to promote capacity building by MS. Chapter VI creates the European Health Data Space Board, which will have the task to facilitate the cooperation between digital health authorities and health data access bodies. Moreover, the chapter includes also provisions related to the joint-controllership groups for EU infrastructures. Chapter VII contains the delegated acts by the EC on the EHDS. Chapter VIII and IX set up the penalties and cooperation as well as the final provisions.

To wrap up, the EHDS will be revolving around three pillars: data governance framework and rules for data exchange; data quality; and data infrastructure that grant interoperability. [110] The data governance framework will build upon the cross-sectorial rules of the DGA. It will aim at enhancing the use and the re-use of personal and non-personal data for both primary and secondary uses in health domain through either legislative or non-legislative measures. As highlighted by the EC study on health data in light of the GDPR, the uneven implementation and interpretation of the GDPR by MSs creates legal uncertainty. Such an uncertainty raises a barrier to secondary use of electronic health data. [80] The data quality pillar, on the other hand, will be aimed at ensuring the quality of data and the quality of the sources of data (EHRs, registries, IT tools) and that such sources shall be interoperable. This basically means ensuring a semantic and technical interoperability. [118] Finally, the third pillar will consist in investments of the EU in infrastructures and technology needed to reach the objectives identified in the other two pillars. Such investments could also be built upon and scale up existing initiatives, such as the eHealth Service infrastructure, the European Reference Networks or the Genomics Projects. [110]

3.2 – Concerns on the Interplay Between the GDPR, the DGA and the Proposed EHDS

Scholars and data protection authorities raised concerns about the lack of consistency between the DGA, the EHDS proposal, and the GDPR. It has been noted throughout the chapter how the difficulty in implementing data protection law in big data context is having a limiting effect on the sharing and re-use of data. This lack of clarity is consequently hampering the exploitation of these technologies [114, 119, 120] In principle, the DGA and EHDS aims at remedying to such limits and in general increasing the sharing of data across EU. Against this background, the re-use of data should be fostered thanks to new data governance framework. However, in the attempt of doing so, the DGA and EHDS clash with some GDPR's requirements and principles. [15, 16] A sort of compliance paradox seems to affect the personal data sharing in Europe. In this sense, the same rules which aim at overcoming the shortage of big data and AI deployment could generate obstacles to data

sharing. This threat will persist as long as some inconsistencies between the new data governance framework and the GDPR are not solved.

The European Data Protection Supervisor was the first authority to release an Opinion about the interplay between GDPR and the EHDS, already in 2020. Through this Opinion the EDPS tried to clarify some critical aspects pertaining personal data availability in the EHDS. [121] Furthermore, more recently, the EDPB and the EDPS have released two additional Joint Opinions. The first concerning the interplay between the GDPR and the DGA [16] and the second on the EHDS. [15] The two Joint Opinions in general state that the DGA[73] and the EHDS proposal lack of consistency with the GDPR. The EDPB and the EDPS, and also several scholars, raised many critical points; however, for the sake of discussion, just some of them will be analysed. The concerns presented here have been organised into four areas of discussion: 1) the definition and the terminology used throughout the DGA and the EHDS; 2) roles, and responsibilities of the new actors; 3) the legal basis for processing personal data and data re-use mechanisms.

3.2.1 – Definitions and Terminology

A first set of issues raised in relation to the DGA proposal concerned the lack of consistency between some definitions and the terminology used in the DGA and the proposed EHDS and the GDPR. In this regard, the Joint Opinion on DGA has identified the following as problematic: "data holder"; "data user"; "metadata". It shall be noted that some of these definitions have been changed in the final version of the DGA, therefore the legislator has solved part of the doubts. Nevertheless, not all the concerns have been removed and some difficult interpretation, or at least a poor readability of the text, is maintained to some extent. Addressing here the concerns raised by the EDPB and EDPS, although partially solved in the final version of the DGA, is useful to understand possible misinterpretation of the legal text.

In the DGA proposal, the definition of data holder 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" was problematic since it seemed to introduce a right upon a legal person holding personal data to grant access and share such data. EDPB and EDPS believes that clarification could be introduced explicitly stating that both the access and the sharing of personal data constitute a data processing under article 4(2) GDPR. [16] However, the final version of DGA did not include the view of the authorities. Though, the definition of data holder has been changed specifying that a data holder is a legal person or natural person "who is not a data subject with respect to the specific data in question" that can grant access to data or even directly share such data[74]. Specifying in the wording of the definition at stake that natural persons in this case do not also include data subjects can solve many interpretation doubts. The previous version of the text of the DGA proposal, would leave room to an interpretation in the sense that data subjects are the natural person allowed to grant access or share their own data. Such an interpretation would create overlapping with the GDPR provisions. The GDPR, indeed, already foresee mechanisms and requirements for data subjects to share their personal data.

[73] At the time of the Joint Opinion the DGA was only a proposal, indeed some of the concerns raised by the EDPB and EDPS have been eventually solved in the final version of the DGA.
[74] Article 2(8) DGA.

Moreover, such an interpretation would even exclude natural person that share non-personal data from the scope of the DGA. [122]

The definition of "data user" has a difficult interplay with the notion of recipient under article 4(9) GDPR. Indeed, a data user is defined as the natural or legal person that "has the right to"[75] use data for commercial and non-commercial purpose. This definition could lead to doubts about the attribution of the roles of data controller, joint controller, or processor under the GDPR to entities falling in the definition of data users. For example, recital 35 DGA[76] explicitly states that data intermediation services are bound to the obligations of data controllers or processor under GDPR, when processing personal data. However, such explicit statement was not foreseen for data users or data altruism organisations in the DGA proposal, although it seems clear that also such entities will process personal data. The final version of the DGA however, has included a similar statement also for data altruism organisations[77]. On the contrary, an explicit statement in this sense seems still missing for data users, for this the readability of the text could be affected in a negative way. Nonetheless, the definition of data users has also been modified from the proposal to the final version of the DGA. The final version of the DGA foresees two cumulative conditions for being considered as data users: 1) having lawful access to certain data; 2) having the right to use such a data. The new version has brought some clarification since it specifies that the right to use the data shall be considered also as having the right to process data under the GDPR. The previous version, on the other hand, as second step for being considered data users, mentioned being authorised for the data use. The authorisation necessary to process data was not better specified, neither in terms of its content, nor as regards who was supposed to provide it. [122] . However, under data protection law, there is no meaningful distinction between "having lawful access to" and "having the right to use" personal data. In the first place, accessing data is already considered a data processing, i.e., accessing data basically means using it. Moreover, under data protection law having lawful access shall mean having a lawful legal basis, as well as respecting all the other principle under article 5 GDPR. [122, 123] Therefore, some doubts still exist in this sense also in the final version of the DGA.

The definition of metadata (article 2(4) in the proposal), read in conjunction with article 11.2 (proposal) could have been interpreted as creating a legal basis for processing metadata. Metadata could also be deemed as being personal data. This concern was stemming from article 11 of the proposal which provides that data sharing services should be allowed to use such data for the development of the data sharing service. [16] However, the final version of the DGA has deleted such definition[78].

The definition of data sharing is also carrier of uncertainty. Namely, where it refers to data sharing as the "provision of data […] for the purpose of joint or individual use of such data […] directly or through an intermediary" [79]. The Joint Opinion believes that such

[75] Article 2(6) of the DGA proposal used the words "is authorised" instead. [109]
[76] Recital 28 in the proposal.
[77] Recital (50) DGA.
[78] Metadata is now mentioned only one time in the text of the DGA, namely in recital (16), where PSB are encouraged to develop harmonised approach and procedures to make data available for scientific research purposes in the public interest. In this sense, such harmonised procedures should also concern metadata.
[79] Article 2(10) DGA.

definition could be at least confusing when it comes to personal data processing. Moreover, there seems to be confusion in the wording of some passages of the DGA which refer to rights and interest of legal persons regarding their data. Indeed, recital (14), article 11(6) and article 19 DGA proposal seemed to treat on the same level rights and interests of individuals on personal data and rights and rights and interest of legal persons non personal data. Such recital and articles, the wording of which has been basically maintained under the approved version of DGA, will be meant use of the same provisions to address situations different to each other. This lack of clarity about the different rights and interests of data subjects on the one hand and the legal persons on the other hand could lead to a provision not solid from a conceptual point of view and difficult to implement as well. [16]

Another area of concern is the legal basis for personal data processing within the new data governance framework. In this sense, the text of the DGA proposal[80], but also of the final version[81], refers in different occasions to the "permission of data holders" for the use of data. The EDPB and the EDPS deem that it is not always clear whether the object of such a permission can only be non-personal data or also personal data. [16] Where personal data are processed it shall be noted that such a permission cannot replace the need to have a legal ground under GDPR. The Joint Opinion stresses the need to specify the necessity to have a legal ground under the GDPR for each data processing within the DGA. [16] Indeed, the final version of the DGA explicitly states in article 1(3) that all data processing involving personal data shall be in compliance with EU data protection law. Therefore, the interpretation that could be provided to "permission" in the DGA is as the business choice of a legal person to let another legal person process personal data as long as there is a legal basis for doing it according to the GDPR. This interpretation indeed seems to be supported by the fact that the approved version of the DGA explicitly states that "permission" means giving the right to re-use non-personal data only[82].

In case of re-use of data for altruistic purposes the re-use will be allowed for pursuing the general interest of the society, i.e., the common good. Besides semantic misalignment where the DGA mainly uses the words "general interest"[83] while GDPR "public interest", these notions are not completely clear. According to article 2(16) DGA purposes of general interest can include, but are not limited to, "healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research purposes in the general interest". Further clarification on these definitions should be provided for the sake of the whole system of legal provisions.

In conclusion, the EDPB and the EDPS generally auspicate for a clarification of the definitions in order to make them consistent with the GDPR and explicitly state that DGA does not amend or remove any definitions pertaining to data protection law domain. Some issues concerning definitions and terminology have been tackled by the final version of the

[80] See recitals (11), (39) and articles 2(10), 5(6), 7(2)(c), 11(11), 19(3) DGA proposal.
[81] See recitals (15), (26), (45), (46), (50), (52) and articles 2(6), (15), (16), 5(6), 5(9), 6(5)(f), 7(4)(d), 12(n), 21(3), (6), 22(1)(a), (b), 25(1) DGA.
[82] Article 2(6) DGA.
[83] The DGA mainly uses the words "general interest", see for example recitals (3), (12), (13), (45), (46) and articles 2(16), (18), (21), 4(2), 15, 16, 18(b), 19(4)(h), 20(2)(b), (c), 21(1)(a), (b), (2), 35 DGA. However, the text of the DGA on some occasions also uses the words public interest, see recitals (6), (16), (11), (16), (24) DGA.

DGA. Nevertheless, the final version still maintains some uncertainty in light of personal data processing.

3.2.2 – Roles and Responsibilities of the New Actors

The roles and responsibilities under data protection law of the actors introduced by the DGA and EHDS is another aspects worth of consideration. Such roles are to be clarified when personal data will be processed within the new EHDS. The risk, in this case, the EDPS highlights, is that individuals will face difficulties to exercise their rights before data controllers. [121] In general, the Joint Opinion on DGA argued for a clarification in the text of the proposal about the roles, in light of data protection law, that actors could play in the DGA. This clarification is needed to avoid ambiguity and improve readability of the text which could lead to misinterpretation and difficulties from a compliance perspective. [16] For example, the EDPB and the EDPS deem doubtful the fact article 5 of the proposal introduced an obligation upon public sector bodies in supporting re-users to obtain consent from data subjects. In this respect, it shall be noted that the final version of the DGA has slightly modified the text of the DGA replacing the word "support" with the expression of making best efforts in providing assistance to data users in seeking consent of the data subject. Such an obligation was indeed not fully specified in the DGA proposal. However, this provision applies as far as it does not create disproportionate efforts to the PSB. Moreover, it shall be noted that the use of consent in this case could be problematic also in light of the imbalance of power between individuals and public authorities. The GDPR transparency principle seems to be put in danger when the DGA does not foresees any obligations upon the public sector bodies to inform data subjects in article 5 DGA, which sets the conditions for data re-use.

From a different perspective, the EDPS suggested the establishment of entities responsible for the collection of data to be re-used within the EHDS. Namely, the EDPS envisaged single contact points at national level which would act as "coordinator between the requests to have access to specific data and the databases relevant for their research activities". On the other hand, the potential re-users of data were recommended to declare and demonstrate to be pursuing specific research objectives with relevant public interest. [121] Actually, the regulatory proposal for an EHDS, in article 36, foresees this role for health data access bodies. [110]

The EDPB and the EDPS also analysed the role of data intermediaries under the DGA in light of GDPR requirements. The DGA set out three different types of data intermediary activities: 1) intermediary activities between data holders (as legal persons) and data users[84]; 2) intermediary activities between data subjects and data users[85]; 3) data cooperatives[86]. As regards the first type of data intermediaries, they should operate as a platform that allows the sharing of data (also personal data) under a bilateral as well as multilateral perspective. The EDPB and EDPS are especially concerned by the situation where these platforms will intermediate between an indefinite number of data holders and an indefinite number of data users. In this scenario, it could be difficult to ensure the respect of some principles of the GDPR, such as data protection by design and by default, as well as the

[84] Article 10(a) DGA.
[85] Article 10(b) DGA.
[86] Article 10(c) DGA.

principle of transparency. The data subjects, in relation to this last aspect, will face difficulties in understanding the purposes of the re-use and the potential impacts stemming from them. [16] In order to be compliant with the data protection principles of purpose limitation, privacy by design and by default, as well as transparency, the platform should "allow a preselection of and prior information about the purposes and users of her or his personal data by and to data subject". [16] As regards data intermediary services that operate between data subjects and potential data users, the Opinion highlight that the proposal does not specify how such service providers are supposed to assist individuals in exercising their rights under GDPR. The EDPB plans to provide guidelines about modalities through which exercising such rights. [16] Finally, as regards data cooperatives, the EDPB and EDPS believes that the notion of data cooperative is still unclear as well as its obligations (even though a definition of the services provided by data cooperatives has been introduced in the last version of the DGA)[87]. In this perspective, it shall be noted that recital 24 of the proposal referred to data that pertain to several data subjects, however such wording is not consistent with data protection law and its definition of personal data. Moreover, it was contradictory when the DGA proposal stated that cooperatives could be endowed with the powers to "negotiate terms and conditions for data processing". Indeed, terms and conditions for processing personal data are as a matter-of-fact rights and freedoms under the GDPR. Data subjects' rights and freedoms cannot be negotiated through any agreements. [16] The final version of the DGA kept the critical definition of data cooperatives as services that support its members in the negotiation of terms and conditions for processing (also) personal data[88].

Continuing the discussion on data intermediaries, the DGA foresees a notification regime for organisation that want to operate as such. The organisation shall respect the requirements explicitly set down in Chapter III DGA in order to play this role. The notification regime is a mainly declarative mechanism, with just a formal control on the respect of the conditions by potential data intermediation services. The EDPB and EDPS have observed how the EC has opted for a maybe too loose system. [16] On the contrary, the system should be more aligned with the principle of accountability. Therefore, data sharing services should also be able to demonstrate their compliance with the essential conditions and not just to declare it within a notification framework. Data sharing services should set forth policies and measures to demonstrate their compliance with the DGA as well as with the GDPR. The EDPB and EDPS claims that in this case a CM or a CoC could help data sharing services demonstrate their compliance. [16]

In conclusion, EDPB and EDPS stress that actors involved in the new data governance framework set out in the DGA and further specified in the forthcoming EHDS will have to find their role in light of the GDPR. The three main roles, besides data subjects, are the public sector bodies that hold data, data intermediaries (could they be data sharing services, data cooperatives or data altruism organisations) and data (re-)users. These organisations will be covering, once DGA and EHDS become both applicable, either the roles of data controller or joint controller or data processor under the GDPR, as long as personal data is involved. Data sharing providers, as well as data holders, will probably have to bear a heavy

[87] Article 2(15) DGA.
[88] *Idem*.

burden in terms of compliance stemming from unsolved misalignment between the DGA and GDPR.

3.2.3 – Legal Basis and Data Re-use

The legal basis for processing personal data in the context of the DGA is another set of problems that affect the personal data processing in the EHDS. In general, the choice of the correct legal basis under the GDPR could be in general problematic. Moreover, looking at the health domain and namely at the re-use of personal data for research purposes finding a proper legal basis could be even more problematic. [120, 124–126] Against this background, the DGA has introduced a new data altruism mechanism, which comes along with a data altruism consent form. The data altruism consent under the DGA is not supposed to introduce a new legal basis for data processing. However, the relationship between data altruism consent and the GDPR consent seems not to be straightforward. On the other hand, the EHDS will probably be deemed as an EU law that constitute a legal basis under article 9(2)(h), (i), and (j) GDPR. Nevertheless, also the use of such legal basis in the context of the EHDS has raised concerns in the EDPB and EDPS view.[121] Finally, in this vein of discussion, it is worth to mention the issues related to the mechanism of further processing compatibility and its relationship the purpose limitation principle of GDPR.

The lack of consistency concerns, in the first place, the role of consent as legal basis under the GDPR and the data altruism consent introduced by the DGA. Indeed, consent under the GDPR shall respect the requirement of specificity, which is stemming from the principle of purpose limitation of GDPR[89]. It shall be noted that relying on consent for sharing health data under the GDPR faces the obstacle of indicating, at the moment of data collection, the purpose of future research in a detailed way. On the contrary, it is often the case in health research that data controllers need to share and re-use data for purposes that were unknown at the time of data collection. The GDPR allows data controller to use a broader type of consent when it comes at research purposes. According to recital (33) a data subject should be able to provide his consent to certain areas of research when the specific research purpose is not identifiable (as long as ethical standards are respected). However, in this case the interpretation provided in the guidelines on consent under GDPR from the EDPB, [127] seems to be rather restrictive.

The data altruism consent under the DGA, on the other hand, allows individuals, as well as companies, to make their data available for re-use for the public benefit, such as scientific research[90]. This consent is defined as altruistic since no compensation is foreseen for who decides to share its data. It shall be noted that sharing data on an altruistic ground is not a novelty per se in scientific research domain. [111] The DGA codifies this concept into hard law provisions, generating problems vis a vis existing data protection legislation. According to Shabani, the data altruism consent, despite meant to be a transparency tool, does not enhance the individuals' power and control on final result of the research. [114] Moreover, it shall be noted that fundamental rights cannot be waived by the data subject not even for altruistic purposes. It means that the GDPR rules on consent shall apply also to data altruism consent. The Joint Opinion then argue for aligning the provision about data altruism with GDPR conditions for obtaining consent. [16] The final text of the DGA

[89] Article 5(1)(b) GDPR.
[90] Article 2(16) and Chapter IV DGA; article 40 EHDS proposal.

indeed provides that the rules concerning consent under GDPR still applies also in case of data altruism consent.

The data altruism consent, as included in the DGA, revolves around the concept of general interest and common good. The altruistic re-use of data shall be indeed performed only for these purposes. Today, it is still not clear the exact meaning of these concepts and their interplay with GDPR requirements[91]. [111, 114, 115] Therefore, clarifying the relationship between the GDPR's consent and the DGA data altruism consent can be instrumental to enhancing research. [111, 114, 115] The data altruism consent could also be seen as an opportunity to harmonise some aspects of the GDPR consent collection process. [111] In particular, data altruism consent could fill the interpretational gap left by EDPB for a broad consent for research purposes. [111] Today, MSs have room to ask for the use of consent for scientific research or allow other legal grounds[92]. Moreover, the data altruism mechanism, as mentioned above, limits the use of data only to "purposes of general interest". Into the category of acceptable "purposes of general interest" is included also scientific research. The Joint Opinion warns that using consent as legal basis for data altruism context, even if it is in line with GDPR requirements, is not free from obstacles. Indeed, since the purpose of data processing will probably be of scientific research (at least in many cases) it will be difficult to identify in advance the specific purpose, i.e., comply with the purpose limitation principle under the GDPR. It means that a consent for processing data for a purpose of "scientific research for general interest" as such is not allowed under GDPR rules. At least, certain areas of scientific research shall be indicated. In light of these considerations, the EDPB and EDPS suggest that the DGA better specifies the meaning of general interest in the data altruism context. Namely, a list of defined purpose of general interest shall be included in the DGA itself.

The definition of data altruism itself seems to require the use of the consent as legal basis: "data altruism means the voluntary sharing of data on the basis of the consent of data subjects […]"[93]. However, as mentioned previously, the consent, as ruled under the GDPR does not fit in many general interest purposes mentioned in article 2(16) DGA. It could be the case that data altruism consent will be used in combination with other legal basis of the GDPR, or as an additional safeguard for data subjects, but not for supporting the lawfulness of the data processing under articles and 6 and 9 GDPR. In this sense, the EDPS envisages the different role for data altruism consent as complementary element to the public interest legal basis.[121]

The use of consent is not deemed to be the best legal basis for many purposes that are mentioned under the umbrella of "general interest" by article 2(16) DGA. However, the DGA, by introducing a data altruism form, might increase the understanding of data altruism mechanisms for data subjects. Being said that, data altruism consent shall not amend principles of the consent enshrined into article 7 GDPR. It means that structural flaws in using GDPR consent for some purposes, e.g., for scientific research, would not be overcame by data altruism consent probably. [114, 115] Therefore, even if the notion of general interest will be deemed the same of public interest, further inconsistency is raised by the

[91] See also above Section 3.1
[92] See also Chapter 3 Section 2.5.
[93] Article 2(16) DGA.

use of the correct legal basis for the specific purpose pursued, especially in case of special categories of data under article 9 GDPR. [122]

Delving into the concepts of general interest and common good, both GDPR and DGA includes the scientific research into the notion of public interest scope. In light of the common good stemming from research, some facilitations for data re-use are introduced in GDPR for carrying out such data processing. Nevertheless, the GDPR itself lack to clarify the notions of "scientific research" and "public interest", and "general interest". This uncertainty is transmitted to the DGA provisions, especially when research projects are financed by private funds. In this regard, is not clear to what extent receiving private funds would affect the character of public interest of the research project. [114] However, receiving private funds is not the only parameter that determines the impact of the research in terms of common good. For example, other aspects to take into account, are the importance of the research questions for the cohort, or the possibility to access research databases. [114]

From a different perspective, the Joint Opinion expresses concerns about the fact that the DGA proposal does not set up strong requirements from technical, organisational and legal point of view for becoming data altruism organisation. It can be noted that the EDPB and EDPS suggest in this case to rely on CMs and CoCs in order to enhance the accountability of data altruism organisations. [16]

The more general topic of the legal basis to be used for health data processing in the context of the new EHDS is another big area that could be discussed in terms of difficult interplay with GDPR. The legal basis for processing data in the research domain are fragmented among MSs, as well as most of the requirements of the GDPR that deal with scientific research and health data. [80] Indeed, MSs have adopted different approaches when allowing re-use of data for scientific purposes under the public interest as legal basis. [114] The fragmentation is due to different ways to organise the health care systems among MSs. Indeed, although the EU primary law (art. 168(2) TFEU) aims at organising the provision of health care services, MSs have the power to adopt the specific policy actions about health care services' organisation. In other words, MSs are those who defines the specific functioning of national health care systems. On the other hand, EU shall only coordinate, support and supplement the actions adopted at Ms level according to the subsidiarity principle. [128] Although this division of powers is rightfully justified by historical, political, and societal differences in the way MSs address the subject at issue, this has inevitably led to differences in Europe. These differences in terms of organisation of the health care systems forced the EU legislator to leave room to national legislators also in terms of data protection requirements. Indeed, as said before, many GDPR requirements and obligations are not perfectly harmonised when it comes to health data processing and to the sharing of health data in EU. In this sense, recital 10 GDPR explicitly states that "a margin of manoeuvre for Member States to specify its rules, including for the processing of "sensitive data" concerning health, biometric and genetic data" is provided. MSs have therefore room to introduce exceptions on legal basis for data processing, on additional safeguards to some data processing, and on exceptions to data subjects' rights. This has been envisaged as a potential obstacle to the creation of the EHDS. [128]

The task of the EHDS should be to provide clarity to the abovementioned uncertainties. The EHDS builds upon the data governance framework of the DGA, though focusing on electronic health data. It means that most of the data processing operations will concern

personal data related to health, which are a special category of data under article 9 GDPR. Therefore, in addition to the standard legal basis under article 6(1) GDPR, a further legal basis under article 9(2) GDPR must be envisaged in order to process such data. In this respect, the EDPS in its preliminary opinion on the EHDS stated that public interest (article 9.2(i)) or scientific research (article 9(2)(j)) should be considered as the suitable legal basis for data processing in EHDS[94]. The current version of the EHDS indeed followed such indication, since the EHDS regulation will constitute an EU law suitable as legal basis according to articles 9(2)(h), (i) and (j)[95]. However, the EDPS is perfectly aware of the fragmentation among MS in data protection law for research purposes. In this sense the EDPS welcomes the desire of the EC to facilitate the development of an EU wide CoC for processing personal data in the health sector.

The problem of the legal basis has been also stressed by the EDPB and the EDPS in their Joint Opinions on the EHDS and the DGA. Namely, in their Joint Opinion on DGA, the two authorities have shut the door to the interpretation that deems the DGA as a legal basis for data processing. It will be a task of actors involved in the data re-use to identify the most suitable legal basis, within those provided by the GDPR, in order to ensure compliance with the GDPR. This is not going to be an easy task, but at least the proposed EHDS regulation seems to provide clarification for the health domain. Indeed, as already said, the EHDS proposal will constitute an EU law that provides enough suitable and specific measure for processing personal data according to article 9(2)(h), (i) and (j)[96]. However, it still means that data processing shall be supported by an appropriate legal basis under article 6(1) GDPR. In this respect, it is possible to remind that EDPS do not deem the consent as the most appropriate legal ground when it comes at data processing for policy making or research purposes. [129] On the other hand, according to the EDPS, article 6.1(e) shall be used as legal ground in the EHDS context. The choice is due to the fact that the main purpose pursued will be the public interest. [121]

It could also be noted that the DGA proposal back then was not clear about its interplay with the re-use mechanism under article 6(4) GDPR. From this point of view, the EDPB and the EDPS deem that the DGA proposal cannot be invoked as Union law constituting necessary and proportionate measure in a democratic society to safeguard the objectives referred to in article 23(1) for a further data processing under article 6(4). Moreover, the EDPB and EDPS do not believe the DGA is suitable for being an EU law that can support a personal data processing under article 6(3) neither (art. 6(1) (c) and (e)).

Moving to the last aspect addressed in this section, from the EDPS Opinion emerged the necessity to make data processing activity compliant with the purpose limitation principles. In the EHDS context, this aspect relates to the necessity of clarification of the purposes before the beginning of the data processing. Complying with the purpose limitation principle has always been however tricky in big data analytics context, especially when further processing is carried out for scientific research. [12] Article 6(4) GDPR, in combination with article 89 GDPR, provides the framework for further use of personal data for this domain. The EDPS therefore believes that the EHDS should clarify the conditions under which personal data are to be re-used in compliance with GDPR framework. [121]

[94] See in this sense also [129].
[95] See recital (37) EHDS proposal.
[96] *Idem*.

About the re-use of certain categories of data, also the Joint Opinion believes the DGA does not set a clear relationship with the GDPR. Indeed, the GDPR already provides rules for re-using data held by public sector bodies while protecting fundamental rights. From this perspective, the two authorities note how article 6(4) GDPR provides already a mechanism for the re-use of personal data. Data re-use within the DGA, as long as it concerns personal data, shall be in compliance with article 6(4) GDPR. Article 5 and recital 15 DGA set the conditions for data re-using under the DGA. This provision seems to be of difficult interpretation in light of the further processing mechanism under the GDPR. Namely, article 5 DGA does not provide explicit indication on the purposes for which the data re-use shall be authorised in light of the compatible further processing mechanism under the GDPR. Moreover, the DGA does not indicate that such specific purposes shall be identified in national level legislation. In this sense, the DGA is not suitable as legal basis under article 6(1)(c) or 6(1)(e) or 6(4) in combination with article 23 GDPR. This aspect has however been clarified in the last version of the DGA through article 1(3), which explains that the DGA does not introduce a legal basis for processing personal data. On the other hand, it has been said that EHDS does introduce a legal basis for processing personal data, although only in the context of electronic health data.

3.2.4 – Legal Hypertrophy, Lack of Consistency and Meta-level Rules.

From the analysis of the present chapter what emerges is that organisations, either private or public companies, will carry a heavy burden for compliance duties and regulatory complexities. This is particularly true for those who operate in the health sector, being them public sector bodies, data intermediaries or health care providers that deploy big data and AI systems. The endeavour of the EC to boost the re-use of data clashes with other measures introduced to protect rights and freedoms of individuals, such as the GDPR. Although the GDPR and the AIA proposal contain mechanisms to ensure the smooth functioning of the internal market and the free flow of data in EU, when the discussion comes to a more practical level of abstraction some consideration should be done. It has been noted how the new data governance framework impose severe duties on the actors involved. These compliance tasks would be difficult to bear especially for PSBs, as well as data altruism organisations. [122]

The health domain moreover is carrier of high risk, both in light of safety of the products that embed AI systems and in terms of data processed. The GDPR as well as the AI act revolves around a risk-based approach. The higher the risk the heavier the compliance duties upon data controllers and AI deployers. The complexity and the heaviness of the legal framework under analysis has therefore lead data protection authorities, but also the EC itself, arguing that some compliance issues would be relieved by using data protection CoC or CM. [80, 121]

Most of the burden in terms of compliance costs will be probably borne by data intermediaries and data holders. This could hamper the final goal of increasing the sharing of data within Europe. Indeed, data sharing providers, according to article 12 DGA have several duties, including also ensuring that different legislations are not infringed during data re-uses. Moreover, data intermediaries could be put in a situation where they have to evaluate in terms of trade-off which policy goals prevail on the others. Such an evaluation is due to the sometimes overlapping and not consistent interplay between the DGA and other legislative acts, such as the GDPR. [123] In order to avoid for the GDPR to become the

"elephant in the room" [123], compliance shall be facilitated and smoothed, but also the approach of stakeholders towards data protection shall be different. Namely, organisation should start to approach privacy as an investment and no more just as a cost.

In order to face innovative technologies legal issues, it has been argued in this first chapter that a regulatory shift has been carried out by the EU legislator in the GDPR. A new co-regulatory approach revolving around risk analysis and the principle of accountability has been followed in designing the rules. However, such regulatory system will not solve compliance burdens if the general co-regulatory approach is not paired to a proper functioning of the practical co-regulatory tools of CoC and CMs. In this sense, the EDPS warns that the EHDS shall ensure that data controllers and processors have organisational and technical security measures to protect personal data, in line with article 32 GDPR. A DPIA would obviously be necessary in most of the cases, since the nature of data and the large scale of data processing will classify data processing as high-risk. The EDPS also recall the CMs under article 42 GDPR as instruments to be used to enhance trust among stakeholders. It would also be useful if GDPR CMs would be co-ordinated with CMs foreseen by the Chapter III of the EHDS proposal for EHR systems and the voluntary labelling for wellness application under article 31 of the EHDS proposal. [15, 16] Besides improving harmonisation through the use of an EU-wide CoC, the feasibility of which is not free from obstacles, the EDPS suggests using a CoC for bringing clarity and building trust among stakeholders and patients. [121] The next Chapter will delve into the discussion of the origin and the meaning of co-regulation concept as well as into CoCs and CMs' specific role and function in GDPR.

Chapter 2 – Co-Regulatory Instruments for Data Processing

1 – Co-regulation

The allocation of power among the state's bodies is a topic discussed in many filed, from law studies to political science. In the same way, it is discussed the choice to involve different actors, other than state bodies, in the regulatory process. The present work does not aim to provide an exhaustive overview in this sense. However, it shall be borne in mind that a paradigm shift is investing the way policies, governance strategies, and even the legislative measures are shaped and applied in the modern world. [130, 131] New phenomena, such as the fast-developing technologies here discussed, as well as globalization, are creating new form of knowledge generation. The means of production of such knowledge are usually handled by the private sector. In many sectors most of the know-how is held by private actors, which led also to the necessity to envisage new form of involvement of the private sector into governmental decision-making. [132] A new concept of "governance" is replacing the old paradigm of "regulation". The new governance concept embeds dynamic processes of attribution of roles during rulemaking. The boundaries between the roles of public actors and private stakeholders in defying the content of policy and governance strategies and even of the content of the regulatory measures is becoming less clear and less neat. [132–134] A clear example in this sense is the incorporation of international technical standards, elaborated by standardisation organisations, into EU legislative measures[97].

Traditionally, the state had always imposed rules to regulate the functioning of the society. The decisional process about the definition of the content of the law was up to State's legislative bodies, as it was the enforcement of the law. This approach is commonly known as *top-down regulation*, or *direct government regulation*, or *state regulation approach*. This approach has in general many positive sides, such as ensuring legal certainty or a sound protection of fundamental rights and freedoms. However, limits of state regulation have become evident in some domains. Among them there are the sectors that are strongly impacted by the technological developments. It is indeed difficult in these cases for governmental bodies and the legislators to have enough knowledge to adopt effective decisions. [135–138]

Complexities brought by innovative technologies put in crisis traditional schemes and paradigms through which the state used to regulate society. [137] The domain of big data and AI is a clear example of technologies that are having huge impact on many aspects of the society[98]. The legal domain of privacy and data protection law has been hugely impacted by the use of big data analytics and AI systems, which have generated a set of transformations in the way many activities are carried out by companies, citizens and by public sector bodies. Today, companies base their business on the possibility to transfer personal data from one country to another in a ubiquitous way. Organisations are more and more

[97] See in this sense the NLF and the use of harmonised standards elaborated by the European Standardisation Bodies, upon a formal request of the EC: https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en (last access: January 2023).

[98] See above Chapter 1, Section 1 for an overview of the impacts that AI and big data may have on the healthcare domain, and consequently also on the application of data protection law in this context.

eager to transfer data within dynamic networks, while the old paradigm of one-to-one data transfer has been overcome. [137] However, it is not just a matter of data flows between different actors, but also how data is used in the decision-making process[99].

The technologies abovementioned are literally changing how we live our lives and how company conduct business. The way companies shape their business models is changing, and the way citizens enjoy products and services, or even the way public services are delivered to citizens is changing as well. [139] As it has been discussed in Chapter 1, big data and AI are paving the way to new great results in the health care domain, for the provision of care, as well as for scientific research, or for the monitoring of public health. On the other hand, the use of such technologies could endanger rights and freedoms of individuals. These changes have made some legislations no more effective or have raised the need for new laws to face the new risks. [140] In this light, data protection law has been re-shaped by the introduction of the GDPR. [141, 142] The aim was ensuring a firm protection of individuals rights and freedoms against the misuse of personal data, which are at the basis of the functioning of big data analytics and AI systems.

Whether the reform of data protection law in Europe has reached the purpose to ensure protection of individuals against risks posed by big data analytics and AI is object of discussion among scholars. [17, 18, 75, 76] However, even though the GDPR is not totally effective against modern risks to privacy and data protection, it constitutes for sure and important steps towards the goal. [18] In this perspective, it is possible at least to state that the GDPR has changed the regulatory approach in data protection law, representing a shift in data protection law in this sense.

The GDPR might be deemed a piece of legislation that embraces and endorses a co-regulatory approach. [73, 143] The GDPR can be defined as a regulatory measure that relies on co-regulation, since it foresees an involvement of private actors next to public authorities throughout the process of application of the law. The GDPR relies on the accountability principle and the risk-based approach, as seen in the previous chapter[100]. These principles require a great involvement and effort from data controllers and processors in terms of interpretation and demonstration of compliance. In order to help to reach out such a compliance, the GDPR has introduced co-regulatory instruments. Such as CoCs, CMs and impact assessments. [144] These instruments shall be elaborated by middle-layer stakeholders and sometimes approved by an authority either at national or EU level. Therefore, despite being voluntary measures, CoCs and certifications can help data controllers and processor in reaching compliance with GDPR principles. As already said, the content of these instruments is elaborated by private stakeholders within the framework defined by the GDPR itself. Afterwards, a more or less formal discussion with DPAs anticipate the final approval from these authorities. [145, 146]

Co-regulation though is a broad concept covering different definitions. Therefore, before moving to the analysis of the specific co-regulatory instruments provided by the GDPR, it is necessary to better elaborate such a concept. The present chapter is organised as follow. The Section 1.1 deals with the definition and the concept of co-regulation, whilst

[99] See above Chapter 1, Section 1.1 about how the use of big data and AI technologies is changing the way decisional processes are adopted in health domain. Data-driven decision-making processes are replacing traditional way to adopt decisions. This shift is key especially domains such as research and policy making.
[100] See Chapter 1, Section 2.1.

Section 1.2 with the classification of co-regulatory instruments according to the mainstream literature. Section 2 provides an overview of co-regulatory tools in data protection law context. At first, Section 2.1 briefly traces back the history of these instruments in data protection law from the DPD to GDPR. Section 2.1.1 in particular analyse the particular use of CoC under the DPD in Italy. While Section 2.1.2 discusses the new and amplified role of these instrument under GDPR, namely in light of the new accountability principle introduced by the GDPR. After that, Section 2.2 focuses on CoC analysing their functioning in light of articles 40 (Section 2.2.1) and 41 (Section 2.2.2) GDPR and providing an overview of some CoCs examples (Section 2.2.3). Section 2.3, on the other hand, performs the same exercise but focusing on CMs. Namely article 42 (Section 2.3.1) and 43 (Section 2.3.2) GDPR are assessed and presented, and an overview of existing examples of these instruments is provided (Section 2.3.3). Section 3 eventually analyses the different roles that can be played, and the effect generated by these instruments on the data protection law context. Namely, the role in terms of compliance facilitators is analysed by Section 3.1. The effect in terms of legal certainty and reduction of legal fragmentation are touched in Section 3.2. Finally, the relationship of these instruments with market dynamics and economic aspects are discussed in Section 3.3. In conclusion, Section 3.4 tries to shed some light on the reason of the sparse application of these instruments in light of the roles and functions highlighted in the previous Sections.

1.1 – Defying Co-regulation

In the first place, co-regulation shall be distinguished from de-regulation. De-regulation is the policy and governance choice that aims at removing any regulatory obstacles to free market exchanges. [147] On the other hand, co-regulation and (in part) self-regulation aim at changing the actors involved in the process of rule-creation, but not at removing the regulatory framework itself. [147] Moreover, co-regulation shall not be approached as an alternative or substitute of top-down regulation, but only as a complementary element.[144] In other words, co-regulation can be defined as the choice, at governance and policy level, to regulate phenomena not just through top-down approaches, but also involving middle-layer actors. It is possible to envisage a co-regulatory approach every time that the State and its governmental bodies intentionally share with industry the task to draft and enforce rules. [137]

Co-regulation is different from early self-regulatory, or to de-regulatory, approaches, because it implies that public authorities keep a role in the regulatory process. The state or the administrative bodies are usually involved in the co-regulatory solutions to ensure that public interests are not overcome by private interests. This choice is meant to ensure greater transparency in the co-regulatory process and respect for fundamental rights. [143] In order to be qualified as co-regulation, it is necessary having strong enough cooperation between the State and private actors. Indeed, according to the extension of the involvement of the state in the process, different types of co-regulatory approaches can be identified. A literature overview of the taxonomies on co-regulation can be found in the next section[101].

It is also important to note that in co-regulatory approaches the private parties that actively elaborate rules do not create rules applicable just to themselves, as it happens in some self-regulation instruments. Indeed, once the co-regulatory instrument is approved by an

[101] See below Section 1.2.

authority, everyone willing to adhere to it can do it. In this sense, the private regulators do not set up rules for themselves, but rather define rules applicable to a whole group. [144] Co-regulation is therefore meant to overcome the limits of State regulation, such as the lack of knowledge about the industry sector, and its technological development, to be regulated. [135–138] In the same way, also the limits of pure self-regulation approach are meant to be tackled by co-regulation.

It shall be noted that self-regulation is often promoted, for different reasons, by private stakeholders as the best solution. For example, one of the arguments provided is the ability of self-made and self-enforced rules to cope with real needs of the industry sector and its technological advancement. Or, for example, the fact that stakeholders would be more prone to adhere to rules developed and tailored by their peers rather than to rules imposed from the top [148]. However, there is more than one reason why it is possible to argue that pure self-regulatory approaches are not the proper solution, especially in the data protection and privacy regulatory domain. Indeed, from a theoretical point of view, it is possible to argue that self-regulation approach is prone to favour companies' interests over interests of the whole society [149]. Companies could be tempted to develop only apparently sound and solid privacy and data protection rules, while, in fact, trying to enforce these rules not very rigorously. [149] Moreover, from a practical perspective, self-regulation does not foresee a formal approval from public authorities, meaning that no regulatory compliance is ensured through self-regulation mechanisms. In other words, even though implementing self-regulatory instruments, companies still have to comply with state laws. [137, 150]. Self-regulation does not reduce the burden upon private stakeholders in terms of compliance costs. For this reason, even industry actors might not find attractive to be involved in self-regulatory processes. [137]

Co-regulation tries to overcome the issues of both state-regulation and self-regulation approaches. Co-regulation indeed relies on industry knowledge for better facing the needs of the specific sector at stake, introducing rules that should be better accepted by the stakeholders involved. On the other hand, the State maintains the role of defining the general framework and principles within which the industry can set rules[151]. Moreover, a public authority is still in charge of the final approval of the co-regulatory instruments and, sometimes, even of their enforcement. [137] Nevertheless, co-regulatory approaches also present shortcomings. Indeed, the negotiation between industry and the government, in practical terms, is not as smooth as in theory. The risk is to either have deals that are too prone towards industry interests, [152] or a too rigid state involvement in the negotiation process. In this last case, the risk is that private stakeholders would lose interest in the negotiation activities because the government is too rigid or too slow in the negotiation process. Indeed, reaching an agreement between the state and private stakeholders on topics with high interests at stake, such as privacy and data protection is not an easy task. This is usually a process that takes a lot of time and resources from both parties. These are all factors that shall be taken into account when developing a co-regulatory approach. [23, 153, 154]

Moving to another aspect of the discussion, this work argues that two dimensions of the concept of co-regulation operating at different levels of the regulatory process can be detected. The first dimension refers to co-regulation as the regulatory method or the governance strategy. The second one refers to the specific set of co-regulatory tools and mechanisms within a specific legislative act. As regards the first dimension, there was an explicit

reference in the better law-making policy action of the EC[102]. [108] Co-regulation used to be officially recognised as an "alternative method of regulation" in the 2003 interinstitutional agreement on better law-making. [130] This provision is no more part of the new interinstitutional agreement as it had been updated in 2016. [131] However, it is worth noting that the point 18 of the 2003 interinstitutional agreement states that "co-regulation means the mechanism whereby a Community legislative act entrusts the attainment of the objectives defined by the legislative authority to parties which are recognised in the field (such as economic operators, the social partners, non-governmental organisations, or associations)". Under this perspective, co-regulation could even be thought as an implementer of the principles of subsidiarity and proportionality[103]. [155] Co-regulation is therefore meant to limit the intervention of the legislative bodies, in this case at EU-level, only to the cases where it is strictly necessary and to the extent that is necessary to reach the objectives of the legal intervention. [156–158] Co-regulatory solutions should therefore be used in order to pursue objectives defined by a legal act and smooth the implementation of these general objective in complex sectors. This regulatory method is meant to move part of the regulatory burden from the legislator and, on the other hand, scale up the skills and the knowledge of private parties recognised in the sector. The second dimension refers to the practical tools that can be elaborated in order to actually shapes co-regulatory rules. These tools changes from one sector to another and from one legislative act to another. They include code of conduct, certification, standards, impact assessment tools, an many others. Sometimes the EC is in charge of supervising the content of such instruments, as it is the case for harmonised standards, sometimes this is a task of national authority or other EU monitoring bodies. This work will be focusing on two specific co-regulatory tools in data protection law: 1) certification mechanisms (CMs) and 2) codes of conduct (CoCs) as envisaged in the GDPR.

In conclusion, co-regulation is an enabler of the principles of subsidiarity and proportionality since it helps legislative bodies alleviate the regulatory burden in complex contexts. Although, the co-regulatory approach has also effects on the addressees of the legal requirements. Indeed, it should provide them with compliance tools, the content of which is shaped by relevant stakeholders under the supervision of the designed public authority.

1.2 – Classifying Co-regulation

In general, co-regulation is classified according to the extent of the involvement of the public authority in the regulatory process, either during the drafting of the content or in the monitoring activity. Therefore, it is possible to see that usually co-regulatory approaches vary from almost pure self-regulation to co-regulatory instruments that entail a very strong involvement of the public authority. Broadly speaking, self-regulation does entail any oversight or approval process from public authorities, while on the other hand co-regulation is characterized by different degrees of such an involvement, according to the specific instrument at stake. [144] However, it has been noted that there are some differences in the classification of these instruments in the literature. This Section does not try to carry out a

[102] The better law-making policy has been replaced by the better regulation agenda, see: https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how_en (last access: January 2023).
[103] See articles 5(3) and (4) TEU.

complete review of all the taxonomies and classifications of co-regulatory instruments. However, it is a useful exercise to highlight the mainstreams classifications that is possible to find in literature.

One of the first taxonomies of self/co-regulation instruments carried out is the one developed by Black in the nineties. [159] According to Black, it is possible to classify self-regulatory instruments looking at the relationship that such instruments have with the State. "Mandated" self-regulation is the situation where a group is put in charge by the government to develop and also enforce the rules within a framework that is defined by the State. "Sanctioned" self-regulation, on the other hand, is the case where the group formulate rules which although shall pass through a governmental approval. "Coerced self-regulation", for Black, is the scenario where industry if forced to self-regulate against the threat of the State to use top-down regulation. While finally, "Voluntary" self-regulation is the situation where a group or a sector decides to regulate itself without any incentives or push from the State. [159] According to Black, the discussion about self-regulation classification should revolve around three key points: 1) what is meant by "self"; 2) what is meant by "regulation"; 3) what is the "involvement of the State" in self-regulatory process. [159]

According to how these three points are shaped then the co-regulatory instrument changes. Indeed, if we think to "self" as a single individual or a single company that "regulates" itself without any "involvement of the State", then the instrument at issue will be not much more than an internal policy of a company. On the other hand, we could think to "self" as a group of company or an industry sector and "regulation" as a set of rules that apply to all the members of the group. Moreover, we can even envisage an active "involvement of the State", which can for example be in charge of the approval of the rules. In this case, the concept of self-regulation would completely be different.

Ayres and Braithwaite, on the other hand, argue that regulatory strategies vary from pure self-regulation to command-and-control measures. In the middle is possible to find co-regulatory solutions. Such solutions can be divided between those that see a group dealing with State mandate to regulate a sector and those solutions that aim at regulating the relationship between one single company and the State. The latter is defined by Ayres and Braithwaite as "enforced self-regulation". [160]

According to Latzer et al., [161] there are forces that are both driving towards more involvement of the public authorities and situations where on the contrary regulatory instruments are more self-oriented. These forces essentially depend on the specific domain and situation. However, Latzer et al. elaborate a taxonomy that starts with co-regulation and finish with "Self-help/restriction by users including rankings to impose restrictions on access to content". In the middle, it is possible to find situations like: "state-supported self-regulation"; "collective industry self-regulation"; "single company self-organization". Marsden further elaborated a detailed taxonomy of self and co-regulatory instruments starting from the work of Latzer et al. [143] In his taxonomy, Marsden identifies twelve level of distribution of authority between public and private parties. At lower level, with no enforcement, there is "pure un-enforced self-regulation" which could represent a model like the one of the videogame SecondLife. In this case, regulatory model is totally informal, and a single player impose its one rules on costumers. Further on the taxonomy, the "Recognised self" is a situation where, although no formal enforcement is foreseen, there is the recognition of a body that is assigned with an informal role of control over policies. In the Marsden taxonomy, CoCs and CMs of GDPR could be included between the categories of

"Approved self" and "Approved compulsory co-regulation". Into these two categories are included those regulatory instruments that entail an ex-ante consultation with the government and an approval procedure. [143]

According to Hirsch, regulatory choices shall be shaped around two main questions.[137] First of all, who is going to regulate? Second of all, at what level of the society are the rules to be applied? According to how these questions are answered, different combination of regulatory choices can be envisaged. Indeed, it is possible to answer to the first question in three different ways: 1) the state is the regulator; 2) the industry will regulate the sector by itself; 3) the government and industry will share the regulatory task. The second question (i.e., at what level to regulate?) has three possible answers as well. Indeed, it is possible to regulate: 1) at the level of a single company; 2) at the level of an industry sector; 3) at the level of the whole economy. [137] Therefore, Hirsch further classifies co-regulation in three sub-classes. The first one is co-regulation at company level, where a single company negotiates with the State the rules. The second one is "sector-based co-regulation" where the rules become applicable to the entire sector because they are negotiated by a representative of the sector. The last one is "economy-wide co-regulation". [137] In light of this last taxonomy, the GDPR approach can indeed be placed in the sector-based co-regulation approach. Especially for CoC, indeed, it is necessary the code owners demonstrates a certain level of representativeness in the sector at stake. Namely, the code owner shall demonstrate that it is able to understand the problems typical of the domain. [162]

2 – Overview of Co-regulatory Instruments in Data Protection Law

It has been mentioned in the previous sections of this work that the EU legislator has adopted a co-regulatory approach when re-shaping the data protection law legal framework. The GDPR, indeed, relies on a risk-based approach and on the accountability principle. This regulatory choice can be considered itself as a co-regulatory choice. The EU legislator only defines principles and general objectives in many parts of the GDPR, while leaving the choice about how to implement them to the addressee of the law. In order to do that, the GDPR itself provide some tools for compliance that can be used by data controllers and processors to facilitate compliance, but also to shape and harmonise the content of concrete actions for reaching the goals set in GDPR's principles. The instruments at stake here are CoCs and CMs, the rest of this section will then provide an overview of these instruments under the GDPR and a quick historical excursus under the previous DPD.

It shall be noted that other parts of the GDPR could be deemed as providing tool of compliance, such as article 35 concerning the DPIA or article 47 on binding corporate rules. However, these instruments are not taken into account in this work given the limited scope of the discussion. The focus will be on CoCs and CMs because they are the instruments that respond to the definition of co-regulation as a solution that is elaborated by a group representing a sector (and not by a single company or a stand-alone group of companies) which includes the participation of a public authority with the task to approve the instrument.

2.1 – Co-regulation in Data Protection Law: from DPD to GDPR

The DPD, which was the former piece of legislation regulating data protection law in EU, already foresaw CoCs as co-regulatory measures. While CMs where not included in the text of the DPD. Article 27 of the DPD states that MSs and the EC shall encourage the draft of CoCs by "trade associations and other bodies representing other categories of controllers". The CoCs under article 27 were meant to "contribute to the proper implementation" of national data protection law, taking into account specific features of various sectors. In the same way of the current GDPR provisions, CoCs under the DPD could have been applicable to a single MS or at Community level.

As for national CoCs, the draft code should have been submitted to the competent DPA for approval. However, the DPD did not provide any specific rule about the modalities for the approval procedure by the DPA. On the contrary, article 27(2) DPD assigned to MS legislator the task to identify such procedural laws. Concerning Community CoCs, the approval procedure was a task of the Working Party Article 29 (hereinafter WP29)[104]. In both cases mentioned above, the DPAs or the WP29 shall evaluate whether the draft CoCs were in accordance with national provisions adopted pursuant to the DPD. However, article 27 DPD leaves room for other kind of evaluation during the decision of approving the CoC by DPAs or the WP29. Article 27(2) and 27(3) DPD indeed provide that respectively a DPA and WP29 shall evaluate, "among other things", that the CoCs were respectful of national data protection law.

It seems that the structure and the tasks played by CoCs were pretty much the same of those played by CoCs under GDPR, though much less specified and defined by the provisions of the DPD. This makes sense if we look at the type of legal instrument under the EU law perspective, i.e., a directive rather than a regulation. Indeed, directives are supposed to leave room for discretionary implementation of the law to MSs. On the other hand, a regulation, like the GDPR shall determine almost all the aspects of the regulated subject matter. However, in 1998 the WP29 has published a guidance which has defined more precisely the procedure for the approval of Community CoCs under article 27(3) DPD. [163] The procedure for the submission of a Community CoC was essentially organised in two steps. The first step was an evaluation of whether the draft CoC can be submitted as Community CoC. This phase included, among other things, the ascertainment that the organisation submitting the CoC was enough representative in at least a "significant number of Member States". In this phase, it was also desirable to demonstrate that consultations with data subjects have been carried out. Moreover, the sector where the CoC was to be applied should have been defined appropriately.

If the draft CoC was deemed to be acceptable in light of the abovementioned evaluations, then the draft CoC was submitted to the members of the WP29. The WP29, at this point, shall in the first place evaluate whether the CoC respect the DPD provisions, and, if applicable, whether it respected the national provision adopted pursuant to the DPD. After that, the WP29 was called to examine the quality and the consistency of the CoC. [21] The draft CoC was indeed supposed to provide sufficient added value to the DPD implementation. The WP29 in this regard specifies that a CoC shall be sufficiently focused on specific data protection issues in a given sector, in order to actually provide clear solutions for data controllers. [163]

[104] The Working Party Article 29 was the body composed by the representatives of the DPAs in EU, it has been replaced by the EDPB when GDPR has been enacted.

Community CoC have been seldom approved under the DPD. In 2003, the Federation of European Direct and Interactive Marketing (FEDMA) CoC has been approved by the WP29. [164] Moreover, some other drafts of Community CoC have submitted for approval to the WP article 29, i.e., the World Anti-Doping Agency (WADA) International Anti-doping standard, the C-SIG Code of Conduct on Cloud Computing, and the Code of Conduct for Cloud Infrastructure Service Providers. [21, 165, 166] However, these draft CoCs have not been approved by the WP29, as not meeting the compliance with DPD provisions or because they were not providing enough sufficient added value. [21] At MSs level, it was a task of national legislator to encourage the drafting of CoCs as well as defining specific rules for their approval by DPAs. The room left by the DPD to national legislators and national DPAs has led to some fragmentation among MSs about how CoCs have been shaped. For example, in Italy there was a particularly harsh interpretation of the role played by DPA in the drafting and enforcement of CoCs[105].

When the DPD was in force, some CoCs have been developed even for the medical field, such as the "Code of practice on secondary use of medical data in European scientific research projects". [167, 168]. Such code of practice was stemming from the necessity to ensure legal compliance in research projects at EU level that involve transborder data processing. The code of practice was submitted to the French DPA "Commission nationale de l'informatique et des libertés" (CNIL)[106] and to the Belgian data protection authority "Autorité de protection des données" (APD) [107]. The final goal was to finally submit the CoC to the WP29 and become a Community CoC. However, the code has never been finally approved by any of these authorities. This code moved from the assumption that many concepts of data protection law under the DPD and the proposed (at the time of drafting the code) GDPR were not clear enough, especially if transposed to the domain of re-use of personal data for scientific research. Therefore, the code's main aim was to help researchers easily understand basic compliance practices in light of data protection law implementation. But the code also aimed at filling some interpretative gaps left by the law and by the DPAs in their role of providing consistent understanding of the law. The code focuses on aspects related to anonymisation and pseudonymisation of health data for research purposes, as well as on the collection of consent and the concept of biometric and genetic data[108]. [167]

It is interesting to note that recital (26) of the DPD explicitly mentions data anonymisations as a possible subject matter of a CoC. This explicit reference is mentioned no more in the GDPR. Anonymisation is indeed a crucial topic in data protection law in general, and even more in the biomedical research field. Anonymisation is therefore one of the potential contents of a CoC for the health domain, as it will be further discussed later in this work[109].

2.1.1 – The Italian Experience of Codes of Conduct

[105] See below Section 2.1.1.
[106] https://www.cnil.fr/en/home (last access: January 2023).
[107] https://www.dataprotectionauthority.be/citizen (last access: January 2023).
[108] The paper, which was presenting the code, concludes with some recommendation for the regulators, see [167].
[109] See Chapter 3, Section 2.1.

It is worth to note that under the DPD the transposition of requirements about CoCs, from the directive to national law, has not always been perfectly aligned with the rationale of the DPD. Indeed, at national level, legislators have sometimes partially used the opportunity to elaborate CoCs for developing de facto mandatory rules. Interesting is the Italian case, where the Italian legislator has put the Italian DPA in charge of developing, only in partial collaboration with private stakeholders, CoCs for personal data processing in some sector. The sectors, explicitly identified by the legislator, included, among others, personal data processing for statistical and scientific purposes, [169, 170] personal data for journalistic purposes [171] or historical research purposes. [172]

The Italian data protection law implementing the DPD[110] assigned to the Italian DPA the task of "promoting" CoCs pursuant article 27 DPD. However, it was possible to read in the article at stake that the involvement of private stakeholders was limited to a rather consultative task[111]. Moreover, the approved CoCs were supposed to be published in the Italian Official Journal and attached as annexes to the law itself, i.e., the D.lgs 196/2003[112]. The compliance with such CoCs were deemed essential in order to ensure the lawfulness of the data processing[113]. A total of seven codes adopted pursuant to article 27 DPD used to be attached as mandatory annexes to the D.lgs. 196/2003.

These instruments were partially discussed with stakeholders of the sector, but the final decisions on the content was basically held by the DPA. Moreover, the respect of these instruments has been made mandatory against a fine from the DPA and the impossibility to proceed with the processing of the personal data. However, after the introduction of the GDPR, the interpretation on the use of such instruments has been partially aligned with the rationale of CoCs under the GDPR. Therefore, according to article 20, D.lgs 101/2018[114], which is the law that has modified the D.lgs 196/2003 in order to adapt it to the GDPR, the role of these instruments has changed.

According to the D.lgs. 101/2018, two of the DPD codes out of seven have been transformed into CoCs under article 40 of the GDPR, as provided by article 20, D.lgs. 101/2018. Namely, the representative stakeholders of the sectors have been called to discuss again the content and submit a revised draft code to the Italian DPA for approval. The first of these two CoCs were concerning the personal data processing which take place within information systems handled by private companies for evaluating the reliability of costumers in payment activities. [173] The second one, on the other hand, were concerning the personal data processing concerning commercial information linkable to physical persons. [174] The transformation of these two codes in CoCs under article 40 GDPR has been explicitly

[110] Article 12 D.lgs 196/2003 version before the modification brought by the D.lgs 101/2018.
[111] Article 12.1 D.lgs 196/2003.
[112] D.lgs. 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).
[113] Article 12(3) D.lgs 196/2003.
[114] 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, nonche' alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati). (18G00129) (GU Serie Generale n.205 del 04-09-2018).

required by the national legislator, which is a peculiar feature at least, if not again against the rationale of the GDPR rules on CoCs. The remaining five codes [169–172, 175] have been kept as something more similar to mandatory regulation rather than co-regulation. Indeed, according to article 20(3) and (4), D.lgs. 101/2018, the Italian legislator has asked to the DPA itself to review the content of these codes and verify its compatibility with the GDPR. Article 2-*quaters* of the D.lgs 196/2003, as adapted to the GDPR, foresees a specific role for these codes. The instruments are now called "deontological rules", and the role of mandatory requirements to be respected when certain data processing is carried out is assigned to them. These codes are indeed again directly included as annex in the D.lgs. 196/2003 and missing to respect them make the data processing unlawful.

It is not clear why the Italian legislator has decided to move two of the seven codes under the umbrella of article 40 GDPR, while keeping the others as *de facto* mandatory rules, without involving private stakeholders in their review and maintaining only the DPA in charge of this task.

2.1.2 – Co-regulation and its Role Vis-à-Vis the Accountability Principle

Under the GDPR, the greater role assigned to co-regulation tools for compliance is testified by the introduction of CMs, which were missing under the previous DPD. The enhanced reliance on this kind of solutions is due to the increasingly important role played by the accountability principles and by the risk-based approach in the GDPR. Indeed, this work argued that if on the one hand such approach allows for more flexibility and technology neutrality, on the other hand, it generates compliance burdens on data controllers and processors[115]. In order to support the compliance activity, especially in complex sectors, the EU legislator have decided to incentive co-regulatory solutions. CoCs and CMs are both instruments meant to fill the gap between abstract rules in the GDPR and demonstration of compliance. These instruments, although different between each other are both meant to better implement GDPR principles, guiding the data controller and processor in the choice of the technical and organisational measures for the protection of personal data.

CoC are explicitly regulated by articles 40 and 41 of the GDPR while CMs are regulated by articles 42 and 43 GDPR. Nevertheless, such instruments are mentioned in many other parts of the GDPR. CoCs and certifications are indeed both considered by article 24 GDPR as instruments that can be used as element to demonstrate compliance with GDPR's provisions in general. In this sense, the accountability function of CoC and certifications is explicitly stated in the GDPR itself. Moreover, both CoCs and certifications are provided as element to demonstrate compliance with some specific requirements of the GDPR. For example, CoC and certifications can be used as element by data processors for demonstrating sufficient guarantees, pursuant to article 28(1) and 4 GDPR. Article 32, as well, foresees the possibility to use CoCs and certifications as element for demonstrating the compliance of technical and organisational measures adopted for data security. Interestingly, only certifications, but not CoCs are explicitly mentioned as element to demonstrate compliance with the principle of privacy by design and by default under article 25 GDPR. Nevertheless, article 40 GDPR suggests the use of CoCs also to demonstrate compliance with article 25 GDPR. Therefore, this can be safely considered just a formal lack of coordination between the parts of the GDPR. In general, it is safe to assume that every provision of the

[115] See Chapter 1, Section 2.1.

GDPR can be subject matter of a CoC or a CM pursuant to articles from 40 to 42 GDPR. The fact that some specific provisions of the GDPR explicitly mention that these instruments can be used as accountability tool probably means that CoCs and CMs dedicated to just few aspects of the GDPR are as acceptable as more comprehensive tools that concerns compliance with the whole GDPR. [22]

Besides being accountability tool for the demonstration of compliance before DPAs, CoCs and certifications also see their *raison d'etre* in the necessity to make of compliance activities not just a cost, but an investment for companies. The GDPR preparatory works [77, 78] highlight how the choice to move from a rigid top-down mechanism of data protection implementation to a more flexible structure was also meant to change the way data controllers approaches data protection and privacy. Due to the technological shift, the GDPR is now based on the concept of risk and the principle of accountability, as said many times already. However, the GDPR was also meant to create a privacy culture and literacy, where the respect of data protection rights and freedoms is considered an important asset to take into account by both commercial partners and customers. In this sense, the goal is to make of data protection co-regulatory tools an enabler of a cultural shift (towards respect of privacy) within business-to-business relationship as well as in business-to-individuals perspective.

As regards the business-to-business relationship, data protection compliance is supposed to become a competitive advantage and a market differentiator. This role is especially assigned to CMs[116]. [22, 176] As regards the relationship with the data subjects or potential customers, certifications should make individuals able to quickly evaluate the level of compliance of a data controller. In this sense, a costumer could be more prone to choose one product rather than another because one provides a certified guarantee that personal data processing concerning it are in compliance with the GDPR. [146] CoCs on the other hand are instruments with less communicative powers towards business partners or data subjects, however they can contribute to enhancing the trust among different actors, especially in a context like the health. [162] Indeed, a CoC, if adopted by all the participants of a research project for example, can stimulate patients in providing their data. The CoC adoption indicates that data are going to be processed within a trusted framework that follow additional compliance rules – indicated in the CoC – which have been endorsed by the DPA or even by the EDPB. [168]

2.2 – Codes of Conduct

CoC are the first of the two co-regulatory instruments touched here, they are not a novelty per se, since the DPD already had foreseen their use. However, as said before, under the GDPR their role and functioning has been further empowered and detailed. CoC are now regulated by two articles of the GDPR, namely articles 40 and 41. Article 40 defines their role as co-regulatory instruments for the enhancement of compliance, while article 41 sets the rules for the monitoring of approved CoC.

A CoCs share with a CMs the role of instruments for compliance and accountability. However, the former is different to the latter since they are supposed to be developed for clarifying broad principles and requirements into sectors where such implementation is particular difficult. [162] In other words, while CMs are meant to guide the data controller

[116] See below Section 3.3.

towards the compliance and release a certification of it, CoC are also supposed to adapt the application of the GDPR to the specific sectorial difficulties. However, it shall be noted that for CoCs is not mandatory to be sectorial, even though most of the time this would be the case. The EDPB has indeed specified that even cross-sector CoCs could be acceptable. In this case, more than one monitoring bodies would probably be involved in the verification of compliance. [145]

It is possible to argue that certification mechanisms have a stronger market orientation, since they are usually paired with mark and seals in order to also communicate to customers and partners the data controller's commitment towards compliance in data protection law. [22] CoCs, on the other hand, are less oriented towards the market dynamics, but are more likely to smooth the interpretation of GDPR into a specific sector. [21]

CoCs, in their activity of facilitating the application of the GDPR, can have a twofold impact. In the first place they facilitate data controllers or processors' compliance, increasing also legal certainty. [145] Namely, they can help data controllers and processors solving compliance issues as well as reducing the risk to get fined from DPAs. On the other hand, CoCs can help the monitoring activities of the DPAs. This role is due to the fact that the compliance of data controllers with CoCs' requirements is monitored by some bodies. These monitoring bodies are supposed to report to the DPA if they became aware of GDPR violations. Therefore, these monitoring bodies operate also, to some extent, as rapporteurs for DPAs. [21, 177]

It shall be noted that there is even a third impact stemming from the adoption of CoCs, which is the enhancement of trust that stakeholders and data subjects can benefit from the spread of CoCs. Trust is supposed to be strengthened by a more consistent application of GDPR's provisions and by a more solid respect of data subjects' rights. [145] According to the EDPB, even a harmonisation function could be played by CoCs. Indeed, such instruments could help to bridge the gap between different MSs implementation of the GDPR. [145] Although very promising, this function might be object of wide debate about its actual effects, as it will be discussed more deeply later in this work[117].

The GDPR dedicates two articles to CoCs. Article 40 deals with the general requirements about CoCs, concerning their role, their elaboration and approval. Article 41, on the other hand, deals with the monitoring of compliance with the CoC's rules and the accreditation of monitoring bodies.

2.2.1 – Article 40 GDPR

Under the GDPR, there has been an extension of the scope of CoCs. [21] Indeed, under article 40 GDPR both data controllers and processors can adhere to a CoC. On the contrary, under the DPD only data controllers could have done it. Article 40 GDPR firstly clarifies who are the subjects that are supposed to elaborate the CoC, i.e., "associations and other bodies representing categories of controllers or processors". On the other hand, "the Member States, the supervisory authorities, the Board and the Commission" shall only encourage the drafting of the CoC, but not proceed to the elaboration of the CoC itself. This specification seems to be meant to overcome possible interpretation of the provision not in line with the concept of co-regulation enshrined into the GDPR. This was the case under the

[117] See below, Section 3.2.

DPD when some MSs have made of CoCs pursuant article 27 DPD an almost hard law instrument[118].

It shall be noted that there is a representativity requirement to be satisfied. In theory, only actors that have enough representativity among a group shall be allowed to submit a draft CoC to the DPA. Nevertheless, such a requirement is not of easy application, as long as it is not better specified in the GDPR when such "associations and other bodies" can be considered representative enough. [21] The decision about whether an association or a body represents enough the sector might become a discretionary choice of the supervisory authority. However, the EDPB guidelines on CoC [145] provides some more detailed information on this point. Namely, in order for a DPA to evaluate the level of representativeness of a stakeholders, the following elements shall be taken into account: 1) "Number or percentage of potential code members from the relevant controllers or processors in that sector"; 2) "Experience of the representative body with regard to the sector and processing activities concerning the code". [162] The representativeness requirement is in line with the concept of co-regulation that the GDPR seems to have endorsed, i.e., rules that are developed by private parties and that are meant to discipline the behaviour of an entire group (or at least those of the group who decide to adhere to the CoC). [159]

As said above, the feature that distinguish CoCs to CMs is the fact that CoCs are meant to specify the application of the GDPR. Article 40 clearly states that point, also listing some possible aspects of the GDPR that could become object of a CoC and therefore benefits from a specification and adaptation into a specific sector. These aspects are listed in article 40(2), however this is not to be considered a closed list of content a CoC shall focus on. Indeed, whichever aspects, requirements, or principle of the GDPR can become content of a CoC. Nevertheless, it could be useful to remind that the GDPR explicitly suggests as content of a CoC, for example, transparency of data processing, pseudonymisation, or the exercise of data subjects' rights. The GDPR then mentions CoC as an element to demonstrate compliance with other specific provisions, such as articles 24, 28, 32 or 35. CoCs indeed as mentioned before, can be used to demonstrate the appropriateness of technical and organisational measures (as in article 24 or 32) as well as that processors or controllers provide acceptable guarantees (as in article 28 or 46.2(e)). CoCs shall also be taken into account and can be brought as an element that mitigate the impacts stemming from the data processing when conducting a DPIA.

Another function of CoCs, that is actually to be considered separately from the others, is the use of CoCs as "appropriate safeguards" under article 46(2)(e) for data transfer to third countries. CoCs can be used for transferring personal data to third countries, however, only if the data controller or processor that receive data in the third country commit itself with binding and enforceable contractual measure (or other legal measures) to comply with the CoC, especially as regards the exercise of data subjects' rights. Moreover, CoC can be used as appropriate safeguards only if they receive the double approval of the national DPA and of the EDPB and if general validity is provided to it by an implementing decision of the EC. [145]

These observations lead us to the fact that there are essentially two kinds of CoCs that is possible to develop under the GDPR: 1) national CoCs; 2) general validity/transnational

[118] See above, Section 2.1.1 on the case of the Italian CoC pursuant article 27 DPD, which were mostly elaborated directly by the DPA itself.

CoCs[119]. The first category refers to CoCs that are supposed to be applied only to data processing within a single MS. In this case, the draft CoC shall be submitted to the competent supervisory authority according to article 55 GDPR. The supervisory authority is then supposed to evaluate if the CoC provide enough safeguards, end, if this is the case, eventually approve it. The second category is reserved to CoCs that are meant to be applicable to data processing in more than one MS. In this case, the competent authority shall, before approving the CoC, submit the draft to the EDPB. The EDPB shall provide its own opinion on the draft CoC and, if it confirms that the CoC complies with the GDPR and provides enough safeguards, submit it to the EC. At this point, the EC, as mentioned before for the transfer of personal data to third countries, can decide to assign general validity to the CoC. [162]

In either case, the organisation submitting the CoC to a DPA or EDPB shall demonstrate the following elements. In the first place, that the CoC "meets a particular need of that sector or processing activity". Then, that the CoC "facilitates the application of the GDPR", as well as that the CoC "specifies the application of the GPDR". The CoC shall then provide "sufficient safeguards", and "effective mechanisms for monitoring compliance"[120]. [145]

As regards the first requirement, the EDPB specifies that a CoC, in order to be considered acceptable, shall explicitly indicate which problems about data protection law implementation it aims to address. Moreover, the CoC shall also be able to indicate which solutions it provides for them. The solutions should not be only beneficial for data controllers, but also for data subjects. As regard the second point, the CoC shall also indicate how it is supposed to facilitate the GDPR application into the specific sector indicated in light of the issues addressed. Then, as third requirements, the CoC shall specify the GDPR. It basically means that the CoC cannot just re-state the GDPR. For example, sector-specific standards shall be indicated in the CoC. Moreover, the CoC, especially if applicable to high-risk sectors shall demonstrate to have appropriate safeguards for rights and freedoms of data subjects. And, finally, the CoC should include systems and a structure for ensuring the monitoring of data controllers and processors' compliance with its provisions. [145] In conclusion, both the EC and the EDPB are required to provide adequate publicity to approved CoC. Namely, the EDPB should keep a register of all the CoCs approved both at national and EU level[121].

2.2.2 – Article 41 GDPR

As mentioned before[122], under article 27 DPD, the process of complying with CoC was not detailed, nor were the characteristics of the bodies in charge of monitoring the adherence of data controllers to such CoCs. These aspects were left to the choice of MSs or DPAs. Article 41 GDPR, on the other hand, establishes the rules for ensuring a proper monitoring activity over data controllers' compliance with approved CoC. It shall however be noted that article 41 applies only to data processing carried out by private sector bodies (article 41.6 GDPR). It essentially means that if a public sector body decides to adhere to an approved CoC pursuant to article 40, the monitoring of its compliance with such a CoC

[119] See also Chapter 4, Section 1.4 on the approval procedure for national and transnational CoCs.
[120] See also Chapter 4, Section 2.1.
[121] See https://edpb.europa.eu/our-work-tools/accountability-tools/register-codes-conduct-amendments-and-extensions-art-4011_bg (last access: January 2023).
[122] See above Section 2.1.1.

is not covered by article 41 rules. In other words, as specified by the EDPB [145], the monitoring of a CoC which concerns data processing carried out by public authorities or bodies is not mandatory to be carried out by an accredited monitoring body. However, the EDPB clarified that this provision shall not be interpreted as removing the monitoring duty upon a CoC. On the other hand, alternative methods of monitoring should be incorporated into the CoC, possibly relying on already existing auditing activities carried out on public authorities. [162]

The monitoring of compliance with the CoCs' requirements by the data controllers and processors that decide to undertake a CoC is assigned to a competent body which has obtained accreditation from the DPA for exercising this task. Article 41 leaves the identity of such bodies intentionally vague. [177] Article 41(2) indeed provides only some requirements that, if met, qualify a body as with appropriate expertise for monitoring the compliance with the CoC. Therefore, in theory, whichever body met such requirements can be accredited by the DPA as monitoring body for a given CoC.

According to article 41, this body shall be competent with the subject matter of the CoC and shall receive an accreditation from the DPA. Some characters of impartiality, professionality, and lack of conflict of interest shall be demonstrated by the applicant body[123]. Namely, the body shall demonstrate independence and expertise. It shall also demonstrate of having procedures in place that enable an assessment of the eligibility of data controller or processor for applying to the CoC. The CoC shall also contain procedure that allows the accredited monitoring body to monitor the compliance of data controllers and processor with the CoC. A body, in order to be accredited, shall then have procedures to handle complaints about infringements of the CoC by the controllers and processors that have adopted it.

The accredited monitoring body shall also be able to take actions in case of infringement of a CoC. For this reason, without prejudice to tasks and powers of the DPA, the monitoring body can suspend or exclude from the possibility to adopt a CoC those data controllers and processors that infringe a CoC. However, the sanctions can be further specified in the CoC itself and could even include monetary penalties. It shall also be noted that the monitoring bodies are also supposed to inform the DPA when sanctions are issued and the reasons why those sanctions have been applied to a data controller. In this sense, the monitoring bodies operate as intermediary bodies in referring to DPAs possible GDPR violations. [177] On the other hand, the supervisory authority can withdraw the accreditation of monitoring bodies in case the conditions for accreditation are no longer respected by the monitoring bodies, or if the actions taken by these bodies infringe the GDPR.

As regards the procedures that data controllers and processors shall follow in order to adhere to the CoC, article 41 indirectly indicates that the adherence to a CoC shall be monitored periodically by the monitoring bodies. Such monitoring activities are decided by the monitoring body itself or defined into the CoC already. [177]

In concrete terms, the monitoring bodies will probably be most of the times the same organisations that draft the CoC[124]. Another option could be that this role is played by

[123] Kamara notes how the lack of conflict of interest could be difficult to demonstrate ex ante, before staring the activity of monitoring the specific CoC. This should rather be also an ex-post evaluation by the DPA after accreditation. [177]

[124] This is the case of the Farmaindustria CoC, see page 13 [232], or the case of EU Cloud CoC, see [233].

conformity assessment bodies[125], which, are bodies that are specialised in performing conformity assessment activities on products, processes, services, systems, installations, or projects, for example, i.e., verify that such objects conform to a set of given requirements. [177, 178] This twofold possibility is endorsed by the EDPB itself, which foresees either an "internal" or an "external" monitoring of CoC. [145]

As for the accreditation procedures, the DPA shall elaborate accreditation requirements on the basis of the element indicated in article 41(2), such requirements shall then be submitted to the EDPB according to the consistency mechanism referred to in article 63 GDPR[126]. Once the applying monitoring body fulfil the accreditation requirements drafted by the DPA, it shall be accredited by the latter.

As noted by Kamara [177] the accreditation of monitoring bodies for CoC is more straightforward than the accreditation procedures for certification bodies under article 43 GDPR. Indeed, for the monitoring of compliance with CoCs there is not the explicit possibility to involve a NAB as accreditation body and therefore there is no need to apply international technical standards, such as ISO/IEC 17065 or the EU law on accreditation, such as the Regulation 765/2008[127]. Another difference respect to accreditation under article 43 for CMs' compliance, is that there is no expiry date for accreditation of bodies monitoring compliance with CoC under article 41. On the contrary the accreditation of certification bodies expires after five years, according to article 43 GDPR.

2.2.3 – State of the Art of Existing CoC

At the moment of writing, the are some CoCs approved at national and EU level, however there is no EU-wide CoC approved so far for health data processing. Especially in the domain of healthcare and biomedical research, the number for CoCs adopted is particularly limited. The few CoCs adopted for the health domain are national CoCs, dealing with either primary or secondary uses of data. Nevertheless, as noted on many occasions in this work, [80] the use and the re-use of data in the health domain is very heterogeneous. This feature essentially brings to the necessity of having many CoCs dealing with different aspects of the domain and the related different data protection issues.

As mentioned before, the EDPB is in charge of making all the CoCs public through a registry. The registry is accessible from the EDPB website[128], and at the moment of writing it shows a total of 6 CoCs adopted. None of these CoCs concern the use of health data[129]. However, browsing some of the DPAs websites it is possible to find out that some other CoCs have been approved at national level by DPAs, although they are not reported on the EDPB registry[130].

---

[125] This is the case for example of the CISPE cloud CoC, see [234].
[126] See article 41(3) GDPR.
[127] Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93.
[128] https://edpb.europa.eu/our-work-tools/accountability-tools/register-codes-conduct-amendments-and-extensions-art-4011_bg (last access: January 2023).
[129] CoCs included in the registry, at the time of writing, concern the following sectors: education; credit/financial; sufficient guarantees from data processors under article 28; commercial information/advertising; smart grid; assurance.
[130] The reason why not all the CoCs and CMs are reported in the registries of the EDPB is not known to the author.

Among these "unreported" CoCs it is possible to notice some CoCs dedicated to the health domain[131]. For example, in Italy, the Garante per la Protezione dei Dati Personali (GPDP) has approved a CoC concerning the re-use of health data for scientific publications and educational purposes[132]. The CoC at stake has been elaborated and submitted to the GDPD by the "Regione Veneto" along with the healthcare provider "Azienda Sanitaria ULSS 9 Scaligera". The CoC has been approved by the GPDP in 2021 with an application scope limited to Italy. [179] Namely, whichever healthcare provider part of the Italian public healthcare system can adhere to it[133].

The CoC focuses on the modalities for re-using personal data of patients, also through anonymisation and pseudonymisation techniques, for educational and scientific dissemination purposes. Therefore, the CoC is meant to provide further clarification, on the re-use of such data by physicians and healthcare professional during seminars presentations, conferences, scientific publications or case studies presentation for educational purposes[134]. The CoC regulates only personal data processing carried out by health care professionals, working within the organisation of the data controller in an autonomous way.

The CoC applies only to personal data already at the disposal of the data controller. Such data could have been collected during diagnostic, care and prevention activities, for example. Moreover, the CoC applies both to anagraphic data (such as name, surname, address, or sex) and health-related data and genetic data[135]. According to article 5 of the CoC, the main aim of the compliance instrument is to regulate the anonymisation and pseudonymisation processes for the aforementioned purposes of scientific dissemination. However, the CoC also provides some standards documentation to be used by the healthcare professional in the contest of data processing at stake. Therefore, for example, a specific form for collecting data subject consent, to be used in case it is not possible to anonymise the data, is attached to the CoC. Moreover, another technical annex of the CoC specifies the anonymisation and pseudonymisation techniques to be followed within the CoC application. The Code essentially provides that it is mandatory to apply anonymisation, or at least pseudonymisation techniques, in order to use the data. The CoC then provide also for a specific conservation time of data processed for such purposes (which is 3 years, according to article 8 of the CoC).

Another CoC adopted at national level, but not reported in the EDPB registry is a code approved by the Agencia Española Protección Datos (AEPD) concerning the processing of personal data in the field of clinical trials and other clinical research and pharmacovigilance[136]. As it is possible to read on the AEPD website, the CoC has been promoted by "Farmaindustria"[137] and is meant to be applied by promoters of clinical studies concerning

[131] An unreported CoC approved by Belgian DPAs is the EU Cloud CoC, see https://www.dataprotectionauthority.be/citizen/the-be-dpa-approves-its-first-european-code-of-conduct (last access: January 2023).
[132] https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9535354 (last access: January 2023).
[133] Article 11 of the CoC.
[134] Article 3 of the CoC.
[135] Article 4 of the CoC.
[136] See https://www.aepd.es/es/prensa-y-comunicacion/notas-de-prensa/aepd-aprueba-primer-codigo-conducta-sectorial-desde-entrada-vigor-rgpd (last access: January 2023).
[137] Farmaindustria is the associations that brings together the majority of pharmaceutical companies based in Spain, see https://www.farmaindustria.es/web_en/ (last access: January 2023).

drugs and Contract Research Organisations (CRO). The CoC is therefore meant to ensure legal compliance of the data processing carried out for research purposes, namely clinical trials, and for purpose of pharmacovigilance, i.e., monitoring activities on possible sides effects of drugs already placed on the market.

Therefore, the CoC is meant to address two different categories of data processing and the related issues. The first one, as already said, is data processing carried out during clinical trials. Under this perspective, the CoC help data controllers comply with GDPR principles, but also carrying out impact assessment activities, identifying the different responsibilities of the participants in the trial activities, identifying the legal basis under articles 6 and 9 GDPR, as well as carrying out the international transfer of personal data. The second category of data processing concerned is the pharmacovigilance. In this sense, the CoC aims at establishing common protocols for the collection of information of adverse reactions to the drugs through different channels, including even social networks[138].

Besides the approved CoCs mentioned above, it is worth mentioning at least two interesting projects that did not reach (yet) the approval of a DPA or the EDPB. The first one has already been mentioned in this work, and it is the "Code of Practice on Secondary use of Medical Data in Scientific Research Projects" which has been developed by a team of experts both from the industry sector and the academia[139]. The CoC has been submitted to the CNIL as well as to the Belgian data protection authority, without receiving the final approval though. The scope of the draft CoC was concerning the processing of personal health data to be used or re-used in collaborative research project (article 1). The subject matter of the draft CoC focused on anonymisation of personal data, as well as the use of pseudonymisation techniques. The CoC was also meant to introduce rules in order to reach compliance within genetic data processing, providing rules from the collection of the samples until the re-use of such data. However, the CoC was developed under the DPD and not the GDPR, and, as said above, it has never been approved by a DPA or the EDPB. [167]

Finally, it is worth to be mentioned another very ambitious project: the BBMRI-ERIC'[140] initiative for a CoC on health research[141]. BBMRI-ERIC, which is one of the main research infrastructure active in Europe, has launched an initiative for a CoC already in 2015. A CoC for health research to be submitted to the EDPB was the ambitious goal of the project. This CoC would have been the first EU-wide CoC for health research and even the first CoC in general to be submitted to the EDPB. In 2017 the consultations with different stakeholders began and in 2018 the actual drafting of the CoC started. However, as it is possible to read on the BBMRI-ERIC website, the project slowed down between 2019 and 2020 because of Covid-19. Nevertheless in 2021 new consultation started again[142]. Even a first draft of the CoC index has been provided, it includes aspects such as anonymization, pseudonymization, the re-use of data and the exercise of rights of participants to the research projects[143]. If ever approved, this CoC could be a landmark case for CoCs

[138] See Introduction of the Farmaindustria CoC. [232]
[139] https://www.imi.europa.eu/sites/default/files/uploads/documents/reference-documents/CodeofPractice_SecondaryUseDRAFT.pdf (last access: January 2023).
[140] https://www.bbmri-eric.eu/ (last access: January 2023).
[141] http://code-of-conduct-for-health-research.eu/ (last access: January 2023).
[142] *Ibidem*.
[143] See http://code-of-conduct-for-health-research.eu/wp-content/uploads/2021/11/Presentation-on-Code-of-Conduct-V02.11.2021.pdf. (last access: January 2023).

under article 40 and 41 GDPR, because of the importance of the promoter of the CoC and because of its EU-level applicability.

2.3 – Certification Mechanisms

CMs are the second co-regulatory instruments explicitly foreseen by the GDPR. They are a real novelty in data protection law, since they were not included in the DPD. These instruments are the first real endeavour of the EU legislator to create a market around data protection compliance. CMs, even more than CoCs, are market-oriented mechanisms. They are supposed to certify that a given data processing is in compliance with GDPR requirements and to communicate such compliance not only towards DPAs, but also towards the market. [153]

Even though certification mechanisms were not explicitly regulated under the DPD, back then there were already examples of data protection certifications. For example, the certification mechanism "EuroPrise"[144] or "Inveo ISDP"[145] were already existent even before the adoption of the GDPR. The EC had indeed commissioned a study with the aim, among other things, to map existing certification mechanisms applicable to GDPR requirements. From this study came up that many certification schemes and standards concerning privacy and data protection were already existing and applicable. However, not all of them fall into the scope of GDPR certifications under articles 42 and 43. Many of these certifications indeed covers only some aspects of privacy and data protection, e.g., the security of information. [176, 180] Anyway, already in 2008, the WP29 stated that, because of the technological aspects directly enshrined into data protection law, it was necessary to identify concrete criteria to be used for auditing data processing and awarding labels of quality. [181]

Certifications mechanisms in general, not only in the data protection domain, are traditionally meant to increase trust among different actors and stakeholders. [153] Certifications and standards are therefore aimed at ensuring the smooth functioning of commercial relationships, from both a B2B and a B2C point of view. Certification mechanisms are regulated in the GDPR by articles 42 and 43. In line with the structure dedicated to CoC, article 42 regulates the content, the role, and the functioning of certification mechanisms. Article 43, on the other hand, is dedicated to monitoring the activities of certification bodies and their accreditation process.

2.3.1 – Article 42 GDPR

According to recital (100), data protection certifications allow data subjects to quickly assess the level of data protection of an organisation. This would lead to greater transparency and trust in the system, as well as facilitate consumers' choice and also promote legal compliance. More in general, CMs, as well as CoCs, are instruments that should promote a data protection culture. [22] This kind of measures should also have impacts on the burden of regulators, reducing the need for prescriptive rules. Of course, this is meaningful only as long as the use of co-regulatory measures does not reduce the level of compliance. [22] It shall be noted however that certification measures contribute to the demonstration of compliance, but do not ensure compliance with the GDPR. In other words, it means that

[144] See https://www.euprivacyseal.com/EPS-en/Home (last access: January 2023).
[145] See https://www.in-veo.com/ (last access: January 2023).

compliance is an activity that shall always happen *ex-ante* the certification process. [22, 176, 180] This interpretation is confirmed by the EDPB itself in its Guidelines on certification mechanisms, where it is clearly stated that certifications do not constitute by themselves compliance. [146]

The compliance assessment within a certification mechanism is carried out against the criteria of the certification scheme. The criteria of the scheme are therefore supposed to transform the general provisions of the GDPR into auditable measures. For this reason, it is important that the criteria are enough clear. This requirement is meant to reduce the objectivity of the auditor in assessing the compliance with the criteria. Ideally, two different auditors, assessing the compliance of a data processing against the criteria, should always produce to the same result. In other words, there should be no room for personal interpretation. While, on the other hand, legal provisions are always object of some interpretation. [22] Once a data controller or processor is awarded with a certification under articles 42 and 43 GDPR, he is contractually bound in respecting the criteria. In case of violation of the criteria, the certification body can even withdraw the certification. However, a violation of the criteria does not automatically mean that also the GDPR has been violated. Indeed, the certification criteria might even introduce stricter requirements than the GDPR itself. [22]

Certification mechanisms under article 42 GDPR are supposed to be encouraged by the Member States, the supervisory authorities, the Board and the Commission. This provision is in line with article 40 GDPR about CoCs. However, as it has been said, the draft of CoCs is a task explicitly assigned by article 40 GDPR to organisations or bodies that can demonstrate some degree of representativeness. On the other hand, article 42 GDPR say nothing about who is supposed to draft the criteria for a certification mechanism under the GDPR. In this sense, the EDPB just says that the criteria shall reflect the requirements of the GDPR and contribute to its consistent application. [146] However, it seems safe to argue that since MSs, DPAs and the EC are asked to encourage the adoption of such mechanisms probably they should not also draft the criteria. However, the interpretation is not straightforward, indeed a certification mechanism directly drafted by a DPA has been developed and approved in Luxembourg[146]. [182] It is interesting to note though that the EDPB in the document stating the procedure for the approval of EU-wide certification schemes, the so-called "European Data Protection Seals" explicitly state that EU-wide certification can be even drafted by DPAs. The document affirms that in that case the DPA shall be considered as a scheme owner. [183]

It is important to note that article 42 explicitly states that CMs are voluntary measures for demonstrating compliance with the GDPR and that the needs of "micro, small and medium-sized enterprises" shall especially be taken into account. However, the adherence to a certification mechanism approved under article 42 GDPR does not reduce the responsibility of the data controller or processor to be incompliance with the GDPR. In other words, to adhere to a certification mechanism under article 42 does not guarantee a presumption of conformity. The adherence is, on the other hand, just an element that can be used to demonstrate compliance. [22] Nevertheless, the GDPR specifies that DPAs shall take into account the fact that an organisation has been awarded with a certification mechanism (as

[146] See the certification scheme developed in Luxembourg called CARPA: https://cnpd.public.lu/en/actualites/national/2022/06/adpoption-gdpr-carpa.html (last access: January 2023).

well as that it has adopted a CoC) when evaluating the imposition of a fine (article 83(2)(j) GDPR).

On the terminology side, has been noted that article 42 GDPR is not always consistent with other articles of the GDPR or with certification systems in other domains. For example, the GDPR seems to refer to "certification", "mark", and "seal" as the same concept and with the same meaning. Nevertheless, there are some conceptual differences between these terms. The GDPR does not define the concept of "certification". The EDB refers to the definition of certifications provided by ISO/IEC 17000:2004, which has been withdrawn by ISO/IEC 17000:2020[147]. On the other hand, marks and seals can be considered as the visual representation of data protection certification, i.e., they are visual indicators that a data controller or processor has been awarded with a data protection certification for its data processing operations. [22, 146]

Recital (100), again, defines the object (i.e., what is going to be certified) of the data protection certifications. The recital at stake indeed mentions that the object of certification shall be products, services or processes. This provision makes the data protection certification fall into a specific certifications area, which is indeed the certification of product, service and process, regulated by ISO/IEC 17065 family. For this reason, cannot be object of GDPR certifications neither persons (e.g., the DPO), nor management systems. Moreover, the only organisation that can apply for a certification are data controllers and processors, not third parties, such as software developers. [146]

A different discussion shall be carried out about the subject matter of the certification (i.e., what is going to be the content of the certification). In this respect, article 42 states that certifications shall be used in order to demonstrate compliance with "this regulation". The statement shall mean that all relevant provisions of the GDPR can be included in the criteria of the certification. Therefore, there seems to be room for both "single-issue" certifications and more comprehensive certification schemes which cover all the provisions of the GDPR. Indeed, the GDPR itself refers to specific requirements where compliance can be enhanced by a certification. For example, like in articles 24, 25, 32, 28 or 42(2) GDPR. In the same vein, the EDPB suggests single cases for GDPR certifications, such as the principles of article 5, the legal basis under article 6, the data subject rights under articles 12-23, article 33 on data breach notification procedures, or privacy by design and by default under article 25, or article 35(7) for the DPIA. [146]

It shall be noted that some certification scheme can also be sector-oriented, therefore certification schemes can target some specific data controllers and processors operating in specific domains, e.g., certification for personal data processing concerning health data for research purposes. [22] Moreover, it has been said that the certification criteria are supposed to contribute to the consistent application of the GDPR and reflect GDPR principles and requirements. For this reason, some other existing standards can be included into the certification criteria. This could be the case of ISO 27000 family[148]. In this case such standards can be used to cover specific areas that contribute to GDPR compliance, such as the management of information security. Moreover, it shall be noted that certification scheme

[147] Here the definition of certification is "demonstration that specified requirements are fulfilled".
[148] ISO 27000 family is a set of standards that covers IT security, cybersecurity, privacy and information security management. For example, the certification scheme Europrivacy foresees the opportunity to scale up on standards already applied within an organisation in order to satisfy some specific requirements of the scheme.

shall not only contain material criteria (i.e., criteria on data protection law) but also procedural criteria. This second set of criteria is meant to indicate, for example, how auditor shall implement assessment procedures, who takes decisions on the certification release, or what documentation shall be produced. [22]

The certification process is not defined by the GDPR in an exhaustive manner; however, the GDPR states that it shall be transparent for both the data applicant (a data controller or processor) and the data subjects. This information must allow the data controller and processor to understand what they are entering into and what being certified will entail. On the other hand, data subjects must be aware of what exactly means for an organisation receiving a certification in terms of enhanced protection of their rights and freedoms.

Even though the certification process is not described in the GDPR it should follow common practices for certification, as indicated in ISO/IEC 17065. Moreover, according to the EDPB guidelines, [146] the assessment shall always be amply documented, and it should report in detail how the criteria have been met. In general, there is a double step: first of all, the auditors provide a first assessment and, afterwards, the assessment is reviewed by a different person working at the certification body. This second subject is going to take the final decision about whether awarding a certification or not.

CMs, as well as CoCs can be used as appropriate safeguards for transferring personal data to third countries under article 46(2)(f) GDPR. However, also in this case it is necessary that the recipient in the third country will commit itself, with a binding contract or other legal instrument, in complying with the certification scheme.

Certifications mechanism can have general application in more than one MS if they receive the endorsement of the EDPB. In this case, the draft certification criteria shall be submitted to the DPA where the scheme owner has its headquarters, which will be considered the competent supervisory authority (CompSa). The CompSa is supposed to carry out a first assessment about the admissibility of the draft certification criteria and if this phase is successful, it shall start a first phase of informal consultation at the EDPB level with all the other DPAs. After the informal consultation phase, the scheme owner is informed about possible shortcomings and issues on the criteria. At this point, if the scheme owner decides to proceed, the phase of the formal submission starts, which leads to the formal approval or rejection of the scheme[149]. [183]

Article 42 specifies that a certification can be issued by either the DPA or by an accredited certification body, and this decision is upon the MS. However, DPAs also have the role to approve the certification criteria. For this reason, the EDPB guidelines [145, 184] recommend transparency to the DPA if the MS assigns to them the role of issuing certifications. In this light, the EDPB also requires for a clear separation of powers within the DPA about these two different tasks. [146, 184] The life span of a certification last three years, with the possibility to renew it, if the conditions under which the certification has been released are still met. However, a certification can be withdrawn by the certification body or the supervisory authority. The EDPB, as it is the case for CoC, has a duty to make available on a registry all the approved CMs in EU[150].

[149] See also Chapter 4, Section 1.4.

[150] Ref registry https://edpb.europa.eu/our-work-tools/accountability-tools/certification-mechanisms-seals-and-marks_en (last access: January 2023).

2.3.2 – Article 43 GDPR

Article 43 deals with the role of certification bodies and, especially with the accreditation process of certification bodies. Certification bodies are the competent bodies that are supposed to issue or withdraw certifications. However, it is crucial that also certification bodies are overlooked by other bodies, the latter are usually called accreditation bodies. The process of "certifying the certifiers" is a key point in building trust in the certification schemes. A lack of oversight on the bodies that issue certifications can deceive consumers and hampering the trustworthiness of the system. [185]

In EU, in general, the accreditation process is regulated by Regulation (EC) 765/2008[151] and by the standard ISO/IEC 17065:2012[152]. Indeed, a definition of "accreditation" is provided by Regulation 765/2008 in article 2(10) as "an attestation by a national accreditation body that a conformity assessment body meets the requirements set by harmonised standards and, where applicable, any additional requirements including those set out in relevant sectoral schemes, to carry out a specific conformity assessment activity". The EDPB recognises that the GDPR does not define the concept of accreditation therefore it refers to the abovementioned definition of accreditation from Regulation 765/2008. [184] In general, the GDPR accreditation process for certification bodies issuing certifications under articles 42 and 43 GDPR takes inspiration, and it is partially built upon such set of rules[153].

In order to be recognised as a certification body, and therefore delivering GDPR certifications, it is necessary to be accredited by the DPA or the National Accreditation Body (NAB). In particular, it is possible to distinguish between three different situations: 1) when the accreditation role is played by the DPA; 2) when the accreditation role is played by the NAB; 3) when the accreditation is issued by both the DPA and the NAB[154]. It has been noted that there is a risk of a potential conflict of interest in the situation where the supervisory authority covers the role of accreditation body, if the same authority is also involved in the certification process. [185] However, regardless of whether the DPA, rather than a NAB, is to deliver the accreditation, the DPA is in charge of a general task. The task at stake is to identify the specific requirements against which evaluate whether a certification body shall receive the accreditation for issuing certification under articles 42 and 43. These specific requirements change if the DPA will also be the accreditation body or if this role is played by a NAB. The DPA shall also make the specific requirements public.

The first situation mentioned above verifies when the DPA is also delivering the accreditation to certification bodies, then the requirements shall only be based on the general points outlined in article 43(2) GDPR. If, on the other hand, the accreditation process is delivered by a NAB the data protection authority shall also draft additional requirements to be applied to certification bodies accredited by NAB. Moreover, the NAB shall also deliver accreditation in compliance with the requirements from Regulation 765/2008 and ISO/IEC 17065:20212.

---

[151] Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93.

[152] ISO/IEC 17065:2012 Conformity assessment — Requirements for bodies certifying products, processes and services, see https://www.iso.org/standard/46568.html. (last access: January 2023)

[153] See recital (100) GDPR.

[154] See article 43(1) GDPR.

Article 43(2) therefore just sets a first set of requirements for being accredited as certification body. These requirements are not data protection requirements, but rather general requirements that essentially deal with independence and expertise of the certification body, the commitment of the certification body in respecting the criteria of the certification scheme, and putting in place procedures for issuing, reviewing or withdrawing the certification. The certification body shall then be able to deal with infringements of the certification scheme, for this, procedures shall be put in place for handling complaints concerning certifications. [146, 184] The additional requirements elaborated by the DPA in case of accreditation delivered by NAB shall essentially focus on data protection aspects, because, on the other hand organisational requirements are defined by Regulation 765/2008 and ISO/IEC 17065:2012. These additional set of requirements is similar to those explicitly indicated by article 43(2) GDPR, but they are more detailed. [185]

The release or the withdrawal of the certification is a decision solely upon the certification body, according to the evaluation carried out during the certification process. Moreover, a certification body's accreditation can last for a maximum of five years. The requirements for being accredited as certification body shall be made public by the supervisory authority and the EDPB. The accreditation body can withdraw the accreditation from a certification body if the conditions are no longer met. [184] It shall also be noted that the EC through implementing acts can specifies technical standards for data protection certification mechanisms, as well as to promote such certification mechanisms.

As a conclusive remark, the GDPR is silent as regards the mutual recognition among MSs of accredited certification bodies. Indeed, according to Regulation 765/2008, once a certification body is accredited in one MS then it shall be recognised as such also in all the other MSs. However, if the accreditation body is the DPA, there is no duty upon such a DPA to recognise certification bodies accredited in other MSs by a NAB, because the GDPR is silent on this point and GDPR is considered *lex specialis* respect Regulation 765/2008. On the other hand, it seems that if a certification body is accredited by a DPA it shall be recognised as such also in all the other MSs. This presumption seems to be stemming from the consistency mechanism according to articles 63 and 64(1)(c). [185]

2.3.3 – State of the Art of Existing Certification Mechanism

At the moment of writing, the EDPB registry on certification mechanisms that has been created according to article 42(8) GDPR contains no record[155]. However, very recently the EDPB has released three opinions on certification mechanisms according to articles 63 and 64(1)(c) GDPR. Two opinions concerned respectively the scheme Europrivacy and the scheme CARPA, both from Luxembourg. The third one concerns the scheme EuroPrise from Germany. The rest of the section will delve into the content of these opinions.

The first opinion concerns the national certification mechanism approved by the Commission nationale pour la protection des données (CNPD), i.e., the Luxembourgish DPA, called "CARPA". [182] This certification mechanism is the first scheme receiving the endorsement of the EDPB under the consistency mechanism pursuant to article 63 GDPR. However, CARPA is not an EU-level CM, since it is applicable only to data controllers and processors established in Luxembourg. Moreover, this certification scheme is not meant to

[155] See https://edpb.europa.eu/our-work-tools/accountability-tools/certification-mechanisms-seals-and-marks_en (last access: January 2023).

be applicable for data transfer to third countries or international organisations under article 46(2)(f). CARPA is a CM directly written by the CNPD. This is a case where the DPA drafted the criteria but at the same time will issue the certification as well, i.e., the DPA is operating as certification body.

CARPA is a general scheme, applicable to whichever sector and domain, however the main subject matter of the scheme are the responsibilities of data controllers and processors (i.e., the accountability principle). Nevertheless, the EDPB has noted that there are some exclusions from the scope of CARPA. Namely, are not covered by CARPA criteria the following situations: "personal data processing specifically targeting minors under 16 years old; processing activities in the context of a joint controllership; processing activities in the context of article 10 GDPR; entities that have not officially designated a DPO (article 37 GDPR)". [182] CARPA is a certification scheme which is supposed to build upon other certification methods and standards[156]. Indeed, the evaluation methods for CARPA's criteria are derived from the International Standard on Assurance Engagements (ISAE 3000 standards)[157], as well as ISCQ1 (quality control of auditing organizations)[158] and ISO 17065 (licensing of certification entities).

The first version of CARPA scheme has been submitted to the EDPB for receiving an Opinion under articles 63 and 64(1)(c), in February 2021. In that occasion, the EDPB has noticed how the first version of the scheme could have posed problems for the consistent application of the GDPR. Therefore, many changes to the certification criteria have been deemed necessary. For example, concerning the terminology used in the scheme, the scope of application of the scheme and the modalities to determine the Target of Evaluation (ToE), the lawfulness of the data processing, or even the respect of principles pursuant to article 5 GDPR. [182] The CNPD has therefore implemented the changes required by the EDPB and, in May 2022, with a decision of the CNPD, the CARPA certification mechanism has been finally approved, becoming the first approved CM in the EU. [186]

The second opinion concerned the "Europrivacy" scheme, which, on the other hand, has become the first European Data Protection Seal approved pursuant article 42(5) by the EDPB, according to its task under article 70 GDPR, in October 2022. [187] The Europrivacy criteria are elaborated by the European Center for Certification and Privacy (based in Luxembourg), which is the scheme owner of the certification mechanism[159]. Therefore, the CNPD, operating as competent supervisory authority, after performing a preliminary assessment of the admissibility of the Europrivacy scheme has submitted the draft scheme to the EDPB for a formal approval, as it is indicated in the EDPB document on the procedure for the approval of EU-wide certifications. [183]

Europrivacy, as well as CARPA, is not supposed to be a certification scheme for the transfer of data to third countries under article 46(2)(f) GDPR. Europrivacy is a general scheme applicable to a wide range of data processing, thus not focusing on a specific sector or domain. However, genetic data processing is not covered by the criteria of the scheme

[156] See https://cnpd.public.lu/en/actualites/national/2022/06/adpoption-gdpr-carpa.html (last access: January 2023).
[157] See https://www.iaasb.org/publications/international-standard-assurance-engagements-isae-3000-revised-assurance-engagements-other-audits-or-0 (last access: January 2023).
[158] See https://www.iaasb.org/publications/international-standard-quality-management-isqm-1-quality-management-firms-perform-audits-or-reviews (last access: January 2023).
[159] See https://www.europrivacy.org/ (last access: January 2023).

for example. The "core criteria" of Europrivacy certification and the "Technical and Organisational Measures (TOMs) check and controls" constitute the main part of Europivacy certification. However, a set of "Complementary Contextual Checks and Controls" make the scheme adaptable to particular technological applications, making the scheme also more scalable and suitable to specific data processing situations[160]. In order to make the set of criteria applicable in different MSs and making of Europrivacy the first European Data Protection Seal, a set of national oriented check and controls is added at the main criteria. The set of additional criteria is called National Obligations Compliance Assessment Report (NOCAR).

The EDPB, in its Opinion addressed to the CNPD, has essentially endorsed all the part of the Europrivacy criteria, with just few smaller remarks, approving the scheme under article 42(5) GDPR.

The last opinion of the EDPB concerns the "European Privacy Seal (EuroPriSe) certification criteria for the certification of processing operations by processors". [188] EuroPrise is a certification scheme developed by EuroPriSe Cert GmbH[161], a legal entity established in Germany, which has submitted the scheme to the Landesbeauftragte für Datenschutz und Informationsfreiheit Nordrhein-Westfalen, the competent German supervisory authority in North Rhine-Westphalia (NRW). The NRW has then asked the opinion of the EDPB under articles 63 and 64(1)(c) GDPR. The EDPB has issued the opinion at stake in September 2022.

The scope of the EuroPrise certification system is not clear enough according to the EDPB, but for sure also this scheme is not meant to provide appropriate safeguards under article 46(2)(f) GDPR. However, the certification scheme should provide sufficient guarantees for a data processor to be in compliance with article 28 GDPR. In this last regard, the EDPB has noted that the CM provides also criteria concerning the relationship between the data processor and a further processor. Therefore, the EDPB suggested to specify that the certification will only be released to the applicant (i.e., the data processor) and not to other data sub-processors. [188]

According to the EDPB opinion, there also seem to be issues regarding the transferring of data towards third countries. Indeed, although the scheme is not suitable for ensuring appropriate safeguards in that sense, it provides criteria that apply also to processors that do not have an establishment in EU or in the EEA. Therefore, the EDPB recommends specifying that every time a data transfer concerning a certified data processor not established in the EU or EEA takes place, it shall be specified that complying with Chapter V GDPR is still mandatory. Moreover, it shall be stressed again that the scheme does not provide appropriate safeguards under article 46(2). [188]

The EDPB is then critical about some requirements concerning the relationship between the data controller and the data processor, as well as about privacy by design and by default criteria in EuroPrise scheme. In general, the opinion of the EDPB concludes that EuroPrise certification, to date, still lack of having full consistency with GDPR requirements, therefore the indicated modification shall be apported to the criteria in order to having them approved. [188]

[160] See also https://www.europrivacy.org/ (last access: January 2023).
[161] See https://www.euprivacyseal.com/ (last access: January 2023).

3 – Role Played by Co-regulatory Instruments in Data Protection Law

As it has been said many times throughout this chapter, CoCs and CMs under articles 40-43 GDPR can be in general defined as tools for compliance. This role is specifically assigned to both the instruments by article 24 GDPR, which states that "adherence to approved codes of conduct as referred to in Article 40 or approved certification mechanisms as referred to in Article 42 may be used as an element by which to demonstrate compliance with the obligations of the controller". It is possible to say then that these instruments cover an accountability function. This function is for sure the main role that both the instruments play, nevertheless there are other functions that can be covered by CoCs and certifications, some of which are explicitly indicated by the GDPR, some others instead are stemming from literature discussion.

3.1 – Co-regulation as Instruments for Compliance

The first function that has been mentioned is the accountability function of CoCs and CMs. This is a role that both the instruments play, as stated in the first place by article 24 GDPR. The accountability function might be divided into two phases: the reaching of compliance, and then the demonstration of it. Neither CoCs nor certifications can ensure compliance or can constitute compliance per se. Although, they are elements that can be used in order to facilitate the compliance activity and its demonstration.

CoCs are supposed to be instruments that further clarify the application of GDPR into specific sector. Therefore, the adoption of a CoC, for sure, should at least help data controllers and processor in their internal activities for reaching compliance with GDPR requirements. CoCs indeed should be used as "rulebook" by data controllers and processors when planning and designing data processing. [145] CoCs should aim at establishing best practices that data controllers and processors can implement in specific situations of particular complexities. In this sense, CoCs are also meant to ensure that compliance is performed in a more cost-effective way. [145] Moreover, CoCs shall be taken into account by the DPA when infringements of GDPR happen. Therefore, those cases when data controllers are called to demonstrate their compliance with specific requirements, such as security measures implementations, the adoption of CoCs shall be taken into account by the DPA. The adoption of a CoC should reasonably also affect the decision of whether to issue a fine or not, and the amount of the fine itself[162].

Certification mechanism, as regard the accountability function, have pretty much the same effects of CoCs. Indeed, almost all the articles of the GDPR which mention the use of CoCs as element to demonstrate compliance with a specific requirement mention certifications as well[163]. However, certifications are statements of conformity, so they certify that the data processing meets some specific requirements, which are in this case the criteria of the scheme, that are then supposed to be consistent with GDPR requirements. [146] However, certifications have a major role in showing accountability to third parties, such as to data subjects, costumers or other business on the market, rather than being a guideline for data controllers and processors about how to internally implement the GDPR.

3.2 – Legal Certainty and Reduction of Legal Fragmentation

[162] Article 83(2)(j) GDPR.
[163] See articles 28, 32, and 46 GDPR.

From another perspective, GDPR co-regulatory instruments can play a role in relation to the consistency and the clarity of data protection law in EU. CoCs indeed are supposed to specify and clarify GDPR requirements in critical sectors. On the other hand, this characteristic does not really belong to CMs. Certifications are indeed not supposed to adapt the interpretation of the GDPR in light of a specific sector. GDPR certifications do not help data controllers in the interpretation of the law, they are rather a tool that demonstrates that compliance has been reached out, they say whether there is compliance or not. Although, they say little about how compliance has been reached. However, it does not mean that it is not possible to develop sectorial CMs. Indeed, certification mechanisms can contain criteria that help data controllers and processors demonstrate compliance especially well into some sectors rather than in others[164].

Moving back to CoCs, they are deemed to play an implementing function, specifying general requirements of the GDPR into specific sectors. By this way, private stakeholders can create their own facultative rules for facilitating the implementation of GDPR in a more effective way. Indeed, there are some data protection issues that are of particular concern in specific sectors and that need further clarification and specification. For example, the concepts of anonymisation and pseudonymisation are considered worthy of being further specified in the domain of biomedical and health related research. [167] CoCs would clarify which is the meaning of these concepts into the specific domain at stake and provide specific best practices to be followed in order to implement them in compliance with GDPR. [167, 168] Moreover, it has been argued that CoCs can also have a positive impact on the problem of legal fragmentation among MSs in some critical domains. [80, 121] For example, the GDPR leaves room for national legislators to introduce further requirements for processing health related data or for data processing in the health domain in general. This has led to a certain degree of fragmentation among the EU about the implementation of data protection law requirements.

In a study commissioned by the DG Health [80] the fragmentation aspect has been particularly stressed, and the development of a CoC has been suggested as a measure to tackle such differences. Nevertheless, it is not clear how in practical terms a CoC could help data controllers and processor overcome different national requirements in terms of data protection. Indeed, as it has been acknowledged in the mentioned study itself, a CoC cannot change or amend any national law provision. However, a CoC could help data controllers dealing with cross-border data sharing in two ways. In the first place, data controllers and processors could be guided by the CoC trough the differences of MS's legislations. Moreover, a CoC could help creating a common understanding of the legal requirements. In this sense, a CoC could contribute to the creation of a common ground in terms of legal interpretation of GDPR rules. It is possible to argue that a sort of "harmonisation function" could be played by a CoC. However, that seems to be far from really harmonising data protection law.

Still concerning legal certainty aspects, the use of CoCs and certification mechanisms can create impacts also on DPAs' activity of legal enforcement. Indeed, the use of such instruments would contribute to the creation of common good practices that are shared

[164] See https://www.i-hd.eu/idhis-information-governance-certification-programme/ (last access: January 2023) for an example of certification mechanisms that focuses on demonstrating compliance with health data processing.

among the members of a group. Moreover, they constitute a gold standard parameter DPAs can refer to when assessing compliance of data controllers. Finally, monitoring bodies (article 41 GDPR) and certification bodies (article 43 GDPR) are called to monitor the adherence of controllers and processors to the CoC or the certification. In doing this, such bodies help DPAs monitor the compliance with the GDPR. Indeed, such bodies are asked to refer to the DPA if during their activity they found any irregularities and violation of the GDPR.[22, 177]

3.3 – Economic Impacts and Marked Related Aspects

Both CoC and CMs are clearly instruments linked somehow to market dynamics. At least in the sense that they are instruments that aim at making of compliance an investment rather than just a cost for organisations. Therefore, like for every other kind of investment, also compliance investments need to generate a return for the investor. Otherwise, companies will juts decline the decision to spend money, time, and human resources on that. Both CoCs and certification mechanisms are costly mechanisms, however they can both generate, at least from a hypothetical perspective, some return on the initial investment for data controllers and processors.

Certification mechanisms are explicitly market oriented instruments. They are meant to communicate to third parties the level of compliance reached by the data controller or processor. This communication is directed to data subjects, which can, by this way, trust the data controller or data processor that has been certified. But the communication is also directed to other stakeholders on the market, such as customers, providers, or competitors. In general certifications are supposed to enhance market transparency. [22] Especially CMs, which are usually paired with marks and seals, aim at remediating to a big problem that characterise the provision of digital and internet-based services: the asymmetric information between users and service providers. Indeed, the users usually do not have the possibility to evaluate some inner characteristics of a service or a product. The level of privacy and data protection of a product or a service is a feature that is difficult to evaluate for an external agent. [105, 189]

Certification mechanisms can help data controllers and processors communicate in business to business relationships, their data protection commitment. In this sense, data protection compliance can even become a market differentiator. [146, 176] However, especially in B2C relationship, it shall not be taken for granted that if costumers are aware of high level of data protection compliance, they will pay a higher price for it. Indeed, even if costumers become aware of a higher level of data protection into a product, they still may choose to buy another product which have less privacy guarantees but provided at cheaper price or with more attractive features. The choice depends on many factors, including the culture of privacy among individuals.

Looking at the CoCs, these instruments are maybe less obviously linked to market dynamics, but anyway their spread is still influenced by the interest that they can generate among stakeholders in terms of economic returns on the initial investment. Indeed, CoC as well as certification mechanisms' success partially relies on how widespread such instruments are among the community. [105] If data controllers and processors do not see any advantage from an economic and market perspective, it seems unlikely that they will rely on these instruments. Anyway, the WP29 has endorsed the idea that data protection levels of compliance can also operate as "market differentiator". [190] In this sense, certification

mechanisms, and to a lesser extent also CoCs, can have a "signalling function", providing competitive advantages to data controllers that decide to adopt such mechanisms. [22, 191]

3.4 – Reasoning on the Sparse Application of Co-regulatory Instruments

In general, the use of these instruments in data protection law has been so far rather limited, although some new CoCs and certifications are starting to emerge on the EU market. For example, two certifications scheme have been approved (CARPA and Europrivacy), and one of them has even been approved by the EDPB as European Data Protection Seal[165]. Moreover, even a sectorial certification for health data processing has been developed (although it has not been approved yet)[166]. CoCs are finding room on the EU scenario with even two examples of CoCs developed for health data processing issues in the context of scientific research[167]. However, it is not possible to affirm that these instruments are playing a key role in the GDPR implementation, yet. The problems related to the use of co-regulatory instrument for compliance with the GDPR can, in general, be related to the following aspects which are discussed throughout the thesis: 1) lack of incentives, especially in terms of returns on the initial investment, for data controllers and processors[168]; 2) complexities related to the procedures of approval and the instruments[169]; 3) unclear legal effects stemming from the application of these instruments[170]. These aspects can be brought as arguments for the limited application of these instruments throughout whichever sector. Nonetheless, some specific limiting factors shall be discussed in light of the specific domain at stake. Therefore, in the next Chapter will be addressed the limiting and the enabling elements of the use of CoCs and CMs in health.

[165] The Europrivacy scheme mentioned above.
[166] The IDHIS scheme mentioned above.
[167] See above Section 2.2.3.
[168] See especially Chapter 3, Section 1.3.6.
[169] *Ibidem*.
[170] See especially Chapter 3, Section 1.3.5.

Chapter 3 – Proposal for a Co-regulatory Model of Governance in Health

1 – Discussion on Co-regulating Health Data Processing

The idea of applying a co-regulatory approach in data protection law has been discussed for years, at least since the adoption of the DPD[171], as it has been argued in the previous Chapter. During the elaboration of the GDPR proposal, there was the impression that co-regulatory instruments were to play an even greater role in data protection law context. Today, however, it seems that such co-regulatory tools are not as popular as hoped by the EC during the preparatory work of the GDPR. [77, 78] In the literature, discussion specifically concerning the health domain and the use of co-regulatory solution for solving compliance issues already took place under the DPD. [167]

There are both arguments that could be brought in support of the use of co-regulatory measures, and arguments that could be used for arguing the contrary. Indeed, if co-regulatory solutions are ideally aimed at smoothing the implementation of data protection law, there are nevertheless some objective obstacles at their widespread utilization.

Looking at the obstacles, a first general concern relates to co-regulatory measures' intrinsic features, such as the high initial costs for developing the instruments, or the difficulties in obtaining the approval by the authorities. [153] A second obstacle is due to features of the health domain in the EU, especially from a legal point of view. An example in this sense are the structural differences in how MSs' organise their healthcare systems at national level[172].

On the other hand, two arguments motivate the use of co-regulatory solutions for enhancing the use and the re-use of health data among EU. These elements could be classified as "motivators" of co-regulation. The first one is related to the nature of the data protection law compliance activities in the health sector (intended as provision of care, as well as health related scientific research). The second one is the necessity to overcome the legal fragmentation among MSs as regards the processing of personal health-related data in medicine and research.

Focusing on the first motivator: a new legal and compliance panorama has been introduced by the interplay between the GDPR and the new data governance strategies. This new governance framework has been envisaged in the communication of the EC "A European strategy for data" in 2020, [14] and implemented, as far it concerns the discussion here, especially by the Data Governance Act (DGA) and the proposed Regulation on the European Health Data Space (EHDS). [110] However, none of the new data governance measures is supposed to amend or modify the GDPR Therefore, coordination between the new legal framework and data protection law shall be investigated, especially when it comes at the secondary use of personal data for research purposes. [115]

---

[171] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

[172] In article 168 TFEU, it is possible to grasp a mix of competences between EU and MSs. However, it is also clear that MSs keep the main competence in terms of organisation on the public health system. On the other hand, the EU can intervene only on limited subjects, see [219].

It has been discussed how the EU policies aiming at enhancing the sharing of data for boosting big data and AI find an obstacle in data protection law[173]. These strategies, more specifically, clash, to some extent, with the implementation of some principles of the GDPR. Therefore, the interplay between GDPR, DGA and the proposed regulation for EHDS is not always straightforward. [16] It should be borne in mind that it would be better to avoid too complex compliance activities concerning the new data governance framework. Otherwise, the actors involved could be disincentivised in engaging with the data sharing activities.

The use of co-regulatory instruments could be crucial in smoothing compliance duties and enhancing the actual functioning, on a practical basis, of the regulatory system. Indeed, co-regulation tools of the GDPR have been envisaged for filling the gap between abstract rules and day by day operations at company level. For example, CoCs have, among their several tasks, also the goal of operationalising GDPR's requirements through the definition of common best practices and guidelines for compliance. [153] In this respect, it could be argued that the new strategy for data will be posing an additional layer of compliance duties upon data controllers and processors. This situation will create a difficult interplay with GDPR compliance, which is already per se complicated in the context big data and AI in health.

The second motivator is constituted by another obstacle on the path to sharing health related data among EU considering the new data strategies, which is the legal fragmentation among MSs. This element has been claimed to be both a motivation for increasing the use of Co-regulatory instruments, as well as an obstacle to the use the same co-regulation approach (at least on an EU-wide level). There is a paradox because it is possible to argue that the use of EU-wide co-regulatory instruments is not, in practical terms, feasible for reasons that will be further investigate in the rest of the Chapter[174]. [80] However, on the other hand, co-regulatory solutions have been deemed by some EU bodies as able of having a sort of "harmonisation function" throughout different MSs legislations. [80, 121, 145, 192] At the same time, the EU institutions themselves have recognised that legal fragmentation cannot be totally overcome by co-regulatory instruments. [80] Indeed, the harmonisation function of these instruments – which is mainly assigned to CoCs, rather than to certification mechanisms – can be related just only to compliance procedures or interpretational activities of the GDPR in light of specific contexts and sectors. On the contrary, is clear that co-regulatory solutions, cannot create a real legal harmonisation at EU level. Therefore, it is not clear how, in practical terms, an EU-wide CoC, can help reduce legal fragmentation among MSs, despite the fact in some cases EU institutions seem to claim so. [16, 121]

The rest of the discussion carried out in this Chapter starts from the general assumption that co-regulation can have a positive effect, at least to some extent, on the implementation of data protection law in health domain. This assumption is deemed valid, at least from a theoretical perspective, on the basis of the literature analysis carried out on this argument. The assumption is then corroborated by the effort of EU institutions to promote this kind of regulatory approach both in general in the EU [108, 130, 131, 155, 193–195], as well as in the specific domain of data protection law. [16, 80, 121]

[173] See Chapter 1, Section 3.2.
[174] See below Section 1.1.

Starting from the assumption that integrating co-regulatory solutions is a desirable goal, the rest of this Section will analyse the elements that could have an impact on the actual use and diffusion of the instruments. The aspects further discussed here are either elements that constitute an obstacle to the use of co-regulatory instruments, or just elements that shall be carefully evaluated for shaping at best the co-regulatory tools. The next sub-section will at first delve into the geographical and material scope of a CoC or a CM in the GDPR context (Section 1.1). Afterwards, the potential impacts that the use of these instruments could have on the behaviour of organisations adhering to them, but also on the ecosystem in general, will be object of discussion (Section 1.2). These are considered the two main elements that would influence the most the effectiveness of co-regulatory instruments as they are now envisaged in the GDPR. The last sub-section, eventually, analyses the concrete expectations and the limits of CoCs or CMs as they are now envisaged into the GDPR. Therefore, here the focus will be more on structural aspects of the legal framework surrounding CoCs and CMs (Section 1.3). In particular, the following aspects are touched: fine reduction (Section 1.3.1); compliance facilitation (Section 1.3.2); legal fragmentation (Section 1.3.4); presumption of conformity (Section 1.3.5); Cost for developing and adopting the instruments (Section 1.3.6); differences and similarities between CoCs and CMs (Section 1.3.7).

The second Section of the Chapter will, on the other hand, just focus on the co-regulatory instrument deemed more appropriate for addressing data protection law issues in health, i.e., CoCs (Section 2). In light of the new data governance framework and the implementation of innovative technologies, the main clutches with data protection law that could be addressed by a CoC are reported here. The aspects touched concern the following topics: anonymisation and pseudonymisation (Section 2.1); data re-use for scientific research (Section 2.2); data subjects' consent and data altruism (Section 2.3); data protection roles and new rights in DGA and EHDS (Section 2.4); legal basis for primary and secondary uses of data in the health domain (Section 2.5); data minimisation, purpose and storage limitation in the big data context (Section 2.6). Finally, some preliminary conclusions pertaining to the discussion carried out in the first three chapters are drawn in Section 3.

1.1 – Geographical and Material Scope

The scope of CoCs and CMs under GDPR is an element to consider from a geographical perspective, as well as in terms of the subject matter of instrument. The geographical scope of the instruments affects the spread of the instruments itself, and therefore its capacity to influence organisations' behaviour. The wider the geographical scope the higher is the number of organisations that will probably adhere to it. Having a high number of organisations adhering to a co-regulatory instrument allows it to develop new best common practices that can contribute to the proper application of data protection law.

On the other hand, the material scope constitutes the subject matter covered by the instruments, i.e., the GDPR requirements that could be specified (CoC) or that could be certified (CMs). Therefore, also the material scope influences the practical effects that CoCs and certification mechanisms produce in the data protection domain. Indeed, each context has its own data protection issues and risks. The data processing carried out in the financial domain will be characterized by compliance issues that are different to those carried out in the health domain. Even within each of these contexts there would be many different "sub-sectors" presenting different data protection issues.

In the health domain is plenty of data protection compliance issues that could be addressed by a specific CoC or certification mechanism. One of the most critical areas in health domain is personal data processing for research purposes. [120, 125] However, it shall be noted that research in medical field is very fragmented in terms of types of research activities that can be carried out, in terms of data processed, and even in terms of individuals involved. [125, 196, 197] Consequently, also the data protection issues will be different. It has been argued in the literature that the data processing could take different shapes in the medical field so that is even difficult to design boundaries and categories on a conceptual level for such data-related phenomena. [48] Essentially, almost each situation has its own unique features in such a way that is difficult to create fixed definitions and categories for big data phenomena in the health domain[175]. Moreover, Lachaud argues that the scope of CoCs is wider than the scope of CMs. The material scope of CoC is not, according to Lachaud, limited to data processing. Otherwise, a CoC could concern whichever aspect that can contribute to a better implementation of the GDPR. [153]

As regards the geographical scope, the GDPR allows stakeholders to develop these instruments either at national or EU level. It could seem straightforward that the best solution is to develop these instruments at EU-level in order to spread gold-standard as much as possible. However, also from this perspective some further arguments could be brought to the discussion. The need to have an EU-wide co-regulation mechanism is essentially motivated by the fact that data processing is changing, and part of the changes are led by the spread of big data technologies. Data processing in big data context often take place in complex networks composed by different actors who share data in a continuous flow. Often the nodes of the network are situated in in different countries. [137] This feature argues for having an EU-wide co-regulatory system that is applicable to all nodes of the network. Especially in the context of health research, sharing data among different research centres situated in different MSs is crucial. For example, in clinical trials or observational studies there are often many research centres participating in a study. Therefore, it is key ensuring the smooth flow of data among them. For example, translational research [197] has the objective to gather insights from clinical trials data and translate the knowledge for improving preclinical and fundamental research in drug discovery. [167] This kind of research looks for new ways to translate laboratory evidence in order to carry out diagnostic and care activities. Therefore, it is necessary to share data outside the study protocol among several research actors; this research activity often also implies that such data processing are not covered by the initial consent, which is, on the other hand, collected at the beginning of the study. [167] Therefore, in terms of GDPR legal basis to be used for these data processing, different solutions shall be found other than consent of the patient. [80] Research centres shall then to deal with different legal specifications at national level for sharing personal data in compliance with the law.

However, developing EU-wide instruments implies receiving an approval from the EDPB and making the provisions of the instrument suitable for every MSs' legal system. Moreover, in the case of CoCs, it is necessary to demonstrate that the instrument has been developed by organisations that are representative enough in the sector where the CoC is supposed to be applied[176]. [145, 146] Nevertheless, both EU institutions [121, 145, 192]

[175] See above Chapter 1, Section 1.1.1.
[176] Article 40(2) GDPR.

and part of the literature [137], seem to suggest that the adoption of an EU-wide co-regulatory instrument could be the path towards having a smoother sharing of health data among Europe. The development of an EU-wide CoC, or certification mechanism, is however problematic because of the differences between MSs in terms of data processing requirements in the health domain. It would be therefore necessary to develop an instrument that take into account the legal fragmentation among MSs. The study commissioned by the EC for assessing the implementation of GDPR in MSs' health system indeed recognises the necessity to consider such differences, if an EU-wide CoC is developed. [80]

Looking at the material scope, it has already been observed how the GDPR in articles 40-43 does not preclude any principles or requirements from being the subject matter of a co-regulatory instrument. Moreover, as said in the previous chapter[177], some requirements of the GDPR are explicitly indicated as possible subject matter of a CoC or a certification mechanism[178]. However, in the medical field, there are some well-known compliance issues. Looking at the few CoCs developed so far for the medical field, it is possible to detect some potential crucial content of a co-regulatory instruments. The relevance of such potential contents will be discussed in the following Sections[179].

In the choice of the material scope of a co-regulatory measure for the health domain is important to understand that the health system is characterized by many different data processing. Especially the secondary use of data is carrier of data protection compliance issues. [125, 198] The issues are due to the big data and AI technologies necessity of a huge amount of heterogenous data to be processed in order for these technologies to perform properly. Such data are indeed collected for different purposes and the data processing are often not included in traditional study protocols. [167, 197] Moreover, as it was discussed before, big data in health domain can take different "shapes"[180]. In terms of the number of data subjects involved, but also in terms of the depth of the observations carried out on a person. There could be therefore data processing concerning a large number of data subjects where, although, the features are not observed in depth. For instance, EHR systems concerns data from a large number of individuals, but the categories of data collected could be just a few, such as the name and demographic data. On the other hand, clinical trials on rare disease could involve much less individuals, but the number of factors observed will be higher[181]. [48] These different features shall be reflected in the co-regulatory instruments.

1.2 – Impacts on Stakeholders' Behaviours

The use of co-regulatory solutions in data protection law is relatively new. Nevertheless, the DPD, as said in the previous chapter[182], already foresaw the use of CoCs. However, their adoption has always been sparse and sometimes detached from the rationale of the

[177] See Chapter 2, Section 2.
[178] See articles 40(2), 25, 28, 32, 35 and 46 GDPR, see also EDPB guidelines on CoCs and certification mechanisms [145, 146].
[179] See further Section 2.
[180] See Chapter 1, Section 1.1.
[181] See above Chapter 1, Section 1.1.1.
[182] See above Chapter 2, Section 2.1.

DPD[183]. The GDPR has amplified the role of co-regulatory instruments as compliance tool in data protection law. However, their use and spread so far is not as popular as hoped[184].

Because of the reasons mentioned, forecasting the actual impacts of these instruments on the behaviour of organisations, and on the society in general, is not an easy task. In part because of the novelty of the use of these instruments in the domain at stake, in part because of the need of empirical analysis that create evidence against which to verify the hypotheses. However, it is possible to borrow some reasoning developed from other sectors. For example, in the sector of sustainability the use of co-regulatory instruments has been affirmed for many years. Certification systems in this domain have been applicable for years and some analogies can even be outlined between environmental and sustainability policies and data protection and technology implementation. [105]

A lesson learnt is that this kind of instruments are successful only as long as they can influence a behavioural change among organisations at sufficient scale. [199] For example, if a standard or a certification mechanism is too costly, or it is too difficult to reach compliance with it, it would be difficult to make its adoption financially viable. Therefore, adhering to such a standard will just be too difficult for organisations, and its spread will probably be limited. Consequently, also the costs for being certified against the standard will be high. On the other hand, setting the bar too low will not induce any incentive in organisations to keep improving once the certification is delivered. [105] That is what is called "adoption paradox". [105] CoCs and CMs under the GDPR are potentially expensive instruments, because their adoption entails a set of costs, such as auditing activities, the adoption of new technical and organisational measures, or just changes in the day-by-day inner operations of organisations. So far, the spread of these instruments has been rather limited, therefore it has not been yet possible to build a system that relies on economies of scale, in order to reduce the costs linked to the adoption of these instruments. Another aspect that has been observed in the sustainability domain, is that if there is the possibility of having many different co-regulatory instruments in competition among themselves. In this case, it has been argued that there could be a "race to the bottom" among providers of co-regulatory instruments. [105] Essentially, it means that developers of co-regulatory instruments could be tempted to develop instruments that are easy to comply with in order get more shares of the market respect to their competitors. Indeed, the more a standard does not entail big adjustments in organisation's operations, the more the organisation will be willing to adhere to such standard. [105]

However, this specific risk is mitigated under the GDPR co-regulatory system because of the existence of the approval mechanism by the DPAs. Indeed, according to EDPB guidelines [162], it is not possible to lower down the level of protection for individuals into the content of CoCs and certifications. Therefore, the bar must not go below to the minimum level of protection granted by the GDPR. Moreover, one of the requirements for a CoC or a CM to be approved is to demonstrate that the co-regulatory instrument contributes

[183] See above Chapter 2, Section 2.1.1 on the Italian case concerning the transposition of DPD provisions on CoCs.

[184] See the registries where all the approved CoC and certification mechanisms are supposed to be published https://edpb.europa.eu/our-work-tools/accountability-tools/register-codes-conduct-amendments-and-extensions-art-4011_en and https://edpb.europa.eu/our-work-tools/accountability-tools/certification-mechanisms-seals-and-marks_en (last access: January 2023). It shall be noted that not all the CoC and certification seem to be promptly published in such registries.

to a proper and consistent application of the GDPR. [146, 162] Throughout this mechanism it should be granted that the use of co-regulatory instruments in data protection law will lead to a grater implementation of GDPR's provisions. By this way, speculative behaviours by organisations, meant to use CMs and CoCs just for gain benefit in terms of compliance demonstration, though without engaging in the actual improvement of their inner processes should be reduced. Anyway, a CoC or a CM that does not bring the bar of compliance enough high, probably will not generate the desired changes in terms of stakeholders' behaviour towards data protection law commitment.

Another aspect that can potentially catalyse positive behaviour in terms of data protection compliance are the "positive spill-overs". Positive spill-overs is a phenomenon that consist in the adoption of positive behaviours and practices also by organisations that decide not to formally adhere to the standards in CoCs or CMs. [199] Indeed, once approved by the DPAs, the content of CoCs and certifications is supposed to be made publicly available[185].

There are some examples of CoCs and certifications the content of which has been made available by the scheme owners itself, in an altruistic perspective, even before the receiving the approval from a DPA or the EDPB[186]. In this scenario, even organisations not adhering to the CoC, or CM can improve their compliance. Nevertheless, the other side of this phenomenon is that there are some actors, so-called "free riders", that would benefit from the system, without participating to its costs. Indeed, such organisations would benefit from the elaboration of CoCs and certifications. Because they can use these instruments for ameliorating their compliance activities and procedures, gaining benefits from different perspectives. But they would not bear the costs related to the elaboration of these instruments or those related to be formally subject to them, such as auditing costs. [24] Therefore, it shall be borne in mind that are not only the "co-regulated" entities, i.e., data controllers and processors who formally adhere to CoCs or CMs, that determines the impacts on the system. On the contrary, the "non-co-regulated" entities as well contribute to the final impacts. Therefore, the mosaic composed by different CoCs and certifications mechanisms, and the interplay between the two kinds of instrument in data protection law, will determine the impacts on stakeholders' behaviours. [105] Crucial is also the role of these instruments in light of the hard law legislation that regulate the domain, i.e., the GDPR. In general, the hard law instrument is supposed to grant the essential level of protection to rights and freedoms, while on the other hand, the co-regulatory instruments are supposed to facilitate reaching compliance. The co-regulatory instruments are then supposed to reward virtuous behaviours, stimulating positive approach towards compliance. The hard law measure, on the other hand, does not reward, but entail mechanisms for punishing those who do not even reach the minimum level of compliance with hard law. [105]

In conclusion, developing hypothesis about the impacts that the use of co-regulatory solutions can have on the actual behaviour of stakeholders is key, but extremely difficult. Even in domains, such as the environmental law and sustainability, where the use of certification and co-regulation is way more consolidated, it is still difficult to reach such conclusions. [200, 201] The reasons are related to struggles in collecting empirical data about

[185] See articles 40(11), 42(8) and 43(6). See also [145, 146].

[186] See, for example, the Italian certification scheme Inveo ISDP 10003 available at https://www.inveo.com/certificazione-isdp-10003-2020-data-protection (last access: January 2023). See also CoCs on cloud computing elaborate from CISPE.cloud, or EU cloud CoC. [233, 234]

the behaviours of companies and organisations, as well as to difficulties to have counterfactuals[187] against which to evaluate the impacts. Some indirect outcomes that these co-regulatory instruments should generate can though be detected in the literature. [105, 200, 201] In this regard, co-regulatory instruments are supposed to 1) contribute to the development of new agreement among stakeholders, while at the same time reducing the conflicts; 2) provide room for experimentation and innovation; 3) introducing best practices that could be then scaled up for shaping policies or regulatory measures. [105]

1.3 – Realistic Expectations from Co-regulation

Great expectations about the role played by these co-regulatory instruments seem to be envisaged in the work of some EU institutions and bodies. [16, 78, 80, 121] However, the use of regulatory strategies other than top-down regulation is often object of critiques from scholarship. [24, 154] This is also due to some failures in regulating markets' behaviour using such instruments in other domains, such as the sustainability domain or the financial sector. [199] Moreover, data protection law co-regulatory instruments present themselves some structural problems in the way they are envisaged in the GDPR[188]. Finally, has it has been discussed throughout the previous sections, the health domain presents some legal and structural constraints that limits the use and the impacts that these instruments can have on the whole ecosystem of stakeholders. It is important then to understand which are the limits and the boundaries within which such instruments can produce benefits.

1.3.1 – Reduction of Fines-related Risks

For sure, CoCs and CMs can potentially reduce the risk for organisations to be fined, or at least they can reduce the amount of the fine, according to art. 83.2(j) GDPR. [24] Fine reduction is an important advantage that organisations that process personal data can benefit from. However, probably, this benefit should be balanced with the costs of adopting the instruments borne by the organisation. As it has been said above, the cost of getting a certification, for example, has been noted to be linked to its adoption rate within a sector. [200, 201] Under this perspective, the more the standards are spread and popular the less they are expensive for organisations. However, it has also been noted that currently there are only few examples of CMs and CoCs approved, or even just under development[189]. Looking at the health domain, there are even less examples in this sense. Only few CoCs concerning the processing of health data have been adopted at national level, while only a project of certification for health data processing is on the table[190]. Probably, organisations

[187] Counterfactuals are Information of how the impacts could have been in absence of the co-regulatory measures. [201]
[188] See also Chapter 2, Section 3.4.
[189] See Chapter 2, Sections 2.2.3 and 2.3.3.
[190] At national level, to my knowledge only two CoCs concerning the health domain have been approved: the Italian CoC for the re-use of health related data for scientific publication and educational purposes available at https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9535354 (last access: January 2023) and the Spanish CoC on Processing of Personal Data in Clinical Trials and Other Clinical Research and Pharmacovigilance Activities, available at https://www.aepd.es/en/informes-y-resoluciones/code-of-conduct (last access: January 2023). As regards certification mechanisms, the only GDPR certification scheme (not approved yet) that focuses exclusively on health data processing is IDHIS – Information Governance Certification

will try to carry out a trade-off evaluation between the costs of adopting a co-regulatory measure and the risk to be fined.

The reduction of the risk to be fined, as well as the reduction of the amount of the fine itself, as real incentive to adopt the instrument, shall also be evaluated against organisations' perception of the likelihood to be fined. In this sense, an element to consider is the number of fines issued by DPAs, as well as the level of risk enshrined into the data processing carried out by the organisation. Indeed, the higher the risk for data subjects, the higher is the fines and the probability to be fined, in case of non-compliance. The health domain is typically considered as a high-risk domain because of the sensitivity of the data processed.

The propensity of the competent DPA to issue fines and the fear of being subject to them is a complex discussion. Indeed, it should include statistics about the different patterns of sanctions issued by DPAs in EU, as well as analysis on the severity of the sanctions in monetary terms[191]. There are some DPAs that are issuing more sanctions than others on average, but this is due to many factors, such as the dimension of the country, or the type of economic activities typically carried out in the MS at stake. For example, there are MSs where the DPA is extremely active in terms of number of sanctions issued, such as Spain or Italy[192]. However, in these countries the monetary sum of the fines is lower than other in countries. In these other MSs, in front of a higher monetary sum of the fines, there are just few sanctions issued. Of course, it means that there are countries where though the DPA issues many fines, the economic value of them is limited. On the other hand, there are countries that issue a lower number of fines but with greater economic impacts.

The situation depicted above is essentially due to the very different economic ecosystems of the MSs in EU. Countries like Luxembourg or Ireland, despite smaller than other countries in geographical terms, have a higher concentration of companies that carry out critical data processing or that process data on large scale. Indeed, looking at the statistics, it is possible note that the highest fines have been issued against big tech companies, such as Google, Amazon, or Meta. For example, in Luxembourg the Commission Nationale pour la Protection des Données (CNPD) has issued 23 fines[193] for a total of € 746,273,600. However, a single fine alone of € 746,000,000 against Amazon, constitutes most of the total of such an amount. On the other hand, as seen above, the Italian or the Spanish DPAs even though issuing many sanctions do not have a high total sum of fines. These patterns are clearly related to the type of economic ecosystems of each MSs, as said before.

It is also interesting to investigate how the fines are distributed among different sectors. The sector where more fines have been issued is "industry and commerce". However, the sector where the highest sum of monetary sanctions has been issued is "media, telecom and broadcasting". The healthcare sector is at the sixth position in terms of number of sanctions issued after: 1) industry and commerce; 2) media, telecom and broadcasting; 3) public sector and education; 4) individuals and private associations; 5) finance, insurance and consulting). As regards the total sum of fines issued, the healthcare is even lower in the ranking of sectors (at the eighth position). Therefore, healthcare seems to be a sector where there is

Programme available at https://www.i-hd.eu/idhis-information-governance-certification-programme/ (last access: January 2023).

[191] Data is from: https://www.enforcementtracker.com/ (last access: January 2023).

[192] *Ibidem*.

[193] *Ibidem*.

a medium number of sanctions issued, moreover the sanctions are not extremely high from a monetary point of view[194].

The analysis should also take into account that the health sector in EU is mainly constituted by public sector bodies, or public undertaking. The core activity of such organisation is not getting an economic advantage from some ethically blamed uses of personal data (e.g., marketing or user profiling), like it happens in other domains, such as media and telecommunication. Nevertheless, also in the health domain many private actors operate, such as pharmaceutical companies or private health care providers. Moreover, the health sector is characterised by very sensitive data and high-risk data processing. In conclusion, it could be argued that nonetheless the health domain is a sector with high-risk data processing, according to the data[195], it does not seem to be a domain where DPAs concentrate their power to enact sanctions the most (especially in terms of monetary sums). Even though a quite high number of sanctions is issued in the health domain, their monetary sum is not extremely high. All these factors can influence the decision of a data controller or processor to adopt a co-regulatory instrument for reducing the risk to be fined or at least reducing the sum of the fine in case of sanction. It is not easy to determine whether data controllers and processor operating in the health domain would be prone or not to adopt such instruments in light of the threat to be fined by DPAs. However, the health sector is a domain where the reputation of organisations is quite relevant, especially for maintaining a trust relationship with patients and the public in general. Scandals involving misuse of health-related personal data are always perceived as something serious, as it happened in the NHS and Google-DeepMind case. [202, 203]

1.3.2 – Facilitation of Compliance Activities (focus on SMEs)

Another realistic benefit stemming from the use of such instruments is the facilitation of compliance activities into an organisation. Throughout the provision of user-friendly procedures for reaching compliance, certification mechanisms and, especially, CoCs can make compliance procedures easier for organisations. As it was discussed before, CoCs are also supposed to provide specification to general requirements of the GDPR. In this sense, CoCs could become very attractive to organisations for defining some principles with difficult interpretation, as well as procedures for complying with them. In this perspective, the GDPR has also tried to move the attention towards SMEs, stating that CoCs (but also CMs) should in particular take into consideration the need of SMEs[196].

The difficulties for SMEs to adopt CoCs has been considered one the weak points of these instruments under the DPD. [24] However, under the GDPR this aspect could change, and it seems more realistic that SMEs could get advantages from the use of CoCs rather than from certification mechanisms. Public health care providers can, from this point of view, be compared to SMEs. Indeed, usually in the health domain, public organisations face shortage of resources to be dedicated to data protection law compliance. Therefore, while big tech or pharmaceutical companies have all the resources to reach out compliance, on the other hand, public hospitals could struggle in the fulfilment of their accountability duties.

[194] *Ibidem*.
[195] *Ibidem*.
[196] See recital (98), articles 40 and 42 GDPR.

These are essentially the reasons why in literature has been argued that CoCs could be instruments more suitable for SMEs, or organisations with limited resources to dedicate to compliance, than CMs. [153] Indeed, for CoCs under GDPR the monitoring procedures seem to be less burdensome than the ones foreseen for certification mechanisms. The latter is indeed carried out by third-party certification bodies through an auditing system inspired by ISO/IEC 17065:2012 international standard[197]. On the other hand, the former is essentially a monitoring activity carried out by accredited monitoring bodies usually identified by the CoC itself. The less strict auditing system, if compared to certifications' one, could make of CoCs a more attractive instruments for some types of sectors and businesses. [153] However, the GDPR foresees in detail the tasks and the features of monitoring bodies. Namely, the accreditation by the DPA, the requirements in terms of independency, expertise and professionality to be satisfied, are identified by GDPR and EDPB guidelines. When the monitoring body is internal to the organisation of the CoC owner, these requirements are more severe. [145]

In particular, according to Lachaud [153], a substantial difference between CoCs and certification lies on the burden of proof during the compliance assessment. Indeed, in CoCs, Lachaud argues, is a task of the monitoring body to verify *ex-post* that the adhering organisation comply with the CoC. As regards certifications, on the other hand, the wording of articles 42 and 43 seems to put the burden of proof upon the adhering controller or processor, in an *ex-ante* evaluation. If the *ex-ante* assessment is not successfully fulfilled by the controller or processor, then the certification seal or mark is not awarded.

In light of the abovementioned observations, there could be debatable follow ups effects. Indeed, there could be a situation where organisations adhering to a CoC and those adhering to certification could get very similar benefits from these instruments in terms of compliance demonstration and facilitation. However, while organisations adhering to certifications shall have passed through a rigorous *ex-ante* auditing activity, organisations adhering to CoCs could not have been subject to any assessment from the monitoring body. [153] Nevertheless, the GDPR states that CoCs shall include mechanisms that allow the monitoring body to verify the conformity of the controllers to CoCs' requirements. Therefore, also CoCs shall anyway be potentially auditable. On the other hand, choosing to adopt a CoC, rather than a CMs, could also have some shortcomings in terms of communicating on the market the compliance commitment, since CoCs do not automatically come along with a label or seal[198]. However, in terms of communication and demonstration of compliance before the DPA, CoCs and certifications produce the same effects, as elements that can be used to demonstrate compliance with GDPR. [153]

In conclusion, it shall be noted that the reference to SMEs in GDPR and EDPB guidelines on CoCs is not straightforward. Indeed, as it has been noted in the literature, [153] the wording of recital (98) GDPR could be interpreted as meaning that a CoC could either be developed for a specific sector or for SMEs in a horizontal way. Indeed, the possibility to draft even cross-sector CoCs is provided by the EDPB guidelines on CoC. [162] Therefore, in theory there seems to be room even for a cross-sector CoC which focuses exclusively on SMEs' needs. This option is nevertheless problematic, and its actual admissibility should

[197] ISO/IEC:2012 – Conformity assessment —Requirements for bodies certifying products, processes and services, https://www.iso.org/standard/46568.html (last access: January 2023).
[198] This is on the other hand the case in certification mechanism.

be further clarified by the EDPB itself. More realistic seems to be the development of a sectorial CoC which focuses, in part or exclusively, on SMEs of a specific sector.

Moreover, it is still not clear according to which logic a CoC, in practical terms, should help SMEs. For example, the EDPB guidelines suggest that CoCs shall be "appropriately scaled to meet the requirement of micro-organisations in addition to small and medium enterprise". However, as it has noted by Lachaud [153] the EDPB does not provide further explanation of the meaning of this concept of "scaling the requirements". It could mean that the objective is "scaling" the level of the requirements, i.e., less burdensome requirements apply to SMEs. Otherwise, "scaling" the requirements could be understood in terms of the nature of the requirements, i.e., the CoC would adopt a different approach with totally different types of requirements for SMEs. Reducing the severity of requirements for SMEs could however be problematic, since sometimes even small businesses process sensitive data, especially in the health domain, maybe even using potentially risky technologies.

1.3.4 – Legal Fragmentation

Doubts remain on the possibility of co-regulatory instruments under GDPR to bridge the gap of legal fragmentation between MSs' different implementation of data protection law. This feature, usually assigned to CoCs, is often claimed by EU institutions and bodies as a real scenario. [80, 121, 162] However, even in the documents of the EU institutions is not always straightforward to understand how this should be happening in practical terms. The legal fragmentation between MSs concerning data protection requirements in the health domain is a problem stemming from the room that MSs have in terms of additional requirements that can be add in some data processing context, as discussed above in Chapter 2, Section 3.2. However, such differences are directly enshrined in the national law implementing GDPR provisions. Creating differences in terms of legal basis required for secondary re-use of data, or special procedure to follow vis-à-vis the DPAs. These aspects cannot clearly be addressed by a CoC or a CM.

1.3.5 – Presumption of Conformity and Burden of Proof

Another crucial point concerns the legal consequences of adopting CoCs and certification mechanisms under articles 40 to 43 GDPR. As it has been said in the previous sections, article 24 GDPR assigns to CoCs and CMs the role of "element by which to demonstrate compliance with the obligations" of the GDPR. From the reading of article 24 GDPR it seems that a presumption of conformity is stemming from the adherence to one of these instruments. However, article 24 leaves some doubts about the correct legal interpretation. Indeed, the article at stake considers these instruments only as an "element" of compliance demonstration. This "element" shall then apparently be combined with other elements of compliance in order to satisfy the accountability principle. Moreover, as stated in article 42(4), the adoption of certification mechanisms, does not reduce the responsibility of the controller or of the processor, and especially does not have any effect on the tasks and powers of the DPA. Therefore, it seems that CMs and CoCs, as envisaged in the GDPR, do not provide a full presumption of conformity comparable to the presumption of conformity that, for example, harmonised standards provide under the NLF context. [22, 180] However, in literature the interpretation seems not to be homogeneous. For example, Lachaud [154] seems to consider certification mechanisms and harmonised standards as producing

the same legal effects. However, other authors, such as Kamara and Leenes, seem to be on another page. [22, 144, 176]

A further element that differentiates CoCs and certification mechanisms from harmonised standards, is the impact on the process of fine imposition of the DPA. According to article 83(1)(j), as it has been discussed above[199], the DPA shall evaluate the adherence to such instruments when evaluating whether to issue a fine, and to which amount the monetary sum should correspond. In this sense, the EDPB argues that the adherence to a CoC or a CMs could either be an element that reduces the fine, or an element that raises the fine amount. [146] Therefore, adhering to a co-regulatory instrument could even become an element that affect in *pejus* the compliance demonstration activities. For example, if a data controller claims to follow a CoC or a certification mechanism, but then after the investigation of the DPA there is evidence that this was not the case, then this could be an aggravating circumstance. [145, 146]

Uncertainty on the legal effects in terms of presumption of compliance could negatively affect the use of such instruments. The legal value of the presumption of conformity is per se unclear in the EU legal system. For example, when it comes to the use of harmonised standards for obtaining the CE mark. In this case, obtaining the CE mark allows manufacturers to place a product on the market. Nevertheless, the manufacturer remains liable, even if the CE mark has been obtained after a third-party assessment. [154]

The EDPB clarifies that obtaining a certification does not constitute compliance with the GDPR by itself. Indeed, demonstration of compliance means also being able to produce supporting documentation, in form of written documents and reports. The documentation should describe how the certification criteria are met. More specifically, such documentation shall contain "reasons, arguments, and proofs resulting from the application of criteria". [146] As regards CoCs, their impact on the burden of proof for compliance demonstration is even more cryptic. Indeed, article 40 GDPR, classify CoCs as instruments for the proper application of GDPR requirements taking into account specific features of different sectors. Article 40 does not identify the demonstration of compliance as the first role of CoCs. Nevertheless, the accountability function of CoC can be inferred from article 24(3) and recital 77 GDPR. Here, the purpose of demonstration of compliance is pursued by certifications as well as by CoCs. The EDPB itself in its guidelines on CoCs [162] states that CoCs have the potential of being element for demonstrating compliance with the GDPR. Therefore, also the adherence to a CoC shall be taken into account by the DPA in evaluating the opportunity to issue an administrative fine, as indicated in article 83(2)(j). This is confirmed also by the WP29 Guidelines for the application of administrative fines as endorsed by the EDPB. [204] In the guidelines, it is said that the adoption of a CoC is an element that the DPA shall consider when deciding whether to issue an administrative fine. Under this perspective, in case of a breach of GDPR requirements, the DPA could even decide that the sanctions applicable by the monitoring or certification bodies (e.g., exclusion or suspension from the CoCs or certification mechanisms) are effective, proportionate or dissuasive enough. Therefore, the application of administrative sanction could not be deemed necessary.

The guidelines on CoCs confirm what stated in EDPB guidelines on certification, i.e., the adherence to a CoC or certification could even be an aggravating factor. [146] Indeed,

[199] See above Section 1.3.1.

the violation of the requirements of these instruments can reasonably indicate the intentional character of the data controller or processor's violation of the GDPR. [204] Anyway, the sanctioning powers of certification and monitoring bodies are without prejudice to the tasks and powers of the DPA. It essentially means that the DPA is not obliged to take into account the sanction already adopted by the monitoring or certification bodies while evaluating to issue a fine to data controllers and processors. Therefore, it is confirmed that the actual legal effect produced by the adoption of a CoC or a certification mechanism in terms of burden of proof and sanction reduction are not straightforward and there is no certainty that adherence to these co-regulatory solutions will lead to benefits during a procedure before the DPA. However, Lachaud [153] noted that the system for monitoring the adherence to CoCs seems to be less strict than the one envisaged for certification mechanisms. Essentially, both instruments, as discussed above[200], seem to produce the same (blurry) legal effects in terms of demonstration of compliance.

1.3.6 – Costs Linked to Co-regulation Development and Adoption

It is necessary to highlight again that the GDPR had made mandatory for these instruments to be approved by either the DPA or the EDPB in order for these instruments to produce their effects. The requirement established by the GDPR to have these instruments approved is crucial. This aspect allows CoCs and certification mechanisms to be allocated in the definition of co-regulation, instead than just self-regulation. It is of the utmost importance to have a check on the content from the authorities for ensuring that fundamental rights and freedoms are respected and in general that the protection ensured through the GDPR is not reduced. [145, 146]

The approval procedure from the DPA or EDPB is also meant to ensure that CoCs and certifications actually bring an added value to GDPR implementation. The mere reproduction of the GDPR requirements into a CoC or certification is not allowed. Despite CoCs and certifications can play different roles, EDPB Guidelines set some specific requirements to be satisfied for submitting these instruments to a DPA or the EDPB. Namely, CoCs shall meet particular data protection needs, facilitate the GDPR application or specify its application. [145] While, in case of CMs, they shall provide a sound system of auditable criteria which specifies the objectives and the implementing guidance for reaching compliance with the GDPR. [146] The GDPR has made the approval mandatory because, among other things, under the DPD some free-rider behaviours have been noted in the literature. [21, 24, 177]

The path to get a CoC or a certification mechanism approved by the DPA or the EDPB could be a long and complicated journey. There are examples of CoCs and CMs that have begun their procedure for obtaining the approval, without though reaching the goal, already under the DPD. For example, the BBMRI-ERIC project for developing and EU-wide CoC for health data processing for research purposes started in 2015 without that the CoC has never been approved so far[201]. Another example of a never approved CoCs, despite its value, is the "Code of practice on secondary use of medical data in scientific research projects"[202]. [167]

[200] See Chapter 2, Section 3.2.
[201] See http://code-of-conduct-for-health-research.eu/ (last access: January 2023).
[202] Available at: https://www.imi.europa.eu/sites/default/files/uploads/documents/reference-documents/CodeofPractice_SecondaryUseDRAFT.pdf (last access: January 2023).

Looking at certification projects, it is possible to notice the length of the procedure to obtain the final approval from the DPA or the EDPB. The first certification mechanisms approved has been the CARPA certification scheme[203]. [182] The works for drafting this scheme started in 2018 and the final approval arrived only in 2022. The Europrivacy scheme has been the second certification scheme approved and the first with EU-wide applicability[204]. The first draft of Europrivacy criteria have been released just after the GDPR came into force in 2018, but its approval only arrived in 2022. The EuroPrise scheme[205] has recently received the Opinion of the EDPB[188], despite the first version of this scheme were circulating already for a while in Europe. [180]

The path to receive the approval for a CoC or a CM is a process that can reasonably last years. This long period is due to the necessity of carefully evaluating the content of an instruments that could be, if not conceived in the proper way, detrimental for individuals' fundamental rights. However, on the other hand, the long and difficult process could discourage organisations from developing such instruments.

A different aspect, but still related to the one just discussed, are the costs that organisations that decide to adhere to an approved CoC or a certification mechanism shall bear. Indeed, adhering to CoCs and CMs, despite being a potentially cost-saving strategy, because of several reasons (for example, like already said in this work, because of the reduction of the risk to be fined, or the facilitated compliance procedures[206]), it could also be a source of high initial costs. There is a paradox, even though only apparent, between the initial costs of adopting a CoC or a CM, and the fact that these instruments are ultimately supposed to make organisations save costs. The paradox is claimed here to be only apparent because these initial costs shall be seen as an investment, with the goal of obtaining an even higher revenue from different points of view[207]. However, such return on the initial investment is not easy to evaluate and it shall be supported by sound empirical evidence. It seems safe to assume though that organisations will not engage themselves in developing, or even just adhering to such instruments, if they do not perceive a concrete final gain in compensation of the high initial costs. Elements, such as the unclear and long procedure to get a CoC or a CM approved, or the unclear legal effect in terms of burden of proof for compliance demonstration, are all aspects that could hamper the use of these instruments. Comparing GDPR co-regulatory instruments with harmonised standards in the NLF can be useful in this situation for understanding the abovementioned problems.

Harmonised standards are part of the NLF and operate as technical specification that implements essential requirements stated in the body of a legislative measure[208]. Harmonised standards are developed by private organisations commonly called Standards Development Organisations (SDO). Therefore, in order to specify general requirements enshrined into a piece of legislation, harmonised standards are developed and put at the

[203] See https://cnpd.public.lu/fr/professionnels/Certification/gdpr-carpa.html (last access: January 2023).
[204] See https://www.europrivacy.org/ (last access: January 2023).
[205] See https://www.euprivacyseal.com/EPS-en/Home (last access: January 2023).
[206] See above Section 1.3.1.
[207] See above Section 1.3.2.
[208] See https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en (last access: January 2023).

disposal of the addresses of the law for reaching and demonstrating compliance[209]. It shall be noted that compliance can also be reached without relying on harmonised standards. However, in this case, a greater production of documentation is required to the addresses of the law, which are usually manufacturers of a product that is to be placed on the market. Under the NLF the law does not define the details of its requirements, but only the essential parts of them, i.e., the essential goals of the law (like of many requirements of the GDPR do). The details of the requirements shall either be decided by the manufacturer by itself, or by following harmonised standards. Like the CoCs and certification under GDPR, harmonised standards shall be approved by an authority, in this case by EC itself[210]. However, adopting harmonised standards, differently than CoCs or CMs, provides a real presumption of conformity to the manufacturers. Therefore, following harmonised standards can generate concrete benefits in terms of documentation that shall be produced, as well as in terms of controls from third party conformity assessment. In this sense, it is possible to argue that harmonised standards are costly procedures for a manufacturer, but with clear and well-defined benefits.

It seems that the EU legislator has chosen to rely on CoCs and CMs taking inspiration from the NLF, and by the role played by harmonised standards into it. However, the mechanism for compliance demonstration through co-regulatory measures in GDPR seems to be less strong than its counterparts in the NLF. Of course, data protection law domain is different to products safety regulation, (where harmonised standards are usually applied). Therefore, a complete comparison is not meaningful. However, it shall be noted that the lack of clear effects in terms of compliance demonstration of these elements could push organisations not to engage with GDPR's co-regulatory instruments. The high initial costs related to auditing and to the change of inner procedures, as well as to long discussions with DPAs for having CoCs or certification mechanisms approved, might discourage the adoption of these instruments.

1.3.7 – Differences and Similarities Between CoC and Certifications

In conclusion of the Section, it can be said that the two instruments, CoCs and CMs, will probably have different impacts and effects on data protection law domain. The future of the two co-regulatory instruments is going to depend on how different these instruments have been envisaged in the GDPR, but also on the different perspectives from which the instruments can be evaluated.

In the first place, it shall be noted (again) that CoCs are not supposed to be paired with seals and marks, whilst this is the case for CMs. Therefore, CoCs are equal in terms of compliance demonstration before the DPA, in light of a less strict monitoring system, and for this reason they may be more attractive than certifications mechanisms. On the other hand, CMs have a more powerful role in communicating compliance on the market. The EDPB guidelines concerning certification mechanisms stress that these instruments are supposed to enhance the transparency in data controller and data subjects' relationship, as well as in B2B situations. Certification mechanisms seems to have then a clearer and more

[209] See https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en#:~:text=A%20harmonised%20standard%20is%20a,to%20one%20of%20these%20organisations (last access: January 2023).
[210] See https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en (last access. January 2023).

explicit role from a market perspective. While, on the other hand, CoCs seems to be more similar to the classic definition of co-regulation measures, i.e., introduction of an additional layer of facultative provisions that further shape the data protection law panorama. [155]

The expectations about the potential effects of these instruments and their potential grade of adoption in EU pass also through these considerations. It shall however be noted that, especially concerning CoCs, their role is not straightforward, therefore it is not easy to forecast their future application in the EU. For example, even though CoCs are deemed to rather be a clarification and specification tool for principles and requirements of the GDPR, the EDPB stress also their role in terms of compliance communication for enhancing trust among data subjects and the public in general. [162]

Looking at the health sector, and at the processing of health data for research purposes, CoCs are not only supposed to clarify some compliance activities for data controllers, but also the information to be provided to data subjects, the safeguards to be applied to this processing or specific security measures. But, on the other hand, the adoption of such compliance activities is made more transparent thanks to the CoC. In this sense, patients and the general public could be confident that their personal data are processed in a proper manner thanks to the adoption of a CoC by institutions participating in the research project.

It seems that the role played by CoCs and CMs can overlap in some occasions. This feature of co-regulatory instruments of the GDPR can become attractive in some situations, but on the other hand could be carrier of confusion and ultimately jeopardize the spread of the instruments. For example, there is no prohibition for data controllers or processors to communicate, for example through their websites that they have adopted a CoC, and leverage on this for gaining competitive advantage on the market. In this sense, there could be confusion between the role played by seals and marks in certification mechanisms and this other kind of communications stemming from the adoption of a CoC. Unfortunately, in the EDPB guidelines on CoCs there is no indication about this possible overlapping with certification mechanisms. The lack of clear rules on this point could have detrimental effects on the market with negative consequences also on data subjects' trust towards these instruments.

2 – Possible Contents of Co-regulatory Instruments (focus on CoCs)

The processing of health-related personal data is a sector where the use of co-regulatory instruments could be particularly useful. This is due to essentially two reasons. The first one is the difficulty to reach and demonstrate compliance with GDPR in light of the complexities of the sector. The second one is the lack of resources to be dedicated to compliance activities in many of the organisations that usually operate in the health domain. Therefore, the health domain is a context where there could be ground for developing co-regulatory instruments. The present work argued that it is necessary to facilitate compliance with data protection law in the health domain, especially when some critical purposes are pursued, such as scientific research. This is even more necessary in light of the new strategy for data launched by the EC. These necessities have been perceived by the EU institutions and the adoption of co-regulatory instruments has been suggested in different occasions. [80, 121] Usually, the co-regulatory tool suggested is a CoC, which is supposed to deal with health data processing. However, in the work of the EU institutions, the precise content of a CoC

is not specified, but, in theory, different topics of interest could be touched to facilitate health data processing in compliance with the GDPR.

The use of a CoC as co-regulatory instruments is motivated by the fact that several notions and tenets of data protection law are of difficult interpretation and application in the health domain, especially when it comes to research activities. Given the heterogeneity of the data processing in the health domain, and the different necessity that could come up in light of the specific purposes pursued, different aspects could be the subject matter a CoC. Indeed, the projects for developing a CoC for health data processing have been extremely different to each other[211]. The purpose of this part of the work is not to list all the possible contents for CoCs in the health domain, but some potential contents that a CoC could address are indicated and briefly analysed in the rest of the Section. The list is not exhaustive and is elaborated looking at the literature works on this matter. The potential topics are elaborated taking into account the main data protection issues in light of the processing of big data and the application of AI systems in health.

The content of a CoC, according to article 40 GDPR, may concern any requirement or principle of the GDPR. Therefore, several aspects of data protection law could potentially be part of a CoC pursuant articles 40 and 41 GDPR. Nevertheless, some requirements seem to be more prone to carry uncertainty in the specific sector of health.

These are the critical points in terms of data protection compliance identified:

- Anonymisation and pseudonymisation of personal data
- Data re-use for research purposes, especially in light of the new data governance framework (Data Governance Act and European Health Data Space proposal)
- Consent collection from data subjects, also in light of the new data altruism consent
- Definition of the role of the actors involved in light of data protection law (e.g., data subjects, data controllers, processors, joint controllership) and the new data governance framework (data holders, data sharing providers, data altruism organisations, etc.)
- Legal basis for personal data processing
- Respect of some data protection principles (e.g., data minimization, purpose limitation, accuracy, storage limitation) in big data and AI context

It shall be stressed again that even though here it is suggested to start form these aspects as possible content of a CoC, this is not an exhaustive list.

2.1 – Anonymisation and Pseudonymisation of Health Data

Anonymisation and pseudonymisation are techniques meant to reduce the possibility to link a piece of information to an identified or identifiable natural person. These techniques are largely used in the health domain, especially for research purposes, in order to protect patients' privacy. [63]

Anonymisation is meant to obtain a high degree of certainty that the data is no longer linkable to a data subject, trying to reach as much as possible the impossibility to re-identify the individual in absolute terms. In other words, the objective of anonymisation is to break in a definitive way the logical and technical link between a piece of information and the individual. [71, 205–207] On the other hand, pseudonymisation is not meant to be definitive

[211] See Chapter 2, Section 2.2.3.

and absolute. The goal of pseudonymisation is rather to make the data just temporarily not linkable to an individual. Namely, the re-identification is supposed to be still possible, but only for those who hold the "re-identification keys". Such keys are the additional information that, if matched to the pseudonymised data, make the individual identifiable again. The re-identification keys shall be kept by those who carry out the pseudonymisation process (either the data controller or data processor) in a secure manner. The keys shall be protected with adequate technical and organisation measures, with the aim to make for third parties impossible re-identify the individuals. [71, 205, 208]

The two techniques, despite similar from an engineering point of view, produce different results from a legal viewpoint. Indeed, in data protection law domain, the two techniques play different roles and produce different legal effect upon who is carrying them out. As regards anonymisation, if properly performed – i.e., in case the risk of re-identification of the data subject from anonymised datasets is deemed acceptably low – it will bring the data processing out of the material scope of the GDPR. On the contrary, pseudonymisation usually do not bring the data processing out of the material scope of the GDPR. Pseudonymisation is indeed considered a mere technical measure for the protection of personal data pursuant to article 32 GDPR. However, a recent decision of the EU General Court[212] has reversed the common approach followed by the EDPS and EDPB so far. Which used to state that, by-default, pseudonymised data cannot be considered anonymous information. The decision at stake rules, on the contrary, that even though if the re-identification keys are not deleted, when data are transferred from a controller to another organization, the recipient organization does not receive personal data if it has no legal means to access the re-identification keys and therefore re-identify the data subjects[213].

Both anonymisation and pseudonymisation find large application in the health domain, especially for scientific research purposes. The two techniques could be part of a CoC because different interpretative aspects could be clarified, leading to more legal certainty for different stakeholders[214].

Moreover, in light of the proposed EHDS, anonymisation and pseudonymisation will play a central role in the new governance framework for data sharing in EU, especially for secondary uses of electronic health data. Indeed, like said in Chapter 1[215], the EHDS when promoting the re-use certain categories of electronic health data held by PsBs, relies on anonymisation and pseudonymisation as technical measure for building the secure environment within which the re-use of personal data is allowed. Under the governance framework envisaged by the DGA and EHDS, indeed, the re-use of personal electronic health data should take place using anonymised data as first choice[216]. Only if the purposes of the re-users cannot be carried out through anonymised data, in case some conditions are

[212] Judgment of the General Court (Eighth Chamber, Extended Composition) of 26 April 2023. Single Resolution Board v European Data Protection Supervisor. Case T-557/20 ECLI:EU:T:2023:219.
[213] The new interpretation could be of particular interest for the research domain, when datasets are shared for secondary uses in a pseudonymised form. In this case if the research re-user commit itself, through a contract for example, not to re-identify individuals, GDPR provisions could not apply to the data processing.
[214] Article 40(2)(d) GDPR explicitly mentions pseudonymisation as a potential content of a CoC. As regards anonymisation, the Opinion 5/2014 of WP29 on anonymisation techniques suggests that anonymisation techniques could be specified by a CoC.
[215] See Chapter 1, Section 3.1.2.
[216] Articles 44, 47 EHDS proposal.

respected, and after a due motivation from the re-users, the re-use through pseudonymised data is allowed[217].

However, it has been noted that these concepts could be subject to different interpretation, depending on several factors. For example, interpretation might change according to the MSs where the processing is carried out or the type of research purpose. [209] Anyway, anonymisation could be a tricky goal to reach because of two reasons. In the first place, it is not straightforward to understand which is the level of anonymisation considered acceptable by DPAs, as it will be discussed later in the following two sub-sections. Moreover, applying strong anonymisation techniques on personal data processed for health research purposes could excessively reduce the value of data, making it not usable for research purposes. [210]

The EHDS make of anonymised data the first choice for having access to electronic health data for secondary uses. Only if the purposes cannot be pursued using anonymised data, then the data user can ask for having access to pseudonymised data. In this sense, the EHDS is in line with the GDPR as regards secondary uses of personal data for research purposes under article 89(1). Indeed, this article provides for the use of non-personal data as long as the purpose can be achieved with such data. However, anonymisation is not a panacea. First of all, simply because it could happen that the research purposes cannot be achieved using anonymised data. The EHDS itself provide for using pseudonymised data if some conditions are respected instead of anonymised data. Besides that, anonymisation interpretative issues relating to the meaning of the concept of anonymisation, as well as to practical procedures through which performing anonymisation could hamper the implementation of the technique in compliance with the GDPR. Both these aspects have already been addressed in CoC projects in the health domain. [167]

The conditions under which a piece of information can be considered anonymous – i.e., not linkable anymore directly or indirectly to an identified or identifiable individual – have been debated for long time in literature. [60, 62, 65] The reason of the great debate is essentially linked to the lack of clarity, both in the data protection legislation and in the interpretative activities of the DPAs, on this theme. Especially in light of the spread of big data in modern society, it seems difficult to evaluate when a piece of information can be traced back to an individual in an objective way. If a strict interpretation of the legal provisions is adopted, potentially almost every piece of information can be deemed personal. [62] Recital 26 GDPR provides indeed that in order to understand whether a data is personal or not, it is necessary to carry out an analysis of the probability that the data can be linked to a physical person. In doing so, all the means that can be reasonably used to identify a person shall be taken into account. The means reasonably likely to be taken into account should include costs and time to be dedicated for re-identifying a natural person, the technological state of the art and future possible technological developments. Two documents enacted by the WP29 contribute to the interpretation of this concept in data protection law. [71, 211] However, without solving all the doubts.

In a nutshell, the concept of anonymisation and the status of anonymous data seems to be strongly related to a case-by-case evaluation that change over time and according to the type of data at stake, as well as to the context into which the data processing takes place. [71] It seems that a risk-based approach shall be adopted when evaluating the appropriate

[217] See article 44(3) EHDS proposal.

data anonymisation techniques. The aim is to reach a result in term of anonymisation where the risk of re-identification is low enough to be considered in compliance with the GDPR and the orientations of DPAs. Moreover, as it has been argued many times in literature, in health research context, even data processed through anonymised data can be subject to ethical concerns and can pose threats to natural persons. [167] This risk is essentially due to the high sensitivity of health data for the life of patients.

However, anonymisation poses problems not only in terms of the expected result, but also in terms of the procedures to follow from a technical and organisational perspective. Indeed, because of the principle of accountability, a data controller or processor is asked to demonstrate, by showing the due documentation, how the anonymised result has been achieved. Because of the risk-based approach, promoted especially by the WP29 [71], it will be necessary to provide proofs that the anonymisation technique adopted is suitable for the data processing at stake, in consideration of the data processed and the purposes pursued. All these evaluations shall then be included into a risk assessment concerning the probability to re-identify the data subjects.

The choice of the correct anonymisation technique is not an easy task however, especially for health data processing. Different techniques have been proposed for anonymising personal data for health research purposes in the literature [72, 167]. Moreover, the WP29 proposes a list of state-of-the-art anonymisation and pseudonymisation techniques. Despite the fact the Opinion has not been update since 2014, it still provides the main reference in terms of DPAs' interpretation.

As regards the concept of pseudonymisation, the EHDS indeed strongly relies on this measure, especially if the secondary uses of data cannot be pursued through anonymised data. According to article 44 of the EHDS proposal, a data user can also ask for having access to pseudonymised data. However, the health data access body keeps the key for re-identify the pseudonymised data and the data users shall commit itself not to try to re-identify data subjects from the pseudonymised data he got access through the health data access body. However, the data user shall provide an explanation on the impossibility to pursue the objectives through anonymous data, as well as which could be the ethical consequences of processing only pseudonymised data[218].

The concept of pseudonymisation is less critic that the anonymisation one, but still not free of doubts. It has been debated in literature if pseudonymised data shared with an organisation that do not have access to the re-identification keys and that commit itself not to try the re-identification of the data, can be considered anonymised data. [212] In this sense, the concept of relative anonymity has even been brought on the table of a CoC. [167] This aspect have acquired even more relevance in light of the recent judgement of the EU General Court mentioned above[219], which seems to support the concept of anonymity in context.

In conclusion, also pseudonymisation aspects can be object of a CoC for health data, especially in terms of guiding the choice about the most appropriate technique to use. Indeed, for pseudonymisation techniques, as well as for anonymisation ones, the only official reference at EU-level is the WP29 Opinion, that has not been updated since 2014.

[218] Recital (50) Article 45 EHDS proposal.
[219] Judgment of the General Court (Eighth Chamber, Extended Composition) of 26 April 2023. Single Resolution Board v European Data Protection Supervisor. Case T-557/20 ECLI:EU:T:2023:219.

2.1.1 – State of the Art of Anonymisation and Pseudonymisation Techniques

A list of the main anonymisation and pseudonymisation techniques is indicated in the Opinion 05/2014 of the WP29 [71], which, as said before, is the main official reference since 2014.

The Opinion at issue indicates the following anonymisation techniques:

- Randomisation techniques (noise techniques, permutation, differential privacy)
- Generalisation techniques (aggregation and k-anonymity, L-diversity/T-closeness)

As regards the pseudonymisation techniques, these are indicated as state of the art in the Opinion:

- Encryption with secret key
- Hash function
- Keyed-hash function with stored key
- Deterministic encryption or keyed-hash function with deletion of the key
- Tokenization

It would be necessary to update the state of the art of these anonymisation and pseudonymisation techniques in order to provide a more useful guidance for data controllers and data processors. Moreover, it would be useful to contextualise the use of such techniques in the health care context. This contextualisation should be meant to detect techniques that fit the health context better than others. There is already a vast literature on the use of anonymization and pseudonymization in the health domain, see, for example: [56, 63, 72, 213, 214].

2.1.2 – Anonymisation and Risk Management

The anonymisation process is deemed to be properly performed as long as the result is data that is no longer linkable to the original data subject. In this case, data does not fall into the material scope of the GDPR. Scholars have discussed two different possible interpretations of the concept of anonymisation under the GDPR. [64, 66, 215] The first one adopts an "absolutistic approach", it means that the anonymised data shall not be linked to the initial data subject at all, with absolute certainty. The second orientation, on the other hand, adopts the "reasonable risk approach". According to such a second approach, it is not necessary that the data is no longer linkable to the data subject in an absolute manner, but even a reasonable low chance of re-identification is accepted.

The first interpretation is also called "zero-risk approach" and it is stemming from an interpretation of some wordings of the Opinion 05/2014 of the WP29. The Opinion seems to implicitly require for the anonymisation process to be irreversible in an absolute manner for both the data controller and any other third party. However, it shall be noted that such a requirement comes up only in some points of the Opinion[220]. Such an interpretation clashes with practical problems. For example, the impossibility for a data controller to be aware of all the information necessary to establish with certainty that no third party is technically able to re-identify the data subjects, for example using additional information

[220] See [71] for example at page 3 "anonymisation results from processing personal data in order to irreversibly prevent identification" or at page 6 "The underlying rationale is that the outcome of anonymisation as a technique applied to personal data should be, in the current state of technology, as permanent as erasure, i.e. making it impossible to process personal data". See further on the contradictory passages the Opinion at stake [64].

retrieved from other data sources. Moreover, such an interpretation clashes with the concept of risk itself. Indeed, the concept of risk and risk analysis never require of bringing to zero the probability that the "feared event" would happen in the future. Finally, a zero-risk approach could lead to an excessive application of the material scope of the GDPR. If any information falls into the scope of the GDPR that would lead to the impossibility to enforce the law. [62]

The second interpretation, which is actually the only one relevant and that is also embraced by most of the mainstream doctrine, foresees that the re-identification risk shall be reduced to an acceptable level, and it shall be calculated on the basis of some parameters. These parameters are the "means" (for example, time, costs, state of the art, future technical innovation) that a motivated third party, or the data controller itself, can reasonably deploy for re-identifying the data subjects[221]. Therefore, the re-identification risk evaluation should be inspired by a subjective approach. Following a subjective approach essentially means that not all the possibilities objectively existing to re-identify a subject can be considered. [61, 216]

Even though this approach is the only interpretation applicable to real word scenario, it does not seem to be endorsed to the same extent by all the DPAs across EU[222]. The Italian DPA, for instance, does not always have a consistent approach, at least from a terminology point of view, concerning the re-identification risk[223].

From a different perspective, embracing the "acceptable risk approach", uncertainty stems from the practical procedures for carrying out the analysis and the management of the re-identification risk. Is not easy indeed, for a data controller, understanding how to build a risk analysis meant to calculate and manage the probability that an individual would be re-identified. In theory, it could be created a risk analysis system based of some "feared events", "threats" and "impacts" in order to calculate the likelihood that the event "re-identification"[224] verifies and to evaluate the severity of its impacts on data subjects' rights and freedoms[225]. [71]

2.1.3 – Anonymisation and Pseudonymisation in Light of GDPR's Principles

Data anonymisation is considered a further data processing under the GDPR. [71] It means that principles of article 5 GDPR shall be applied also to the anonymisation process, at least in theory. Moreover, the compatibility test under article 6(4) GDPR shall be successfully carried out also for further data processing concerning data anonymisation. [217] In light of this observation, it seems that anonymisation, differently to pseudonymisation,

---

[221] Recital 26 GDPR.

[222] Different approaches can be noticed between the French DPA (more absolute approach) and the UK DPA (more relativistic approach). See [205, 228] for the Information Commissioners Officer (hereinafter ICO) and https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles (last access: January 2023) for the CNIL.

[223] See for example, GPDP, doc web 2020 – 9356568, p. 12; GPDP, doc web 2018 – 8998319, p. 3; GPDP doc web 2018 – 8997404, p. 4; where the GPDP requires that anonymisation will be "irreversible", or that personal data would be made "anonymous in a definitive way".

[224] ISO 31000:2018 identified the following elements as expression of the concept of risk: 1) risk sources (threats), 2) potential events (feared events), 3) consequences (impacts), 4) likelihood (probability). [89]

[225] Opinion 05/2014 indicated three "feared events" that could potentially lead to the re-identification of the data subjects: linkability, singling out and inference, see p. 11-12.

is not just a security measures, and this is due to the fact that anonymisation leads the data out of the GDPR material scope. Pseudonymisation, on the other hand, is explicitly listed among the security technical measures that data controllers can apply to reduce the risk enshrined into a data processing[226]. It is not clear, in other words, whether the data controller or data processor can freely apply anonymisation techniques to fulfil some principle of the GDPR, such as the security of the processing, minimization, storage limitation. Otherwise, if anonymisation can only be considered as a personal data processing based on an appropriate legal basis under articles 6 and 9 GDPR. If this approach is followed, anonymization shall also be carried out only in light of a specific purpose detected in advance and communicated to the data subjects.

2.1.4 – CoC on Pseudonymisation and Anonymisation

A CoC concerning data anonymisation and pseudonymisation could have impacts on: 1) detecting and clarifying which are the anonymisation and pseudonymisation techniques more appropriate for the specific context, starting from the state of the art of such techniques; 2) providing indications as regards the cases where is more indicated to use anonymisation rather than pseudonymisation, also in light of their different roles under the GDPR and the new data governance legislation; 3) providing guidelines about how to build the risk analysis procedures; 4) ethical issues stemming from the use of anonymised or pseudonymised data, also in light of the EHDS requirements for data users; 5) quality and value of data after anonymisation for research purposes.

2.2 – Data re-use for Scientific Research

The data re-use for scientific purposes in the medical field is subjects to tensions between the necessity to re-use big amounts of data (especially collected for care purposes) and some principles of the GDPR which aim at limiting the re-use of personal data to some conditions. This trade-off is exacerbated by the spread of new technologies that revolve around the collection and the re-use of big data for purposes that is not always possible to detect in advance. This new paradigm of data use is the so-called data driven decision making process[227]. [31, 139]

The re-use of personal data for scientific purposes, even if formally supported by the GDPR[228], is often, in practice, difficult. In this case, a significative compliance burden relies on some stakeholders in the medical filed. Other problems are also brought by additional requirements adopted by national legislators, thanks to the room left by the GDPR in the context of health data processing. For example, the Italian legislator has introduced in articles 110 and 110-*bis* D.lgs. 196/2003 a mechanism essentially based on consent as the main legal basis for research purpose in the medical, biomedical and epidemiological filed.

To boost the sharing and the re-use of data for medical research, the EC has recently adopted a new strategy for data. [14] The purpose of the strategy, as said many times in this work, is to increase the exploitation of big data analytics and AI technologies. Among

[226] Article 32 GDPR.
[227] See above Chapter 1, Section 1.1.2.
[228] Article 5(1)(b) establishes that the re-use of data for scientific research should be considered as compatible with the original purposes, as long as the conditions under article 89 are respected. To demonstrate the respect of article 89 conditions is not always easy for data controllers.

the instruments stemming from the initiative, the DGA has already been adopted as a regulation with the aim to boost the sharing of personal and non-personal data held by public sector bodies, especially for general interest purposes, such as scientific research. Moreover, stemming from the initiative of the EC is the "Common European Health Data Space" (EHDS). The EHDS will have the task to specify some requirements of the DGA into the domain of electronic health data. [110]

The new strategy of the EC is meant to introduce mechanisms to incentivize the sharing and re-using of data. Examples in this sense are the "data sharing services" intermediaries that will have the task to facilitate the sharing of data, enhancing also trust among parties involved. Additionally, a new form of consent, i.e., "data altruism consent" has been introduced. This new consent form will be used for altruistic re-use of data from natural and legal persons[229]. Anyway, the new measures of the DGA and EHDS are not supposed to amend any GDPR's requirements or principles. Therefore, organisations involved in the new data governance framework shall still be compliant with GDPR requirements. From this last perspective, it has been noted by the EDPB and EDPS how there is a lack of consistency between the GDPR, the DGA, and the EHDS. Such a lack of alignment could bring to legal uncertainty and difficulties in reaching compliance. [16] Namely, looking at the recently adopted Joint Opinion of the EDPB and EDPS on the EHDS it seems that further clarification shall be provided in the context of data re-use. The EDPB and EDPS had indeed noted that according to article 34(1) of the proposed EHDS there is a broad array of purposes, for which electronic health data can be re-used[230]. The lack of definition of these purposes could create a difficult interplay with GDPR mechanism for data re-use for scientific research purposes, and consequently also a difficult interplay and compatibility with article 9(2) GDPR. [15] However, for sure, the EHDS provides interesting opportunities for biomedical research, including the possibility to process electronic health data for testing, training and the evaluation of algorithms in medical devices, AI systems and digital health applications[231].

An open question left by the current version of the EHDS proposal concerns the inclusion of for-profit research projects. [115] The inclusion of this type of research projects in the room provided by the EHDS for secondary uses of data is still not clear, as also noted by the EDPB and the EDPS in their Joint Opinion. [15] Indeed, only recital (41) of the EHDS states that secondary uses of data for general interest include also research carried out by "public, private and non-profit entities". On the other hand, the operative part of the EHDS (i.e., the articles) is silent on this point, leaving doubts on a subject which is object of discussion since the DGA and the GDPR approval. [114, 115]

Another critical point concerns the terminology about the re-use of data. Namely, about the meaning of the concepts of "general interest" and "public interest". The DGA, setting the grounds for the new data governance framework, states that the re-use of personal data shall pursue purposes of general interest, especially in the context of data altruism. Then, the EHDS expand the concept to the secondary uses of electronic health data. For example, recital (41) reads: "access to data for secondary use should contribute to the general interest of the society". However, the wording "general interest" is not used in data protection law.

[229] See Chapter 1, Section 3.1.
[230] Especially under article 34(1)(f) and (g) EHDS proposal.
[231] Article 34(g) EHDS proposal.

The GDPR, uses the terminology "public interest" when it comes to the legal basis for processing personal data[232]. Therefore, doubts are raised whether the two concepts are interchangeable or if there are any differences among them. [15, 16, 111, 114, 115] At least, the differences from a terminology perspective could create confusion. Moreover, neither the DGA nor the EHDS provide a definition of scientific research in the general interest. The previous version of the DGA also included examples of research falling under the notion of "general interest"[233]. Unfortunately, the approved version of the DGA includes no more such explicative references.

In order to understand the concept of scientific research and shed some light on its possible relevance in terms of general interest, it is necessary to look at the GDPR. According to recital 159 GDPR the purpose of scientific research should be interpreted in a "broad manner", including, but not limited to, technological development and demonstration, fundamental research, applied research and privately funded research, studies conducted in the public interest and in the area of public health. Despite the recital at stake mentions privately funded research, there has been debate in literature whether the for-profit feature of research could affect the public interest character of research. [114, 115]

As regards the legal basis for secondary uses, it shall be noted that the EHDS will constitute an EU law providing suitable and specific measures to safeguard rights and freedoms of data subjects under article 9(2)(h) and (j) for processing personal data for public interest in the area of public health (article 9.2(h)) and for scientific research purposes (article 9.2(j)). In this last case, it shall be noted that such legal basis can only be used in conjunction with article 89(1) GDPR. According to article 89(1), appropriate safeguards shall be implemented for using such a legal basis. The safeguards shall include both technical and organisational measures, including minimization and pseudonymization, or even anonymisation if the purpose can be achieved in this way. The EHDS seems to be in line with this approach requiring indeed anonymisation or at least pseudonymisation[234] and that data users can access data only within secure environments[235]. [114, 115] Nevertheless, which are the specific safeguards to be implemented is a decision left to the data access body.

It can be noted that at national level, MSs can introduce further conditions according to article 9(4) GDPR for the processing of genetic data, biometric data and data concerning health. Such further conditions shall be coordinated with the data processing that will be carried out in the EHDS for the purposes indicated in article 34 EHDS proposal, which on the other hand will be based on article 9(2) exemptions stemming from EU law. [15]

2.2.1 – CoC for Data Re-use for Biomedical Research

Data controllers and data processors could need additional guidelines about the re-use of health data in light of the new data governance mechanisms. The points to be clarified about the secondary use of electronic health data in DGA and especially the EHDS are numerous. In the section before some examples have been provided. Any of them could be part of a CoC; for example, the inclusion of a specific research project in the set of research that can be considered of general interest seems to be a topic that will carry uncertainty among stakeholders.

---

[232] Article 6(1)(e) GDPR.
[233] See Recital (35) DGA proposal.
[234] Articles 44 and 47 EHDS proposal.
[235] Article 45 EHDS proposal.

The specific safeguards to implement, pursuant to article 89 GDPR could be specified in a CoC. The choice of the appropriate legal basis under article 6(1) and 9(2) GDPR for processing data in the EHDS is another important topic. Such guidelines could be clarified into a CoC, as also suggested by the EDPS. [121] Moreover, also the coordination between data processing carried out for the purposes under article 34 EHDS proposal and the further condition introduced at national level under article 9(4), could be clarified in a CoC.

2.3 – Data Subjects' Consent and Data Altruism

DGA and EHDS has formally recognised the concept of data altruism into a piece of legislation, therefore another scenario for data re-use has been introduced. Data altruism is defined by article 2(16) of the DGA as:

"the voluntary sharing of data on the basis of the consent of data subjects to process personal data pertaining to them, or permissions of data holders to allow the use of their non-personal data without seeking or receiving a reward that goes beyond compensation related to the costs that they incur where they make their data available for objectives of general interest as provided for in national law, where applicable, such as healthcare, combating climate change, improving mobility, facilitating the development, production and dissemination of official statistics, improving the provision of public services, public policy making or scientific research purposes in the general interest".

It could be noted the concept enshrined into the data altruism framework, i.e., sharing data without a compensation, is not a novelty per se. The idea of sharing personal health information for scientific research on altruistic basis is a well-known scenario among researchers. [111] Collecting data from patients willing to share it because they have a particular commitment in promoting research is key opportunity. This scenario can indeed help researchers overcome different obstacles, such as the limits of anonymisation of health data. However, relying on altruistic commitments can also have shortcomings. For example, a higher risk of bias relies on dataset composed of data collected from a certain set of population that share some motivations in sharing their data. [218] Moreover, the GDPR itself gives the opportunity to data subjects to provide consent for processing personal data for scientific purposes, without receiving a compensation whatsoever. [16]

However, the DGA has introduced a new type of data intermediation service just for the data altruism sector. Data altruism organisations are legal entities that seek to promote objectives of general interest, making available personal and non-personal data[236]. A key point is that data altruism organisations shall not establish commercial relationship with data subjects or data holders that provide data to them. By this way, data subjects and data holders can provide respectively their personal or non-personal data to data altruism organisations. The provision of data is meant for the creation of data pools to be used for purposes of general interest or common good, such as the purpose of scientific research[237].

Data subjects can provide their consent through a data altruism form. Such a consent will be collected by data altruism organisations that operate as data intermediaries for data shared through this modality. The role of data altruism organisations is not totally clear, even though it seems they shall act with the general objective to grant trustworthiness in the system and transparency. In order to enhance trust and transparency, data altruism

[236] See Recital (1) DGA.
[237] Recital (21) DGA.

organisations are required to respect some requirements, such as processing personal data within secure environments and in respect of ethical standards[238].

The data altruism consent shall be used for the collection of personal data from natural persons for purposes of "general interest", among which there is also scientific research, as said above. Nevertheless, the DGA specifies in recital (50) that personal data collected within the data altruism context shall be considered based on consent under article 6(1)(a) or 9(2)(a) GDPR, according to the procedural rules for consent under articles 7 and 8 GDPR. Therefore, the DGA does not introduce a new type of consent, or a new legal basis for processing personal data, but it relies on existing rules for consent under GDPR[239]. The EDPS has then provided for the opportunity to use a different legal basis under article 6(1) for the secondary use of electronic health data, namely the public interest purpose (6.1(e) GDPR)[121]. In this case, it could be possible to process data in the context of data altruism, relying on a legal basis other than consent, nevertheless it seems that the data altruism consent shall be collected anyway.

The EHDS proposal brought only few more clarifications about the use of data altruism consent when electronic health data is processed. The EHDS proposal identifies the general role of health data access bodies in supporting the data altruism mechanism[240]. Moreover, the EHDS specifies that when electronic health data are collected, a secure processing environment shall be ensured by data altruism organisation[241].

The first doubt raised in literature concerns the actual impact and benefits of the data altruism consent. There is discussion whether this new consent is just another layer added to the already existing requirements for processing health data for secondary uses. Otherwise, if it could create an actual harmonisation in the modalities to collect consent from data subjects for secondary uses of data for general interest.[111] Moreover, the data altruism consent has to deal with a difficult interplay with the rules in GDPR about consent and with the GDPR notion of public interest and scientific research. In that sense, as said above, if the legal basis supporting the collection of data from individuals is consent, even in case of data altruism, it shall comply with GDPR's rules on consent. Therefore, the risk is to just replicate the already existing problems of data protection law about consent in the context of DGA and EHDS. [122] Indeed, using consent for secondary uses of personal data for scientific research purposes could be problematic in terms of defining the exact research purposes, in light of the restrictive interpretation of the EDPB of the notion of "broad consent". [127] The role, covered by data altruism consent, of bringing additional safeguard for enhancing transparency and trust among data subjects could be even jeopardized in this context. In case the legal basis used for processing data collected in the data altruism context is not the consent, there would still be problems related to the exercise of data subjects' rights. For example, it is not clear what are the legal consequences if a data subjects decides to withdraw the data altruism consent when the legal basis used for processing data is a different one under article 6(1) GDPR. [114]

The DGA promote the re-use of data in an altruistic way for purposes of general interest and common good. However, these concepts are not properly defined neither in the DGA nor in the EHDS nor in the GDPR. The EDPB and EDPS have strongly suggested to clarify

[238] Recital (46), article 21 DGA.
[239] See article 25(3) DGA.
[240] Articles 37 and 39 EHDS proposal
[241] Article 40 EHDS proposal.

these notions. [16] In particular, the EDPB and the EDPS are concerned by the broad legal uncertainty that could stem from the re-use of personal data for such a broad purpose. Indeed, personal data re-use shall, irrespective of the legal basis, be in compliance with the purpose limitation principle under article 5(b) GDPR. [16] Recital (159) GDPR sheds some lights on the concept of scientific research, as it has been said above in this chapter, however it is still critical to understand when a scientific research project can be considered of general interest under the DGA. Namely there are doubts on how the type of funding of the research (profit or non-profit), rather than the impact of the outcome of the research on the public, can affect such evaluation. [111, 114, 115] In conclusion, the critique raised against the new data altruism consent is that it introduces no novelty respect to the GDPR rules on consent. On the other hand, the risk is to create an overlap with the mechanisms of the GDPR, without though coordinating perfectly the two legislative measures.

2.3.1 – CoC for Consent and Data Altruism

The envisaged goal of data altruism seems to be the enhancement of trust, transparency, legal certainty and individual control over data. However, the exact role and function played by data altruism organisations is not well-defined in the DGA, neither are the functions of the data altruism consent. [122] Therefore, given the lack of definition of the legal framework surrounding data altruism in DGA, an interpretative effort in light of data protection law is necessary. [116] The alignment of the data altruism consent with the GDPR consent shall be better understood and defined. In this case, a CoC might help data controllers and processors reach compliance, for example guiding them in understanding whether a research project could be considered of general interest under the DGA and EHDS.

Another critical point is the withdrawal of data altruism consent from data subjects, especially when consent under GDPR is not the legal basis for processing personal data. In this case, a CoC could commit data controllers in deleting data when such consent has been withdrawn, even though the legal basis is not consent under GDPR. Article 21 DGA establishes that data altruism organisations shall provide tool for obtaining consent or data permit in an altruistic optic, as well as for withdrawing consent or permission. A CoC could provide guidelines for developing such tools in the most appropriate way, especially for data subjects' rights. Finally, the EHDS provides that for processing personal data, data altruism organisations shall comply with requirements in article 50 EHDS, which aim at creating a secure environment for data processing. The definition of the specific measures in order to build a secure environment could be defined with the help of a CoC.

2.4 – Data Controllership

The DGA has introduced a set of new actors that play different roles in light of the new data governance framework. The identification of the responsibilities that such actors have, under data protection law, is not always straightforward. Such problems are also generated by the lack of perfect alignment and compatibility with GDPR's definitions and roles. The DGA has introduced a new definition of "data holder", which is a

"a legal person, including public sector bodies and international organisations, or a natural person who is not a data subject with respect to the specific data in question, which, in accordance with

applicable Union or national law, has the right to grant access to or to share certain personal data or non-personal data"[242].

This definition of data holder could for example create confusion. Namely, where it states that a data holder is a legal person has the right to grant access or to share personal data. As noted by the EDPB and EDPS, the GDPR follows a different paradigm. According to such a paradigm, data subjects have a right of protection upon their data. Namely, data protection rights are constituted by the set of requirements that data controllers have to respect when processing personal data. [16] On the other hand, no explicit right to grant access to data is assigned to data controllers in the GDPR. Scholarship has discussed indeed what kind of right a data holder can actually exercise on personal data. It shall be noted that personal data is not usually considered a property; property, on the other hand, is an essential part of access/sharing rights. It has been noted that the DGA follows, from a literal point of view, a right-based approach. This approach seems to be adopted where the DGA states that data holders have a right of granting access and sharing data. These rights do not seem to be ultimately created by the DGA, though. [122] Anyway, such confusion in the terminology can hamper the readability of the piece of legislation, as well as its actual implementation. [16]

Another new role is the "data user", which is defined under the DGA as

"a natural or legal person who has lawful access to certain personal or non-personal data and has the right, including under Regulation (EU) 2016/679 in the case of personal data, to use that data for commercial or non- commercial purposes;".

The interplay between this definition and the definition of data recipient under GDPR is again not straightforward[243]. Moreover, an interpretative effort could also be necessary for identifying whether a data user would act as data controller, joint controller or processor. [16] In the same vein, the role played by data altruism organisations is not clear in terms of their actual responsibility in light of the GDPR, as already said above in this work. [122] Consequently, it could be difficult to foresee which would be the role of data altruism organisations under data protection law. The same issue applies to data intermediation services under article 10 DGA.

For example, as noted by the EDPB and EDPS, looking at the EHDS, it is not always clear who is in charge of some of the tasks assigned to data holders and data users. Especially for the exercise of natural persons' rights. This is due to a difficult interpretation between the definition of data holder in the DGA and its counterpart in the EHDS. [15] That situation could make difficult understand which are the organisations that shall make data listed in article 33(1) available for re-use.

Moreover, some tasks of data holders could be just difficult to implement, especially in light of similar rights under GDPR. For example, article 3(9) EHDS provides that "natural persons shall have the right to restrict access of health professionals to all or part of their electronic health data.". On the other hand, the GDPR already provides a right to restrict data processing to data subjects, namely under article 18 GDPR. Under article 18 GDPR, the right to ask for a restriction of personal data processing can be exercised by data subjects

[242] Article 2(8) GDPR.

[243] Article 4(9) GDPR reads: "'recipient' means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not.".

only in case some specific conditions apply[244]. Therefore, since article 3(9) EHDS does not make any specific reference to article 18 GDPR, it is not clear how the two provisions could be applied on practical terms.

2.4.1 – CoC for Data Controllership

The definitions of the roles introduced by the DGA and to some extent in the EHDS seem not to be completely compatible and aligned with the roles and the definitions under the GDPR. Moreover, even between the DGA and the EHDS there are some inconsistent definitions. This lack of coordination could hamper the smoot interpretation and implementation of the legal framework. A CoC could guide data controllers and processors in reducing the legal uncertainty providing common interpretation of complex interplay between DGA, EHDS and GDPR. For example, in the detection of the organisations falling into the definition of data holder under the EHDS. Otherwise, even providing some guidelines about how to implements rights of data subjects, such as the right to obtain a restriction of data processing by health professionals in light of the GDPR. Finally, the identification of the roles under data protection law perspective, i.e., data controller, processor, or joint controller, is essentially a case-by-case analysis. The implementation of such analysis, and especially its demonstration of appropriateness under the principle of accountability is not an easy task. A CoC could provide a "rulebook" with indications about how to carry out the analysis.

2.5 – Legal Basis for Primary and Secondary Use of Data in Health

In the health context, the choice of the proper legal basis for processing personal data for primary care, as well as for secondary uses (e.g., research), is not a trivial task. In general, as "primary use" of data is meant data processing for the provision of care and services to the patient. As "secondary use" of data, on the other hand, is meant the use of data for research and innovation or for enhancing the functioning of the health care system, or policy making[245]. The EHDS itself distinguish between primary and secondary uses of data[246].

The GDPR has left room to national legislators for the introduction of specific requirements for data processing in health domain. This space left to national legislators include also the appropriate legal basis for such data processing. The choice is motivated by the necessity to create different regimes in light of the differences that historically characterise national health care systems. And in light of the lack of full competence of the EU legislator as regards the organisation of the national healthcare systems according to article 168 TFEU. [219] All these differences have created a fragmented legal landscape among EU MSs, especially as regards the legal basis needed for personal data processing. A study conducted by the EC has highlighted how at national level there is fragmentation about the

[244] Namely, if: "(a) the accuracy of the personal data is contested by the data subject, for a period enabling the controller to verify the accuracy of the personal data; (b) the processing is unlawful and the data subject opposes the erasure of the personal data and requests the restriction of their use instead; (c) the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defence of legal claims; (d) the data subject has objected to processing pursuant to Article 21(1) pending the verification whether the legitimate grounds of the controller override those of the data subject.".
[245] See Chapter 1, Section 1.
[246] Article 2(2) (d) and (e) EHDS proposal.

legal basis for processing personal data in health care both for primary and secondary uses. [80] Especially in transborder sharing of data the fragmentation of legal basis constitutes a big obstacle for boosting big data uses. Nevertheless, the new EHDS will constitute an EU-level legal basis for personal data processing.

2.5.1 – CoC on Legal Basis for Primary and Secondary Use

A CoC, especially if coordinated to other CoCs in the creation of a European framework of GDPR CoCs, could introduce more legal certainty for the data controller. However, it shall be clarified that legal fragmentation in data protection law cannot be solved just by CoC. A CoC indeed cannot amend or modify national law, but only create a harmonisation of the interpretation of hard law requirements. Nevertheless, the EC has conduct studies that shows how stakeholders support a wider use of CoCs in order to contrast the negative effects of legal fragmentation in health care domain. [80] A CoC could help data controllers in choosing the proper legal basis for processing personal data in health domain, both for primary and secondary uses.

2.6 – Data Minimisation, Purpose and Storage Limitation in Big Data Context

Technologies that work on the elaboration of big amount of data, often generally known as AI, are frequently at odds with some principles of the GDPR. [12] Namely, the principles of data minimization, purpose limitation and storage limitation are difficult to apply in big data context. Indeed, the added value of such systems is the possibility to analyse big volume of data without having in mind a clear vision of which specific function of the decision-making system will be supported by the analysis. In other words, it is hard, if not impossible, to define in advance the exact final purposes of the data processing. On the contrary, the purpose limitation and storage limitation principles requires the data controller to specify the purposes before starting the processing and to delete data, or at least anonymise it, once the objective is reached[247].

As regards the data minimisation principle, it entails that only the minimum amount of data is collected and processed in light of the declared purposes. From this point of view, it is possible to note that in big data contexts it is necessary to collect and process as much data as possible in order to increase the decisional and predictive capacity of the system. However, it shall also be borne in mind that the amount of data to be collected shall always be considered in light of the quality of the data. Just collecting and processing huge amounts of data is not efficient. [139]

It shall also be noted that data minimisation could also negatively affect AI performance in terms of bias reduction. Indeed, it is possible to note a trade-off between trustworthy and equity in AI systems and data minimisation principle. Moreover, using data anonymisation techniques as data minimisation measures could lead to an excessive reduction of the data value. If the minimisation technique impacts too heavy the quality and the value of the data, then the AI system performance will not be efficient. [72, 220] In this perspective, the coordination with the new AI act proposal is crucial. The AI act provides several measures meant to ensure the quality of the data and the absence of bias, respecting at the same time the GDPR's provisions[248].

[247] See article 5 GDPR.
[248] See Title III, Chapter 2 AIA proposal.

2.6.1 – CoC on Data Minimisation, Purpose and Storage Limitation

A code of conduct could define standards and guidelines in order to better perform the trade-off between such data protection principles and the efficiency and the trustworthiness of AI systems. Moreover, the several links between the AI act and the data quality principles of the GDPR will create an overlap between the two instruments, a CoC could help stakeholders give interpretation of the AI act in compliance with the GDPR.

3 – Preliminary Conclusion

The first three Chapters of the thesis have carried out a theoretical analysis on opportunities as well as shortcomings related to CoCs and CMs in data protection law. It has been observed that a shift in the regulatory approach by the EU legislator has characterised the come into force of the GDPR. Accountability activities have essentially changed, moving from formalistic procedures to more heavy risk-based assessments. Moreover, some principles of the GDPR are still at the odds with big data and AI functioning. Being in compliance with data protection law provisions is a struggle for many organisations of the health domain. This assumption seems to be true also in light of the new layer of regulation introduced by the DGA and the forthcoming EHDS.

Using middle-layer co-regulatory solutions seems to be more than desirable for easing the interplay of the GDPR with the new data governance, as well as for enabling a greater use of big data and AI solutions in health-related domains. Some specific potential contents of a CoC for the use and re-use of personal data in health has been highlighted in this Chapter. Many concepts could be clarified, such as those of anonymisation in GDPR, or scientific research for general interest under the DGA. However, also more operational procedures could be specified into a CoC. For example, how to perform a proper pseudonymisation procedure or how to choose the proper legal basis according to national legislations.

The obstacles to the development and to the spread of these instruments have been identified as well. In general, the GDPR seems to lack a strong enough approach in relying on co-regulation. The GDPR indeed adopt these instruments but without assigning to them a real central role in data protection law. The comparison with harmonised standards is explicative in this sense. GDPR's co-regulatory solutions seems to lack of proper legal effects in terms of compliance demonstration before DPAs. This aspect essentially makes difficult to convince stakeholders investing time and resources in developing and adopting such instruments. Which are therefore seen as costly facultative instruments by stakeholders without concrete effect on the legal perspective. Moreover, some obstacles are posed by the conditions that characterise the health domain itself. A complex legal fragmentation across EU health systems and the fact that most of the actors of the system are public sector bodies or public undertaking make even more difficult the spread of EU-wide co-regulatory solutions.

The conclusion of this first theoretical part of the thesis is though that co-regulatory solutions are potentially useful instruments that can produce a number of benefits for the whole ecosystem. This instruments indeed have a role of trust enablers, as well as compliance facilitators, having positive consequences for both individuals and organisation and even for DPAs. Therefore potentially, as discussed in this Chapter these instruments are carrier of a shift in stakeholders' behaviour towards privacy-wise organisational decisions.

Unfortunately, the only real force pushing for the use of these instruments is a market related one: data controllers and processors shall decide to invest in these compliance instruments in light of an economic return. Since the legal effects in terms of burden of proof are blurry, the only reason that could really motivate an investment in CoCs or CMs by stakeholders is a competitive advantage on the market. The competitive advantage on the market could be generated in different ways. In the first place, through CMs, seals and marks organisations can show their commitment in privacy aspects and gain shares of the market. Moreover, trough CoCs organisations can better perform compliance activities, avoiding cost related to fines and litigations related to data protection. In conclusion, the whole ecosystem and society can benefit from a smoother and more responsible application of the law in this domain. Indeed, the sharing of health data for primary and secondary uses is carrier of benefits for everyone. The next Chapter will deal with the specific steps for developing a hypothetical CoC for the use and re-use of personal data in health domain. Both procedural aspects and a potential structure of the content of the CoC will be analysed.

Chapter 4 – Proposal for a Code of Conduct for Health Data Processing

1 – Road Map of a Code of Conduct for Health Data Processing

The last chapter of the thesis will delve into the practical procedures that shall be pursued in order to develop a CoC for the processing of personal health data, under article 40 and 41 GDPR. The steps detected by this chapter are either grasped from other projects of CoCs for data processing in EU, or explicitly required by the GDPR requirements on CoCs and the EDPB Guidelines.[145]

The GDPR and the EDPB guidelines establish, in the first place, some formal parts that shall be included in the content of the CoC. If the body of the CoC misses to contain such parts, then the CoC cannot even be submitted to the DPA for the assessment. These requirements represent therefore the essential part a CoC must contain. However, these parts for a CoC are a necessary condition, although not sufficient, for the CoC to be approved. Besides the mandatory and formal parts, the CoC shall then satisfy other requirements in terms of objectives that the content of the instrument shall reach. If such objectives are not fulfilled, the DPA, after the forma assessment, shall not approve the CoC. These objectives are indicated in the EDPB guidelines and constitute the criteria against which the DPA carries out its assessment of the CoC. Essentially, the DPA adopts its decision of approving or rejecting the CoC upon the evaluation it.

The rest of the Section 1 will detect the preliminary steps for the drafting of a CoC. These steps are the writing of the scientific basis for the CoC, namely in the form of a White paper (Section 1.1), and the definition of the expected results in terms of improvement of GDPR implementation (Section 1.2). Namely, the results expected are: 1) meeting a particular need (Section 1.2.1); 2) facilitating the effective application of the GDPR (Section 1.2.2); 3) specifying GDPR's requirements (Section 1.2.3); providing sufficient safeguards (Section 1.2.4); having an effective monitoring system (Section 1.2.5), To prepare the submission to the DPA, then the phase of consultation with stakeholders is presented (Section 1.3). Finally, the procedure for the submission of the draft CoC to the DPA (Section 1.4) is discussed as a whole, although then distinguishing between national (Section 1.4.1) and EU-wide CoC (Section 1.4.2). Section 2 will delve into the analysis of the structure that a CoC for health data processing might adopt. In the first place, the code owner(s) shall include the mandatory and formal parts of a CoC (Section 2.1), which have been mentioned above, consisting in the explanatory statement (Section 2.1.1), the representativeness of the CoC (Section 2.1.2), the material scope (Section 2.1.3), the submission to the DPA (Section 2.1.5), the oversight mechanism (Section 2.1.6), consultation activities (Section 2.1.7), compliance with national legislation (Section 2.1.8). Furthermore, a possible structure of the rest of the CoC is proposed taking inspiration from other projects of CoCs (Sections 2.2 and 2.3). The structure of the CoC proposed is divided between two main parts: 1) guidelines (Section 2.2.1); and checklist (Section 2.2.2).

1.1 – White Paper

The first step for developing a CoC is to identify a real need of clarification in terms of data protection issues. Therefore, the first passage could be to develop a white paper where

the data protection issues of the specific sector are analysed[249]. For the health domain, it shall be demonstrated how the application of data protection law generate doubts and difficulties in terms of GDPR implementation and compliance with legal requirements. The studies conducted in the present work, and the literature cited throughout the thesis, could be used as starting point for demonstrating that the processing of health data is carrier of data protection complexities. Detecting a real need for developing a CoC is a requirement explicitly indicated in the EDPB guidelines in order to pass the assessment of the DPA. [145]

Health data processing represent a domain where, as it has been argued in the present work, compliance issues often arise. The need to clarify data protection issues is even more remarkable when some technologies are implemented in the health domain. The implementation of AI technologies, which revolve around the use of big data analytics, has exacerbated the friction between the voracious need of personal and the constraints posed by data protection law to personal data elaborations. [18, 26, 125, 142]

Moreover, the EC is launching an ambitious programme for incentivising the sharing and the re-use of data across EU in order to boost the deployment of AI and big data analytics technologies. [14] The new legal provisions, part of which have already been adopted (e.g., the DGA[250]), introduce a new layer of rules that create a new data governance framework. This new set of provisions will naturally have a strong interplay with data protection law. [122] The way data protection law is implemented is about to change, especially in the health domain. The use of electronic health data has been identified as the first strategic sector where the new data governance framework, introduced at general and horizontal level by the DGA, will be further specified by the proposed EHDS. [110, 209]

The new data governance framework provided by the DGA and the new sectorial rules that are to be introduced by the EHDS, will strongly impact the way data protection law is applied in health domain, both for primary and secondary uses of data. In Chapter 1, Section 1 the difficulties in terms of data protection law compliance brought by the use of AI and big data in the health domain have been highlighted. On the other hand, the complexities about the difficult interplay between data protection law and the new DGA and EHDS has been discussed in Chapter 1, Section 3.2. Moreover, some more specific data protection issues have been further investigated in Chapter 3, Section 2. The discussion about the problems indicated in Chapter 3 has been done on a general level because the specific aspects of data protection law that can be included into a CoC would change according to the composition of the stakeholders who participate in the drafting of the CoC. On the contrary, it is not possible to define ex-ante in a detailed way the content of a CoC. This decision heavily relies on the specific objectives decided among the stakeholders and the interests carried out by the same. However, some general critical points have been indicated as main issues of data protection law in the domain at stake. This existence of these problems should however be enough to argue that a CoC in this domain is needed in light of some objectively difficult implementation of data protection law.

[249] Many of the projects on CoC and certifications even when carried out by industry actors rely on academic research investigations, see for example [167].
[250] Regulation (Eu) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

1.2 – Definition of the Content

The definition of the specific data processing issues addressed into a CoC shall be decided on a case-by-case analysis when the stakeholders participating to the project gather together and start discussing. However, as said before, some main critical issues of the domain can be highlighted and sketched in Chapter 3, Section 2. Moreover, according to the EDPB guidelines, [145] a CoC shall include some essential formal parts, the lack of which make the CoC not admissible for the DPA's evaluation. This mandatory elements of the CoC will be further analysed later in this chapter in Section 2.1. The CoC's content, anyway, shall pursue some specific goal and produce some essential effects directly indicated in the EDPB guidelines. These objectives to reach out are the already mentioned criteria that will be evaluated by the DPA in its substantial evaluation on the CoC content. The rest of this section will briefly deal with such of these objectives.

1.2.1 – Meet a Particular Need

The first criterion to be satisfied has been already mentioned[251]. Namely, the CoC shall meet a particular need in terms of data protection issues into a specific sector. In this sense, it has been shown how the use of a white paper that refers to literature can be useful from this perspective. It could be noted that the EDPB and the EDPS suggest scientific health research as a possible sector where a CoC could find room to be applied.[121, 145] Therefore, it should not too difficult to argue that a CoC could be used to meet a particular of this particular domain. Nevertheless, the use of a scientific preliminary research could be an added value to bring before the DPA.

1.2.2 – Facilitate the Effective Application of the GDPR

The second criterion that will be evaluated by the DPA in assessing a draft CoC is whether the CoC facilitate the effective application of GDPR requirements[252]. This requirement essentially is meant to ensure that the content of the CoC is sufficiently oriented towards a specific sector. The specific needs of the sector at stake shall be taken into account and the application of the GDPR in the sector shall be improved in practical terms. The specific terminology of the sector shall be included in the CoC. The CoC should be readable by the practitioners of the sector. Moreover, as suggested by the EDPB, the general requirements could even be further specified in light of the sectorial features. On the other hand, also the specific risk of the sector must be addressed by the CoC. As a result, the rationale of the requirements should be better implemented than in case of the simple GDPR implementation. [145]

1.2.3 – Specify the Application of the GDPR

This criterion is meant to evaluate whether the CoC can actually improve the application of GDPR in the specific sector at issue. In order to improve the application of GDPR, the CoC shall clearly indicate which are the aspects of data protection law it aims to focus on. Moreover, the CoC shall also indicate clear and realistic solutions for solving the problems detected. In this regard, the CoC shall be enough specific and precise. On the other hand, the CoC shall not just re-state GDPR provisions. [145]

[251] See above Section 1.1.
[252] See recital (98) GDPR.

The CoC must therefore set up specific standards for the sector, through which the implementation of GDPR shall be implemented in a practical manner. The content of the CoC shall be concrete and enforceable. In this regard, it is important that the CoC deals with concrete scenarios of the sector at stake, even providing examples and best practices. The EDPB guidelines, in this sense, stress again that the content of the CoC shall try to include terminology from the sector and avoiding, on the other hand, legalistic wordings. [145]

The EDPB guidelines consider crucial the ability of the CoC to create operational rules for implementing the data protection principles of article 5 GDPR. The CoC, in doing so, shall take into account all the sectorial guidelines and opinions adopted either at national or EU-level on the implementation of data protection law applicable. Also, the jurisprudential orientation shall be taken into account when developing a CoC. [145]

1.2.4 – Sufficient Safeguards

Article 40(5) and the EDPB guidelines require that the submitted CoC provides enough "appropriate safeguards"[253]. The CoC shall demonstrate that thanks to its adoption, rights and freedoms, especially in light of the risk of the sector, are preserved. The code owner(s) is required by the EDPB guidelines to provide enough evidence on this point. Therefore, it would be desirable that the CoC provides even higher safeguards standards than just the GDPR requirements. In this regard, as suggested by recital (99) GDPR, it would be useful to include also consultation with data subjects' representatives, in order to understand the proper way to mitigate high-risk situations. [145]

1.2.5 – Mechanisms for an Effective Oversight of the CoC

Another essential element is that the CoC contains mechanisms for a proper monitoring activity over its requirements. This aspect is a mandatory requirement for the submission of the CoC to the DPA. A CoC shall therefore include mechanism for the monitoring and the enforcement of its own rules. In order to ensure a monitoring activity over the CoC's content, the CoC itself shall identify a monitoring body. Such monitoring body shall therefore be responsible of ensuring compliance with the CoC and adopting measures in case of infringement of the CoC's rules. [145]

It shall be noted that, according to article 41(6) GDPR, the monitoring activities of a CoC shall not be pursued by a monitoring body when the data processing are performed by a public body. However, the EDPB has specified that this exemption does not mean the CoC should not include mechanisms for effective oversight. Therefore, it only means that the monitoring shall not be performed by a monitoring body, but still some mechanisms for monitoring the compliance shall be present in the body of the CoC. [177] At this point, it could be natural posing the following question: what is the purpose of having a mechanism for the effective oversight compliance if there is nobody performing the monitoring activity? The monitoring systems, besides detecting the monitoring body, might identify auditing activities and a system for handling complaints and even a system of sanctions and remedies, as explicitly suggested by the EDPB. The sanctions can even include monetary fines, if the code owners can demonstrate the rationale of such sanctions and that they are operationally feasible. [145]

[253] See also recital (98) GDPR.

1.3 – Involvement of the Stakeholders

The involvement of and the interaction with stakeholders of the sector is another key step of the CoC's development. The Code owner(s) shall demonstrate to have a certain grade of representativeness in the sector at stake. However, neither in the GDPR nor in the EDPB guidelines there is a clear guidance about how to evaluate such representativeness. The EDPB guidelines provide that the CoC must be submitted by an association or a consortium of associations or by other bodies representing the categories of controllers or processors. This provision is essentially what is already stated in article 40(2) GDPR. However, the EDPB guidelines at least provide a non-exhaustive list of possible code owners, such as "trade and representative associations, sectoral organisations, academic organisations and interest groups." There is a rather big room for the discretionary decisions on this point for the DPAs. The only two concrete criteria explicitly indicated are: 1) "the number or percentage of potential code members from the relevant controllers or processors in that sector; 2) experience of the representative body with regard to the sector and processing activities concerning the code". [145] This lack of specification has been observed also in literature, raising critiques because it leaves too big room for discretional decisions to the DPA. [153]

It shall be noted that the EDPB guidelines ask to the code owner to demonstrate such a representativeness. Namely, it shall be demonstrated that the code owners are capable of understanding the needs of the stakeholders that operate in the sector where the CoC is to be applied. Demonstrating wide participation of stakeholders already during the preparatory work of the CoC can be used as evidence of representativeness. Therefore, just few criteria are suggested by the EDPB for carrying out the evaluation of the representativeness feature. While, on the other hand, the code owners seem to have a rather big task of demonstrating that they are enough representative in the sector.

From a different perspective, the EDPB guidelines explicitly ask to demonstrate that the code owner had had enough consultations with stakeholders, also including data subjects' representatives. As it is also stated by recital (99) GDPR, the code owner shall carry out consultation activities at an "appropriate" level and, if it feasible, such consultations shall also include discussions with data subjects. Again, it is not completely clear how such consultations should be carried out and to what extent the content of the CoC should reflect the opinions and the positions of data subjects. In this regard, the EDPB guidelines state that the code owner shall provide demonstration of having carried out the consultations at stake, specifying at what level the consultations have been carried out, as well as the nature of the consultation. The EDPB further specifies that the consultation activity is recommended between the parties that act as code owner. However, also clients and commercial partners of the code owner(s) should be heard. Therefore, it seems that all the stakeholders who could be impacted by data processing regulated by the CoC should be consulted. If these consultations do not take place, because it is not feasible doing it, then the code owner shall provide explanation about such lack of feasibility. [145]

1.4 – Submission to the DPA

The procedure for submitting a CoC changes according to the geographical scope of the CoC. Therefore, whether the CoC is meant to have national of transnational scope. Either way, the CoC shall be submitted to a the national DPA which will be qualified as

Competent Supervisory Authority (CompSA). The CompSA will carry out a revision of the mandatory conditions for the admissibility of the CoC. After this point the procedure changes according to the territorial scope of the CoC.

1.4.1 – National Codes of Conduct

In case of national CoCs, the CompSA is the DPA to which the code owner shall submit the CoC, it is also the competent DPA for the controllers that adhere to the CoC. In this sense, the jurisdictional scope of the CoC must be clearly indicated by the code owner. As already said, at this point, the CompSA (if the CoC is deemed admissible[254]) shall start the evaluation of the CoC's content, in accordance with the procedures provided by national law. The CompSA shall draft an Opinion concerning the draft CoC in consideration of the criteria indicated in the EDPB guidelines. [145] In doing so, the CompSA can even provide feedback to the code owner that can be used for an eventual second submission. This step is important especially if the decision of the CompSA is to refuse the approval of the CoC. In case of refusal, the process is over, and the code owner can in case choose to submit again the CoC but integrating the feedback from the CompSA decision. On the other hand, if the CompSA decides to approve the CoC, the CoC shall be registered and published. The CompSA is responsible for such publication and registration tasks, for example through its website. Moreover, according to article 40(11) the EDPB is asked to make all the approved CoCs publicly available through a register[255].

1.4.2 – Transnational Codes of Conduct

In case of transnational CoCs, the code owner(s) shall submit the draft CoC to a DPA that will act as CompSA, and this will be the main authority for the approval of the CoC. The task of this CompSA is (as for national CoCs) to evaluate the admissibility of the CoC. In case the CoC is evaluated as admissible, before proceeding with the assessment, the CompSA shall notify all the other DPAs about the submission of the transnational CoC. The other DPAs shall at this point decide if they consider themselves as "concerned DPAs" in the context of the CoC according to article 4(22)(a) and (b) GDPR[256].

Once the admissibility of the CoC is ascertained by the CompSA, and once the CompSA has notified about its decision all the other concerned authorities, an informal cooperation procedure for the assessment of the CoC starts. The cooperation procedure begins with the CompSA notifying the other SAs of its position on the admissibility of the CoC. Following the notification, the other SAs shall appoint a maximum (depending on the complexity and the scope of the CoC) of two co-reviewers. The co-reviewers will therefore assist the CompSA in the assessment of the CoC's content. Namely, they are supposed to provide comments on the content of the CoC (within 30 days from their confirmation as co-

[254] Therefore, if the draft CoC contains all the mandatory elements indicated above in Section 2.1 of this chapter.

[255] See https://edpb.europa.eu/our-work-tools/accountability-tools/register-codes-conduct-amendments-and-extensions-art-4011_en and https://edpb.europa.eu/our-work-tools/accountability-tools/certification-mechanisms-seals-and-marks_en.

[256] According to article 4(22) GDPR a "concerned supervisory authority" is a DPA that is concerned by a data processing because either "a) the controller or processor is established on the territory of the Member State of that supervisory authority" or "b) data subjects residing in the Member State of that supervisory authority are substantially affected or likely to be substantially affected by the processing".

reviewers). These comments shall be taken into account by the CompSA in carrying out its assessment. Nevertheless, the EDPB guidelines specifies that the final decision, according to article 40(7) GDPR, whether the draft decision can be submitted to the EDPB, is a task solely of the CompSA. In case the CompSA decides that the CoC shall be approved and submitted to the EDPB, the procedures shall follow the mechanisms under articles 63 and 64 GDPR[257].

The decision as to whether approve or refuse the CoC shall be adopted by the CompSA within a reasonable time, keeping the code owner regularly updated. Moreover, the basis upon which the decision is adopted shall be clearly indicated. If the CompSA decides to refuse the CoC, as in the case of national CoC, the procedure will be over, and it will be up to the code owner whether to re-submit another draft CoC later. The only additional stage, in case of transnational CoC, will be the duty upon the CompSA to communicate to the others SAs about its decision and the reasons motivating it. On the other hand, if the decision is to submit the draft decision to the EDPB, the CompSA shall, in the first place, circulate its draft approval decision among all the concerned SAs. At this point, the SAs will have 30 days to respond and bring concerns and issues to be discussed before the EDPB, in line with the consistency mechanism pursuant to article 64 GDPR.

According to article 64(4) GDPR, the view of the other SAs concerned shall be included in the draft decision submitted to the EDPB in the framework of the consistency mechanism. After this step, the EDPB is required to issue an Opinion concerning the CoC submitted, as provided by article 70(1)(x) GDPR. The Opinion of the EDPB shall be issued following the procedural rules of the Board and the consistency mechanism under article 64 GDPR, as it has been said already. According to article 64(5) GDPR, the EDPB shall communicate its Opinion to the CompSA and to the other SAs concerned. The CompSA, at this point, shall, according to article 40(5), decide whether to maintain or amend its own decision on the CoC. In case the CompSA does not agree with the Opinion of the EDPB, the procedure in article 65(1) applies[258]. If the Opinion of the EDPB is positive, the Board shall submit its Opinion to the EC according to article 40(8) GDPR. The EC shall decide, through implementing acts, whether the CoC have general validity within the EU. In conclusion, in the same vein of national CoC, the EDPB is supposed to collate all the CoC in a registry in order to make all the CoCs publicly available under article 40(11) GDPR.

2 – Structure of the Code of Conduct

Looking at the structure of the CoC, i.e., how the content of the CoC is organised among its parts, it is possible to divide between the mandatory parts of the CoC and the rest of it. The mandatory parts of the CoC, as it has been said already, determine the admissibility of the CoC to the substantial evaluation from the CompSA. The essential parts of a CoC are enlisted in the EDPB guidelines 1/2019 and consist of ten requirements which will be analysed hereafter in the next Section 2.1. Besides these formal parts, the CoC shall then

[257] Article 63 GDPR defines the "consistency mechanism", according to which "In order to contribute to the consistent application of this Regulation throughout the Union, the supervisory authorities shall cooperate with each other and, where relevant, with the Commission, through the consistency mechanism as set out in this Section". Article 64(1)(b) and (c) GDPR, on the other hand, define the role of the EDPB in issuing Opinions concerning draft CoCs and CMs.
[258] See article 65 GDPR "dispute resolution by the Board".

regulate the subject-matter or material scope. In other words, the subject matter of the CoC is the part of the CoC that directly deals with the data protection issues in a substantive way. The subject-matter of the CoC is the part of the CoC that shall satisfy the requirements provided by the EDPB in terms of objectives, as discussed above. Despite the EDPB provides some objectives to reach, no indication about how to organise in practical terms this part of the CoC is provided. Therefore, in light of the room that the code owner(s) has in managing this part of the CoC, a solution is proposed in Section 2.2 of this chapter.

2.1 – Essential Parts

The first elements that a CoC shall indicate are, as said above, the mandatory elements of the CoC, which are listed by the EDPB guidelines and that are hereafter discussed.

2.1.1 – Explanatory Statement and Supporting Documentation

The CoC's content shall be introduced by an explanatory paragraph where some information is summed up. Namely, the purpose of the CoC and the scope of the CoC shall be specified, as well as the expected results in terms of facilitation of GDPR application. [145] Essentially, this part of the CoC is meant to clarify the rationale of the CoC, as well as its basis, the benefits it can bring and to whom the beneficial effects will apply. Moreover, this part of the CoC could include a summary of the research at the basis of the CoC, plus further information on the actor(s) participating to the drafting of the CoC. [145]

2.1.2 – Representativeness

As said above, the CoC, in order to be deemed admissible to the DPA's assessment of the content, shall indicate its representativeness[259]. Therefore, the content of the CoC shall provide enough information to demonstrate that the code owner(s) is able to represent the sector at stake and understand its actual needs in terms of data protection law compliance. The means through which such demonstration of representativeness shall take place however are not extremely clear, as it has been mentioned already.

From the analysis carried out here, it could be argued that this is not an extremely strict requirement. [145] Situations very different among each other have been considered as satisfying in terms of representativeness. For example, an Italian CoC concerning the reuse of health data[179] has been deemed admissible in light of the role played by its code owner in the Italian health system. Such CoC has been elaborated by the "Regione Veneto"[260], which is the regional government of Veneto. In Italy, indeed, the organisation of the health systems is largely delegated to regional authorities[261]. The regional governments are therefore asked to organise and control the provision of care at regional level. In light of that, it could be argued for sure that a regional government is representative enough within its own Region itself, but not at national level. Nonetheless the Italian DPA has deemed the CoC admissible and approvable. On the other hand, looking at another CoC,

[259] See above, Section 1.3.
[260] In collaboration with a specific public hospital operating in the same region.
[261] The organisation of the National Health Care System is divided between the State, Regions, and other local entities, see Legge 23 dicembre 1978, n. 833 Istituzione del servizio sanitario nazionale. (G.U. Serie Pregressa , n. 360 del 28 dicembre 1978).

i.e., the Farmaindustria CoC approved in Spain[262], the situation is totally different. In this case, the national CoC has been elaborated by Farmaindustria[263], which is an association representing at national level in Spain the pharmaceutical industry. In this case, it seems more straightforward to understand why the Spanish DPA has deemed the code owner as representative in its sector.

2.1.3 – Data Processing and Material Scope

The CoC shall be clear enough concerning its material scope. Therefore, the CoC shall explicitly state which are the data processing and the characteristic of such data processing object of the CoC. Essentially, the CoC shall clearly indicate, already at the beginning of it, for example into a dedicated section, which are the data processing issues addressed into it. The categories of data controllers and processors the CoC is meant to govern shall be clearly indicated as well. [145]

2.1.4 – Territorial Scope

Co-regulation mechanisms under GDPR can either have national or transnational applicability. This is a choice that strongly affect the content of the CoC as well as the effects on organisations that decide to follow the CoC. Therefore, the geographical scope shall be clearly indicated at the beginning of the CoC. Moreover, the procedure for obtaining the approval from the DPA is different according to whether the CoC is national or transnational.

If the CoC is transnational, i.e., it is meant to be applied in two or more MSs, then the MSs where it is supposed to be applied shall be clearly indicated. In this case, also the concerned SAs shall be stated in this part of the CoC. It shall then be noted that, for being qualified as transnational CoC, it is not necessary for the CoC to cover data processing activities that constitute a "cross-border data processing"[264]. In order to be qualified as transnational, a CoC only need to regulate data processing activities carried out by controllers or processors in different MSs. Therefore, as clarified by the EDPB Guidelines, when a CoC is adopted by an organisation at national level, but the CoC is meant to regulate data processing that the organisation carries out in other MSs, then the CoC shall be considered as transnational. On the other hand, the EDPB provides, as an example, the case of a CoC approved only at national level but adopted by an organisation that also carries out cross-border data processing. In this case, the organisation cannot claim to benefit from using a CoC for trans-border data sharing. It can only claim the adherence to the CoC for data processing that fully take place in the MSs where the national CoC has been approved.

In general, communicate transparently the territorial scope of the CoC is crucial. First of all, because, as said above, the procedure for its approval changes according to that.

---

[262] See https://www.aepd.es/es/prensa-y-comunicacion/notas-de-prensa/aepd-aprueba-primer-codigo-conducta-sectorial-desde-entrada-vigor-rgpd (last access: January 2023).

[263] See https://www.farmaindustria.es/web_en/ (last access: January 2023).

[264] According to article 4(23) "cross border processing means either: (a) processing of personal data which takes place in the context of the activities of establishments in more than one Member State of a controller or processor in the Union where the controller or processor is established in more than one Member State; or (b) processing of personal data which takes place in the context of the activities of a single establishment of a controller or processor in the Union but which substantially affects or is likely to substantially affect data subjects in more than one Member State."

Moreover, transparent communication of the territorial scope is also meant to avoid misleading interpretation about the potential impacts of the CoC by data subjects or other stakeholders.

2.1.5 – Submission to the DPA

The code owner(s) shall choose the appropriate DPA to which submitting the CoC, ensuring that that is the competent authority (CompSA) in accordance with article 55 and recital (122) GDPR. The choice of the CompSA is not that problematic for national CoC. Otherwise, in case of transnational CoC, the code owner(s) might have a margin of choice. The articles of the GDPR dealing with CoC do not provide clarification about how the code owner(s) shall identify the CompSA in case of transnational CoC. However, Appendix 2 of the EDPB Guidelines provide some parameters for guiding the choice of code owners in this sense. Despite not being exhaustive, some factors to take into account by code owner(s) for the choice of the CompSA are: a) the MS where take place most of the data processing ruled by the CoC; b) the MS where are located most of the data subjects affected by the CoC; c) the MS where are located the code owner(s)' headquarters; d) the location of the monitoring body's headquarters; e) the MSs where a DPA has launched specific policy and initiatives concerning the sector to be regulated by the CoC at stake. It shall be noted that these are not strict criteria, as stated by the EDPB itself, and that the code owner shall anyway be able to motivate its choice properly. [145] The choice of the CompSA is anyway a really important step to be carefully evaluated. Indeed, the CompSA, beside conducting the main evaluation of the CoC, will also act as single point of contact with the code owner(s). Moreover, the CompSA will provide accreditation to the monitoring body and will monitor its activities as lead DPA. [145]

2.1.6 – Oversight Mechanisms and Monitoring Body

A CoC, in order to be approved, shall contain a mechanism for monitoring the compliance of organisations that undertake the CoC. Moreover, the CoC, if aiming at regulating data processing carried out by private organisations, shall also identify a monitoring body. On the other hand, a monitoring body is not required when a CoC content concerns data processing carried out by public authorities or public bodies[265]. In case a monitoring body is required, the CoC shall indicate mechanisms that enable the monitoring body to conduct its function of monitoring of compliance, as indicated in article 41 GDPR. However, it shall be noted that even though CoCs involving public sector bodies does not need a monitoring body, it still needs to identify a system for monitoring the compliance with the CoC. [145] The monitoring body (or bodies, in case they are more than one[266]) shall receive the accreditation from the CompSA. The accreditation is the certification from the DPA that the monitoring body satisfies the requirements for exercising its tasks[267].

2.1.7 – Consultation

Consultations with relevant stakeholders shall be carried out during the drafting of the CoC. Such consultations shall be documented and included in the CoC itself. Providing information about the consultation that took place during the drafting of the CoC shall be

[265] See article 41(6) GDPR.
[266] As this possibility is indicated in EDPB Guidelines 1/2019.
[267] See Chapter 2, Section 2.2.2.

included among the mandatory elements of the CoC. In this part of the CoC, the code owner(s) should indicate how discussion with relevant stakeholders of the sector have been conducted. Also, if possible, data subject's representatives should be included in these discussions. For example, the code owner(s) should be able to indicate if and how they have included in the CoC's content the opinion and the position emerged from other stakeholders during the consultation activities. In this regard, the EDPB guidelines recommends to code owner(s) to carry out both internal and external consultations. Therefore, demonstrating that consultations took place both among the member part of the associations acting as code owner(s), and with external stakeholders. These external stakeholders could be the clients of these organisations for example. In case such consultation did not take place, the code owner(s) should be able to demonstrate the lack of feasibility of it. [145]

2.1.8 – National Legislation Compliance and Language

The code owner(s) shall provide proof that the draft CoC is in compliance with national legislation applicable in the sector at stake. This requirement shall be intended as that the CoC must be in compliance with national data protection law or any other law applicable in the sector. As for the processing of health data, it is a domain already strongly regulated both by data protection law and other laws. In particular, the CoC shall take into account the provisions for processing health data at national level introduced in the space left by the GDPR for such purposes to national legislators. This point is critical especially for transnational CoC in the health domain, as it has been discussed in Chapter 3, Section 1.1. Indeed, being the health domain really fragmented from a legal perspective, an EU-wide transnational CoC shall take into account all these differences. This could essentially hamper the smooth functioning of the CoC. Or at least, it could be a heavy burden upon the code owner(s). In this regard, it shall be noted that an EU-wide CoC should probably identify more than one monitoring body. The reason is that it could be difficult to find a single body with the competences for monitoring compliance with so many national legislative frameworks. Therefore, besides the complexity during the drafting phase of an EU-wide transnational CoC, also the cost related to its monitoring could be elevated.

It shall also be noted that national CoC should be submitted in the national language of the CompSA, unless differently provided by the CompSA itself. Otherwise, transnational CoCs should be submitted in the language of the CompSA, or in English.

2.2 – Subject Matter of the Code of Conduct

Besides the formal mandatory parts of the CoC, which have been highlighted in the previous section, the CoC then shall discipline its subject matter. Therefore, the CoC shall indicate how the GDPR provision in the specific sector are further specified in order to reach compliance with data protection law. This part of the CoC shall satisfy the criteria indicated in the EDPB guidelines, which are 1) meeting a particular need of the sector; 2) facilitate the compliance; 3) specify the application of GDPR; 4) provide sufficient safeguards; 5) provide mechanisms for monitoring compliance with the CoC itself. However, these are just the objectives the CoC shall pursue.

As regard the subject matter of the CoC, it has been said that a CoC could cover different aspects of data processing, including essentially whichever aspects that could improve the implementation of the GDPR. [153] Moreover, it is not specified how this part of the CoC shall be organised.

The possible aspects that a CoC in the health sector could cover have been indicated in Chapter 3, Section 2. The rest of the section will indicate a possible organisation of the content, in order to specify the application of the GDPR in the sector, but also facilitate compliance activities. Namely, the proposed CoC could take the form of a set of guidelines, which should be meant to specify the meaning of GDPR provisions. Moreover, a checklist with a set of checks and controls could be used for facilitating the compliance activities of data controllers and processors[268]. Nevertheless, it shall be noted that the specific way to organise the content of a CoC should be discussed among the stakeholders participating in the drafting activity. Each CoC will have, therefore, its own structure. It should be clear that the structure here proposed is not claimed as the best or the only way to organise a CoC, but just a possible solution.

2.2.1 – Guidelines

The first part of a CoC should contain practical guidance. This part of the CoC is therefore meant to specify the application of the GDPR into the sector, but also to facilitate compliance with the law. In order to specify the application of GDPR into a specific sector, in the first place, some terminology could be clarified by the CoC. Therefore, a first part of the CoC could be a section dealing with the terminology applicable in the sector. The inclusion of a glossary might be the first step for specifying and, at the same time, clarifying the application of the GDPR in health domain. Most of the CoCs developed either in health domain or in other sectors include a glossary section[269]. The glossary should explain in plain language the meaning of data protection related terminology, in order to allow also non legal experts to understand the rationale behind the terms.

A second key part of the CoC is constituted by guidelines that explain in further detail the rationale of GDPR provisions and how to implement them in a better way. This part of the CoC shall contain practical explanation of the principles, trying as much as possible to translate them into the specific context and into practical operation of organisations adhering to the CoC[270].

Finally, as also suggested by the EDPB guidelines, a CoC might contain concrete examples of procedures and situations where the requirements are applied in a proper way. By this way, it is possible to provide further clarification to non-expert from legal perspective about how to implement the law. [145]

2.2.2 – Checklist

In conclusion, it has been noted that several CoCs come along with checklists. In such checklists there are checks and controls that help the data controllers and processor to reach compliance. Usually, these checklists are linked to certification mechanisms and standards

[268] Several CoC, although not in the health domain, have been organised in that way, see for example in the domain of cloud computing the CISPE CoC at https://cispe.cloud/code-of-conduct/ (last access: January 2023) or the EU Cloud CoC at https://eucoc.cloud/en/home (last access: January 2023).

[269] See for example the Farmaindustria CoC at https://www.aepd.es/en/informes-y-resoluciones/code-of-conduct (last access: January 2023) or the Code of Practice for data re-use of research https://www.imi.europa.eu/sites/default/files/uploads/documents/reference-documents/CodeofPractice_SecondaryUseDRAFT.pdf (last access: January 2023).

[270] See for example the Farmaindustria CoC, pages 26 and 27, where all the GDPR are explained in the context of clinical trials and clinical research. [232]

that can contribute to data protection law implementation[271]. Although, checklists are usually part of certification systems, it could be useful to use these systems for the quick identification of measures to apply even the context of CoC. Therefore, a data controller or processor could be guided, thanks to the checklist, with detailed information about practical measures to implement throughout the guidelines of the CoC.

To sum up, the first step (guidelines) is meant to create a proper understanding of the principles and tenets of the law within organisations. Then, in the second step (checklists), each principle could be linked to a set of checks and controls. The check and controls should then guide the data controller or processor in the application of specific technical and organisational measures. Often the checklists are used for the application of technical and organizational measures under article 32 GDPR, indeed, these checklists could also be replaced by already existing standards, such as ISO 27000 series for the security of information management[272].

[271] See, for example the CISPE CoC https://cispe.cloud/code-of-conduct/ (last access: January 2023).
[272] *Ibidem*.

Conclusion

The use of technologies based on big data analytics, such as AI systems in the health domain, can bring several advantages to clinical practice and research results. [6, 44, 51, 221] The possible AI applications in health care are numerous and promising in different areas, from diagnostic to public health or even assistance to the surgeon during operations. [57, 222, 223] Moreover, AI systems are also deployed in the context of observational studies or clinical trials. [7, 224]

The deployment of such technologies in medicine takes place however within a complex legal framework, which aims at safeguarding the safety of patients as well as protecting rights and freedoms related to privacy and personal life. Furthermore, an additional layer of legislation has been introduced by the EU legislator. The new data governance framework stemming from EC's strategy has the specific objective of incentivizing the sharing of data, especially in some contexts, such as the health domain. The strategy aims at boosting the use of big data technologies. The DGA has been the first piece of legislation stemming from the EC's strategy for data to be adopted. The DGA indeed set out a new framework for incentivising the sharing and the re-use of certain categories of data held by public sector bodies. Moreover, the sharing of electronic health data will be object of further rules and conditions as provided by the EHDS regulation. Compliance issues are therefore stemming from two sources: 1) the new layer of legal provisions for data sharing and its difficult interplay with GDPR; 2) the dichotomy between the main features of big data processing and the opaqueness of AI systems and some GDPR principles.

Against the above-mentioned scenario, the present work has analysed the potential application of regulatory models other than top-down regulations in the medical filed for enhancing data protection law provisions implementation. Namely, the two co-regulatory instruments provided by the GDPR, CoCs and CMs, have been analysed. From the analysis, it came up that the application of such co-regulatory measures essentially has some shortcomings, as well as promising results. The analysis has tried to take into account the characteristic of the sector at stake, i.e., the processing of personal data in the health domain. The sector analysed is carrier of issues and difficulties in terms of data protection law, which could be partially solved by the use of co-regulatory measures, as it is indeed envisaged by EU institutions and bodies. [80, 121] Nevertheless, some obstacles to the disposal of these instruments rely in the way the legislator has shaped these instruments into the GDPR but also among the features of the sector analysed itself.

At the EU level, the use of alternative regulatory solutions, has been promoted as general regulatory and policy strategy since the adoption of the better regulation approach. [108, 131] To some extent, the use of alternative regulatory solutions that include the involvement of private stakeholders, although under the supervision of public authorities, is an implementation of the principles of subsidiarity and proportionality. [155, 158, 225] The EC has been promoting such solutions as general approach in shaping the way legal act at EU level are written and implemented. The GDPR is an example of legislation carrier of a co-regulatory approach.

The GDPR follows a co-regulatory approach since it just sets out general principles and broad requirements that only identify objectives but do not specify how to reach them out. The choice of the way for reaching compliance with them, i.e., the specific technical and

organisational measures to implement, are left to the addressee of the law. The data controller (or processor) is therefore called to detect the proper measures to be implemented, on the basis, most of the time, of a risk-based analysis. Only through a proper documentation activity of such analysis concerning an evaluation of the risks involved in the data processing, the data controller can demonstrate compliance with data protection law and therefore fulfil the accountability principle. [18, 81–83]

Adopting such a regulatory approach, the EU legislator has decided not to deal with the details concerning the technical and organisational measures that an organisation shall perform in order to be considered in compliance. This effort has been moved upon the addressee of the law. This choice is meaningful in light of the complexities brought by the use of innovative technologies which are constantly re-shaping the way individuals and organisations behave in many sectors. Public authorities have indeed difficulties in regulating fast-evolving technologies, because of a lack of knowledge and resources. On the other hand, private stakeholders often have more know-how about innovative technologies. Moreover, there are arguments in the literature that claim how stakeholders might be more committed to stick to the rules that are elaborated by themselves, rather than by the State. [137] However, adopting such an approach is raising the burden upon data controllers and processors in terms of compliance costs. Indeed, the effort required to private parties in meeting regulatory objectives could be too onerous and therefore make the co-regulatory approach counterproductive.

For this reason, another layer of the concept of co-regulation can be envisaged, which is constituted by the compliance tools that fill the gap between the general and abstract rule or principle and the day-by-day implementation of the law into organisational procedures. The above-mentioned concept of co-regulation is an essential part of the more general co-regulatory approach that has been discussed before. Indeed, only if the second layer of co-regulation works properly the entire ecosystem of rules can produce beneficial effects. Otherwise, the addressee of the law could be left alone facing an enormous duty of compliance that he does not have the resources to fulfil. [20]

At this point, the key question is whether the co-regulatory tools enshrined into the GDPR fit the purpose of facilitating compliance activities left wide open by the more general co-regulatory approach in data protection law. Indeed, the present work has tried to assess the issue looking at the roles that such co-regulatory solutions can have in data protection law domain. Which are, as it has been said, smoothing compliance activities, but also creating legal certainty and creating competitive advantages on the market. Having in mind the role of such instruments it is also possible to note how the current application and diffusion of CoCs and CMs is still objectively rather limited[273].

Moving the discussion towards the sector of health data: this is one of the most critical domains in terms of risks and complexities related personal data processing. To investigate the reasons of the shortage of co-regulatory tools in this particular domain is key. Indeed, if, in general, it is possible to identify some problems common to every sector because they are directly related to the features of such co-regulatory instruments as envisaged in the GDPR. And these problems can essentially be related to the lack of incentives in terms of

[273] See the registry of the EDPB https://edpb.europa.eu/our-work-tools/accountability-tools/certification-mechanisms-seals-and-marks_en and at https://edpb.europa.eu/our-work-tools/accountability-tools/register-codes-conduct-amendments-and-extensions-art-4011_en (last access: January 2023).

return on the initial investment stemming from the adoption of a CoC or a CM, or to the long procedure for obtaining the endorsement from the DPA, or even to the unclear legal effects produced by the instruments[274]. However, the necessity to look at a specific sector is due to the peculiarities of each sector that make the application of data protection law so different from a domain to another.

The present work argues that the success or the failure of the use of such instruments in the health domain pass through a twofold analysis: 1) the choice of the scope of the co-regulatory instrument, both from a territorial and material point of view; 2) hypothesis on the possible impacts on stakeholders' behaviours.

The analysis of the geographical scope of the instrument is crucial because of the differences between MSs in terms of organisation of the health system. Such structural diversities also lead to differences in data protection law domain when it comes to the processing of health related, biometric and genetic data. These differences essentially create an obstacle to the development of co-regulatory instruments at EU level in the health domain. Different legal provisions at national level hamper the realisation of transnational instruments that have the goal of further specifying GDPR rules. For example, an EU-wide CoC, especially for facilitating the sharing of data among MSs for research purposes shall have to take into account several differences, in terms of legal basis for processing personal data, exercise of data subjects right and additional safeguards to be implemented. [80]

The elaboration of such a CoC should be therefore carefully evaluated, taking into account possible differences at national level. For instance, creating a "multiple layer CoC". Where at the first layer of compliance specification there is the implementation of the GDPR rules not subject to MSs additional provisions and another layer that take into account MSs' specifications. However, such a wide co-regulatory measure might be hard to develop because it would be difficult to gather enough stakeholders willing to take part to the project, with enough expertise for each MSs. Moreover, the costs and the investments related to this instrument would be for sure elevated.

As regards the material scope of the co-regulatory measure, the thesis has focused on just one instrument, i.e., CoCs rather than CMs. This is essentially due to the fact that the development of a CoC for the health domain and for the sharing of health data for research purposes is incentivised at the EU level by different EU institutions and bodies. [121] But, the choice is also due to some endeavours to develop such a CoC which have been launched, already. Although, none of them have never reached, so far, the final approval from the data protection authorities[275]. Therefore, several possible content of a CoC for health data processing have been detected. The data protection law issues that could become part of a CoC in the domain of health are numerous. Some of them have been suggested by studies commissioned by the EC [80], others are gathered from an analysis of the literature or from previous projects of CoCs in this domain.

The content of such a CoC therefore can include, for instance, aspects of anonymisation, pseudonymisation of personal data. In this sense, the meaning of these concepts could be clarified by a CoC in light of the specific domain at issue. Moreover, the procedures and techniques that could be used to properly implement anonymisation and pseudonymisation in compliance with the GDPR can be object of a CoC. Other aspects that can be touched

[274] See Chapter 3, Section 1.3.5.
[275] See above Chapter 2, Section 2.2.3.

by a CoC are the re-use of personal data for scientific research, or the new concept of data altruism and its relationship with consent under GDPR.

In general, a CoCs for health data processing should look at the evolving situation in terms of big data and AI application in the domain, as well as at the increased need for data sharing between MSs. However, the specific content of a CoC cannot be speculated in advance because too many factors could influence its content. The content of a CoC will be for sure influenced by the owners of the CoC and the interests pursued by them. Nevertheless, some general possible topic can be detected in advance, as it has been done in the present work.

Moreover, it shall be borne in mind that the content of a CoC shall be aligned with the EDPB guidelines as it has been widely discussed in this Chapter. Therefore, in defying the content of a CoC the stakeholders participating at its draft shall be able to demonstrate that the CoC is meant to meet a particular need in the first place. This first requirement, as discussed before, it is easy to demonstrate since the health system present several data protection compliance issue. Moreover, the content of the CoC shall facilitate the effective application of the GDPR as well as specify its requirements. The content of a CoC should indeed create high standards of compliance in the sector without losing of sight the objective of creating doable procedures and practices that can be performed by actors in the domain, without too high costs.

As regards the impacts on stakeholders' behaviours, this is one of the most critical points to discuss, because the use of co-regulatory solutions is a quite new phenomenon in data protection law. Therefore, there are few, if none, empirical data on the use of CoC or CMs in data protection law domain, especially concerning the effects that these instruments produce on stakeholders and on the ecosystem in general. However, some insights can be borrowed by other domains where the use of co-regulatory solutions, especially certifications have been spread for many years already. for example, some authors have suggested to take inspiration from the environmental domain. [105, 136]

The thesis has tried to look at some realistic expectations of impacts that could be produced by the use of these instruments. Moreover, having said that, all the speculations shall be confirmed or refuted by empirical analysis. The first point that would affect the diffusion of such instrument in the future is their ability to reduce the risk of fines. The discussion on this point is linked to the legal effect that these instruments have in the context of data protection law in terms of burden of proof and on the presumption of conformity. Unfortunately, the GDPR seems not to assign a strong presumption of conformity to co-regulatory solutions. Therefore, from this perspective, CoCs and certification mechanisms seem to be quite flawed. On the other hand, DPAs must take into account that a CoC or certification mechanism have been adopted, when deciding whether to issue a fine or the amount of it, according to article 83(2)(j) GDPR. However, it is not clear what this requirement exactly means in practical terms.

Another important point of the analysis is whether the legal fragmentation typical of the health domain will be an obstacle for the development of such instruments, especially CoCs. Otherwise, if these instruments will enable a reduction of the differences at national level.

In conclusion, a paradox seems to affect the use of co-regulatory solutions. Indeed, the development and the adoption of these instruments is costly in terms of monetary and human resources. However, the adoption of this instrument is ultimately promoted as a cost

saving solution or at least as enhancer of competitive advantages on the market. Essentially, the adoption of a CoC or a CM is an investment that can bring benefits to many stakeholders; although, as all the investments, it shall lead to a financial return for the organisations. To date, given the limited spread of these instruments it could be assumed that data controllers and processors are still sceptical about using these instruments in data protection law. However, also this aspect shall be analysed in light of future empirical data. For sure, the process for obtaining the approval from a data protection authority is a long and complicated process, and also some structural aspects of these instruments, as envisaged in the GDPR, are maybe discouraging their use.

It can be argued that the analysis carried out here has reached the following conclusions, which although cannot probably be considered definitive. Indeed, these conclusions should be rather used to address the future reflection on the topic. Namely, what came up is that the shift from a regulatory approach perspective, has made the use of meta-level rules such as co-regulatory solutions of the utmost importance. This is in order to fill the gap between general principles of the GDPR and technical and organisational measures. The use of these measures is supported and motivated by the more general co-regulatory approach in the better regulation strategy. However, the use of these instruments seems, so far at least, not as spread as it was hoped. Therefore, their impact in terms of compliance facilitation is limited. This situation is exacerbating the struggles that some actors are facing in terms of compliance duties. Especially in data intensive domains, such as the health domain, where the use of innovative technologies that rely on big data is becoming crucial. In this sense the lack of co-regulatory solutions seems to be hampering the proper functioning of the whole ecosystem. On the other hand, despite a theoretical assumption of necessity of these instruments, no empirical data is available for demonstrating the actual potential impacts in data protection law of these instruments. This is due to the relatively novelty of the use of such instruments in data protection law. Therefore, the caveat of this analysis is that the theoretical assumptions shall then be confirmed on an empirical level. Nevertheless, some hypotheses can been draft looking at other domains that have a similar regulatory approach and that apply co-regulatory measures as well.

The essential findings of this work are that for sure in light of the regulatory approach adopted by the EU legislators in data protection law, the use of such instruments is key for ensuring the smooth implementation of the law. Nevertheless, the instruments seem to have some structural flaws in terms of certainty of their legal effects. The instruments are indeed only facultative measures that can be adopted by data controllers and processors. Indeed, there is no obligation to be certified or to adopt a CoC for placing a product, or a service on the market, for example. The incentive for data controllers and processors in using the instruments is essentially linked either to competitive advantages on the market (especially using seals or marks), or to the reduction of risks to be fined, or finally to improvements in compliance activities procedures from an inner perspective.

Despite the potential impacts and roles envisaged for these instruments, the actual use of these instruments shall be evaluated against their capacity of becoming attractive for stakeholders on a practical level. This is a discourse also related to the culture of privacy and data protection that the GDPR will be able to create among organisations and individuals. If individuals do not perceive privacy and the protection of their personal data as an important asset, organisations processing personal data will not be engaging in the adoption of co-regulatory measure. Such instruments, indeed, although potentially beneficial for the

ecosystem and even economically advantageous for the organisation itself in the medium long term, generate an initial investment in compliance. The future of these instruments is strongly tied to the future approach of organisations towards data protection compliance. Namely, whether organizations will keep looking at data protection law as a mere cost or rather also as an investment.

Bibliography

1. Mehta N, Pandit A (2018) Concurrence of big data analytics and healthcare : A systematic review. Int J Med Inform 114:57–65. https://doi.org/10.1016/j.ijmedinf.2018.03.013

2. Kruse CS, Goswamy R, Raval Y, Marawi S (2016) Challenges and opportunities of big data in health care: A systematic review. JMIR Med Inform 4:. https://doi.org/10.2196/medinform.5359

3. OECD (2014) Data-driven Innovation for Growth and Well-being. Interim Synthesis Report.

4. Habl C, Renner AT, Bobek J, Laschkolnig A (2016) Study on Big Data in Public Health, Telemedine and Healthcare

5. He KY, Ge D, He MM (2017) Big data analytics for genomic medicine. Int J Mol Sci 18:. https://doi.org/10.3390/ijms18020412

6. Bates DW, Saria S, Ohno-Machado L, et al (2014) Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. Health Aff 7:1123–1131. https://doi.org/10.1377/hlthaff.2014.0041

7. Auffray C, Balling R, Barroso I, et al (2016) Making sense of big data in health research : Towards an EU action plan. Genome Med 1–13. https://doi.org/10.1186/s13073-016-0323-y

8. S. Berne A (2019) Big Data in the Pharmaceutical Sector. European Pharmaceutical Law Review 3:37–45. https://doi.org/10.21552/eplr/2019/1/9

9. Moustris GP, Hiridis SC, Deliparaschos KM, Konstantinidis KM (2011) Evolution of autonomous and semi-autonomous robotic surgical systems: A review of the literature. International Journal of Medical Robotics and Computer Assisted Surgery 7:375–392

10. Wagner M, Bihlmaier A, Kenngott HG, et al (2021) A learning robot for cognitive camera control in minimally invasive surgery. Surg Endosc. https://doi.org/10.1007/s00464-021-08509-8

11. Kassahun Y, Yu B, Tibebu AT, et al (2016) Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions. Int J Comput Assist Radiol Surg 11:553–568

12. Zarsky TZ (2017) Incompatible: The GDPR in the Age of Big Data. Seton Hall Law Rev 47:995–1020

13. European Commission (2010) A Digital Agenda for Europe

14. European Commission (2020) A European Strategy for Data. Brussels

15. EDPB-EDPS (2022) Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space

16. EDPB, EDPS (2021) Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act)

17. Mantelero A (2016) Personal data for decisional purposes in the age of analytics : From an individual to a collective. Computer Law & Security Review: The International Journal of Technology Law and Practice 32:238–255. https://doi.org/10.1016/j.clsr.2016.01.014

18. Mayer-Schonberger V, Padova Y (2016) Regime Change? Enabling Big Data Through Europe's new Data Protection Regulation. Columbia Sci Technol Law Rev XVII:315–335

19. Becker R, Chokoshvili D, Comandé G, et al (2022) Secondary Use of Personal Health Data: when is it "Further Processing" under the GDPR, and what are the Implications for Data Controllers. Eur J Health Law 30:129–157. https://doi.org/10.1109/JBHI.2015.2450362

20. Mantelero A (2017) Responsabilità e Rischio nel Reg. Ue 2016/679. Le nuove leggi civili commentate 1:144–164

21. Kamara I (2020) Article 40. Codes of Conduct. In: Kuner C, Bygrave LA, Docksey C (eds) The EU General Data Protection Regulation (GDPR). Oxford University Press, pp 716–724

22. Leenes R (2020) Article 42. Certification. In: Kuner C, Bygrave LA, Docksey C (eds) The EU General Data Protection Regulation (GDPR). Oxford University Press, pp 732–743

23. Lachaud E (2018) The General Data Protection Regulation and the rise of certification as a regulatory instrument. Computer Law and Security Review 34:244–256. https://doi.org/10.1016/j.clsr.2017.09.002

24. Vander Maelen C (2020) Codes of (Mis) conduct ? An Appraisal of Articles 40-41 GDPR in View of the 1995 Data Protection Directive and Its Shortcomings. European Data Protection Law Review 6:231–242. https://doi.org/10.21552/edpl/2020/2/9

25. Russell S, Norvig P (2021) Artificial Intelligence. A Modern Approach, Fourth Edition. Pearson

26. Pasquale F (2013) Grand Bargains for Big Data: The Emerging Law of Health Information. Maryland Law Review 72:682–772

27. Wang H, Xu Z, Fujita H, Liu S (2016) Towards felicitous decision making: An overview on challenges and trends of Big Data. Inf Sci (N Y) 367–368:747–765. https://doi.org/10.1016/j.ins.2016.07.007

28. Janssen M, van der Voort H, Wahyudi A (2017) Factors influencing big data decision-making quality. J Bus Res 70:338–345. https://doi.org/10.1016/j.jbusres.2016.08.007

29. Duan Y, Edwards JS, Dwivedi YK (2019) Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. Int J Inf Manage 48:63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

30. Panesar A (2021) Machine Learning and AI for Healthcare. Big Data for Improved Health Outcomes. Apress

31. McKinsey&Company (2011) Big data: The next frontier for innovation, competition, and productivity

32. European Parliament (2017) European Parliament resolution of 14 March 2017 on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement (2016/2225(INI))

33. Reinsel D, Gantz J, Rydning J (2018) IDC Data Age 2025. The Digitization of the World. From Edge to Core

34. Agcm-Agcom-GPDP (2020) Indagine Conoscitiva sui Big Data

35. AGCM, AGCOM, GPDP (2019) Big data joint survey. Guidelines and policy recommendations.

36. Kahn MG, Raebel MA, Glanz JM, et al (2012) A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Med Care 50:. https://doi.org/10.1097/MLR.0b013e318257dd67

37. Kahn MG, Callahan TJ, Barnard J, et al (2016) A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. eGEMs (Generating Evidence & Methods to improve patient outcomes) 4:18. https://doi.org/10.13063/2327-9214.1244

38. Chen H, Hailey D, Wang N, Yu P (2014) A review of data quality assessment methods for public health information systems. Int J Environ Res Public Health 11:5170–5207

39. Panesar A (2019) Machine Learning and AI for Healthcare: Big Data for Improved Health Outcomes. Apress Media LLC

40. Reinsel D, Gantz J, Rydning J (2017) IDC Data Age 2025: The Evolution of Data to Life-Critical. Don’t Focus on Big Data; Focus on the Data That’s Big.

41. OECD (2019) Health in the 21st Century Putting Data to Work for Stronger Health Systems

42. Dinov ID (2016) Volume and value of big healthcare data. J Med Stat Inform 4:1–7. https://doi.org/10.7243/2053-7662-4-3

43. He KY, Ge D, He MM (2017) Big data analytics for genomic medicine. Int J Mol Sci 18:. https://doi.org/10.3390/ijms18020412

44. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare : promise and potential. Health Inf Sci Syst 2:1–10

45. Scruggs SB, Watson K, Su AI, et al (2015) Harnessing the heart of big data. Circ Res 116:1115–1119. https://doi.org/10.1161/CIRCRESAHA.115.306013

46. Ghani KR, Zheng K, Wei JT, Friedman CP (2014) Harnessing Big Data for Health Care and Research : Are Urologists Ready? Eur Urol 66:975–977. https://doi.org/10.1016/j.eururo.2014.07.032

47. Baro E, Degoul S, Beuscart R, Chazard E (2015) Toward a Literature-Driven Definition of Big Data in Healthcare. Biomed Res Int 1–10

48. Shilo S, Rossman H, Segal E (2020) Axes of a revolution: challenges and promises of big data in healthcare. Nat Med 26:29–38. https://doi.org/10.1038/s41591-019-0727-5

49. Comandè G, Schneider G (2018) Regulatory Challenges of Data Mining Practices: The Case of the Never-ending Lifecycles of "Health Data." Eur J Health Law 25:284–307. https://doi.org/10.1163/15718093-12520368

50. OECD (2017) New Health Technologies

51. OECD (2013) Strengthening Health Information Infrastructure for Health Care Quality Governance

52. OECD (2015) Health Data Governance

53. Aurucci P (2019) Legal issues in regulating observational studies: The impact of the gdpr on Italian biomedical research. European Data Protection Law Review 5:197–208. https://doi.org/10.21552/edpl/2019/2/9

54. Deverka PA, Majumder MA, Villanueva AG, et al (2017) Creating a data resource : what will it take to build a medical information commons ? Genome Med 9:1–5. https://doi.org/10.1186/s13073-017-0476-3

55. Evans BJ (2011) Much Ado about Data Ownership. Harv J Law Technol 25:69–130

56. El Emam K, Arbuckle L (2013) Anonymizing Health Data. O'Reilly, Sebastopol, CA

57. AHSN Network, Deapartment of Health & Social Care (2018) Accelerating Artificial Intelligence in health and care: results from a state of the nation survey

58. Rumsfeld JS, Joynt KE, Maddox TM (2016) Big data analytics to improve cardiovascular care: Promise and challenges. Nat Rev Cardiol 13:350–359. https://doi.org/10.1038/nrcardio.2016.42

59. Zhang D, Maslej N, Brynjolfsson E, et al (2022) Artificial Intelligence Index Report 2022

60. Leenes R (2007) Do They Know Me? Deconstructing Identifiability. University of Ottawa Law & Technology Journal 4:135–161

61. Finck M, Pallas F (2020) They Who Must Not Be Identified - Distinguishing Personal from Non-Personal Data Under the GDPR. International Data Privacy Law 10:11–36. https://doi.org/10.2139/ssrn.3462948

62. Purtova N (2018) The law of everything . Broad concept of personal data and future of EU data protection law. Law Innov Technol 10:40–81. https://doi.org/10.1080/17579961.2018.1452176

63. Langarizadeh M, Orooji A, Sheikhtaheri A (2018) Effectiveness of Anonymization Methods in Preserving Patients ' Privacy : A Systematic Literature Review. In: Health Informatics Meets eHealth. pp 80–87

64. El Emam K, Alvarez C (2015) A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques. International Data Privacy Law 5:73–87. https://doi.org/10.1093/idpl/ipu033

65. Ohm P (2010) Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation. UCLA Law Review 57:1701

66. Stalla-bourdillon S, Knight A (2017) Anonymous Data v. Personal Data - a False Debate: an EU Perspective on Anonymization, Pseudonymization and Personal Data. Wisconsin International Law Journal 34:

67. Su J, Shukla A, Goel S, Narayanan A (2017) DE-anonymizing Web Browsing Data with Social Networks. In: International World Wide Web Conference Committee (IW3C2)

68. Rocher L, Hendrickx JM, de Montjoye YA (2019) Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun 10:. https://doi.org/10.1038/s41467-019-10933-3

69. de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD (2013) Unique in the Crowd: The privacy bounds of human mobility. Sci Rep 3:. https://doi.org/10.1038/srep01376

70. Narayanan A, Shmatikov V (2008) Robust de-anonymization of large sparse datasets. In: Proceedings - IEEE Symposium on Security and Privacy. pp 111–125

71. Article 29 Data Protection Working Party (2014) Opinion 05/2014 on Anonymisation Techniques

72. Corrales Compagnucci M, Meszaros J, Minssen T, et al (2019) Homomorphic Encryption: The 'Holy Grail' for Big Data Analytics and Legal Compliance in the Pharmaceutical and Healthcare Sector? European Pharmaceutical Law Review 3:144–155. https://doi.org/10.21552/eplr/2019/4/5

73. Pagallo U, Casanovas P, Madelin R (2019) The middle-out approach: assessing models of legal governance in data protection, artificial intelligence, and the Web of Data. Theory and Practice of Legislation 7:1–25. https://doi.org/10.1080/20508840.2019.1664543

74. Marelli L, Lievevrouw E, van Hoyweghen I (2020) Fit for purpose? The GDPR and the governance of European digital health. Policy Studies 41:447–467. https://doi.org/10.1080/01442872.2020.1724929

75. Moerel L (2014) Big Data Protection: how to make the draft EU Regulation on Data Protection Future Proof. 1–68

76. Moerel L, Prins C (2016) Privacy for the homo digitalis. Proposal for a new regulatory framework for data protection in light of Big Data and Internet of Things. 1–98

77. European Commission (2012) Safeguarding Privacy in a Connected World A European Data Protection Framework for the 21st Century

78. European Commission (2012) Executive Summary of the Impact Assessment SEC(2012)73 final

79. Solove DJ (2013) Introduction: Privacy Self-Managemet and the Consent Dilemma. Harv Law Rev 126:1880–1903

80. European Commission (2021) Assessment of the EU Member States' rules on health data in the light of GDPR. Brussels

81. Gellert R (2017) Why the GDPR risk-based approach is about compliance risk and why it's not a bad thing. In: Schweighofer E, Kummer FsC (eds) Trends and Communities of legal informatics: IRIS 2017 -Proceedings of the 20th International Legal Informatics Symposium. pp 527–532

82. Gellert R (2018) Understanding the notion of risk in the General Data Protection Regulation. Computer Law and Security Review 34:279–288. https://doi.org/10.1016/j.clsr.2017.12.003

83. Gellert R (2016) We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection. European Data Protection Law Review 2:481–492

84. Gellert R (2018) Understanding the notion of risk in the General Data Protection Regulation. Computer Law and Security Review 34:279–288. https://doi.org/10.1016/j.clsr.2017.12.003

85. van Dijk N, Gellert R, Rommetveit K (2016) A risk to a right? beyond data protection risk assessments. Computer Law and Security Review 32:286–306. https://doi.org/10.1016/j.clsr.2015.12.017

86. Article 29 data protection working party (2014) Statement on the role of a risk-based approach in data protection legal frameworks

87. Covello VT, Mumpower J (1985) Risk Analysis and Risk Management: An Historical Perspective. Risk Analysis 5:103–120. https://doi.org/10.1111/j.1539-6924.1985.tb00159.x

88. Bernestein PL (1996) Against the Gods: The Remarkable Story of Risk. John Wiley and Sons, New York

89. ISO - International Standardisation Organisation (2018) ISO 31000:2018 - Risk management - Guidelines

90. Article 29 Data Protection Working Party (2013) Opinion 04/2013 on the Data Protection Impact Assessment Template for Smart Grid and Smart Metering Systems ('DPIA Template') prepared by Expert Group 2 of the Commission's Smart Grid Task Force Adopted on 22 April 2013 - WP205

91. Article 29 Data Protection Working Party (2011) Opinion 9/2011 on the revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications - WP180

92. Murphy T, Whitty N (2009) Is human rights prepared? Risk, rights and public health emergencies. Med Law Rev 17:219–244. https://doi.org/10.1093/medlaw/fwp007

93. Commission Nationale Informatique & Libertés - CNIL (2018) Privacy Impact Assessment (PIA) - Knowledge Bases

94. Commission Nationale Informatique & Libertés - CNIL (2018) Privacy Impact Assessment (PIA) - Methodology

95. European Union Agency for Network and Information Security - ENISA (2017) Handbook on Security of Personal Data Processing.

96. European Commission (2020) White Paper On Artificial Intelligence - A European approach to excellence and trust. COM(2020) 65 final

97. Independent High Level Expert Group on Artificial Intelligence (AI HLEG) (2019) Ethics Guidelines for Trustworthy AI. European Commission 1–39

98. Independent High-Level Expert Group on Artificial Intelligence (2020) Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment

99. 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

100. European Commission (2021) Annexes to the 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

101. Palmieri S, Walraet P, Goffin T (2021) Inevitable influences: AI-based medical devices at the intersection of medical devices regulation and the proposal for AI regulation. Eur J Health Law 1–18. https://doi.org/10.1163/15718093-bja10053

102. Veale M, Borgesius FZ (2021) Demystifying the Draft EU Artificial Intelligence Act. 1–26

103. Van Kolfschooten H (2022) EU Regulation of Artificial Intelligence: Challenges for Patients' Rights. Common Market Law Review 59:81–112

104. McFadden M, Jones K, Taylor E, Osborn G (2021) Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation. Oxford Commission on AI & Good Governance

105. Matus KJM, Veale M (2022) Certification systems for machine learning: Lessons from sustainability. Regul Gov 16:177–196. https://doi.org/10.1111/rego.12417

106. Leistner M (2021) The Commission's vision for Europe's Digital Future: Proposals for the Data Governance Act, the Digital Markets Act and the Digital Services Act-A critical primer. Journal of Intellectual Property Law & Practice 16:778–784

107. European Commission (2020) Shaping Europe's Digital Future

108. EU Commission (2002) COM(2002) 275 final. European Governance: Better Lawmaking.

109. European Commission (2020) Proposal for a Regulation on European Data Governance (Data Governance Act). Brussels 1–43

110. European Commission (2022) Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM(2022) 197 final 1–122

111. Shabani M (2021) The Data Governance Act and the EU's move towards facilitating data sharing. Mol Syst Biol 17:. https://doi.org/10.15252/msb.202110229

112. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al (2016) Comment: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:1–9. https://doi.org/10.1038/sdata.2016.18

113. Rieder G (2018) Tracing Big Data Imaginaries through Public Policy: The Case of the European Commission. In: Rudinow Sætnan A, Schneider I, Green N (eds) The Politics and Policies of Big Data, 1st ed. Routledge, London, pp 89–109

114. Shabani M (2022) Will the European Health Data Space change data sharing rules? Science (1979) 375:1357–1359. https://doi.org/10.1126/science.abn4874

115. Shabani M, Yilmaz S (2022) Lawfulness in secondary use of health data. Interplay between three regulatory frameworks. Technology and Regulation 8:128–134. https://doi.org/10.7189/jogh.08.010502

116. Kruesz C, Zopf F (2021) The Concept of Data Altruism of the draft DGA and the GDPR: Inconsistencies and Why a Regulatory Sandbox Model May Facilitate Data Sharing in the EU. European Data Protection Law Review 7:569–579. https://doi.org/10.21552/edpl/2021/4/13

117. Martinsen DS, Schrama R, Mastenbroek E (2021) Experimenting European healthcare forward. Do institutional differences condition networked governance? J Eur Public Policy 28:1849–1870. https://doi.org/10.1080/13501763.2020.1804436

118. Stellmach C, Muzoora MR, Thun S (2022) Digitalization of Health Data: Interoperability of the Proposed European Health Data Space. In: Studies in Health Technology and Informatics. IOS Press BV, pp 132–136

119. Bentzen HB, Castro R, Fears R, et al (2021) Remove obstacles to sharing health data with researchers outside of the European Union. Nat Med 27:1329–1333

120. Eiss R (2020) Confusion over data-privacy law stalls scientific progress. Nature 584:498–498

121. EDPS (2020) Preliminary Opinion 8/2020 on the European Health Data Space. Brussels

122. Baloup J, Bayamlıoğlu E, Benmayor A, et al (2021) CiTiP Working Paper Series White Paper on the Data Governance Act White Paper on the Data Governance Act

123. Gellert R, Graef I (2021) TILEC Discussion Paper The European Commission's proposed Data Governance Act: some initial reflections on the increasingly complex EU regulatory puzzle of stimulating data sharing

124. Comandè G, Schneider G (2022) Differential Data Protection Regimes in Data-driven Research: Why the GDPR is More Research-friendly Than You Think. German law Journal 23:559–596. https://doi.org/10.1093/idpl/ipaa005/5813830?redirectedFrom=fulltext

125. Starkbaum J, Felt U (2019) Negotiating the reuse of health-data: Research, Big Data, and the European General Data Protection Regulation. Big Data Soc 6:. https://doi.org/10.1177/2053951719862594

126. 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. https://doi.org/10.1038/s41431-019-0386-5

127. EDPB (2020) Guidelines 05/2020 on consent under Regulation 2016/679

128. Kiseleva A, de Hert P (2021) Creating a European Health Data Space: Obstacles in Four Key Legal Area'. European Pharmaceutical Law Review (EPLR) 5:21–36. https://doi.org/10.21552/epIr/2021/1/5

129. EDPS (2020) A Preliminary Opinion on data protection and scientific research

130. European Parliament, Council, Commission (2003) European Parliament, Council, Commission interinstitutional agreement on better law-making (2003/C 321/01). Official Journal of the European Union 1–5

131. European Parliament, Council of the European Union, European Commission (2016) Interinstitutional Agreements of the European Union and the European Commission on Better Law-Making. Official Journal of the European Union 1–14

132. Ladeur K-H (2021) Governance, Theory of. Max Planck Encyclopedias of International Law 1–25

133. Keohane R (2002) Power and Governance in a Partially Globalized World, 1st ed. Routledge, London

134. Slaughter A-M (2009) A New World Order. Princeton University Press

135. Gunningham N (1998) Environmental Management Systems and Community Participation: Rethinking Chemical Industry Regulation. UCLA Journal of Environmental Law and Policy 16:. https://doi.org/10.5070/l5162018936

136. Hirsch DD (2006) Protecting the Inner Environment: what Privacy Regulation can Learn from Environmental Law. Georgia Law Review 41:

137. Hirsch DD (2013) In Search of the Holy Grail: Achieving Global Privacy Rules Through Sector-Based Codes of Conduct. Ohio State Law J 74:1029–1070

138. Haufler V (2001) A Public Role for the Private Sector: Industry Self-Regulation in a Global Economy. Carnegie Endowment for International Peace

139. Organisation for Economic Co-operation and Development (2015) Data-Driven Innovation. Big Data for Growth and Well-Being.

140. Federal Trade Commission (2016) Big Data. A Tool for Inclusion or Exclusion?

141. Polonetsky J, Tene O (2013) Privacy and Big Data: Making Ends Meet. Stanford Law Rev 66:25–33

142. Tene O, Polonetsky J (2012) Privacy in the Age of Big Data: A Time for Big Decisions. Stanford Law Review Online 64:63–69

143. Marsden CT (2012) Internet co-regulation and constitutionalism: Towards European judicial review. International Review of Law, Computers and Technology 26:211–228. https://doi.org/10.1080/13600869.2012.698450

144. Kamara I (2017) Co-regulation in EU personal data protection: the case of technical standards and the privacy by design standardisation "mandate." European Journal of Law and Technology 8:1–24

145. EDPB (2019) Guidelines1/2019 onCodes of Conduct and MonitoringBodies under Regulation 2016/679. Version 2.0

146. EDPB (2019) Guidelines 1/2018 on certification and identifying certification criteria in accordance with Articles 42 and 43 of the Regulation. Version 3.0

147. Price ME, Verhulst S (2000) The Concept of Self-Regulation and the Internet. In: Waltermann J, Machill M (eds) Protecting our children on the internet: Towards a new culture of responsibility. pp 133–198

148. Gunningham N, Rees J (1997) Industry self-regulation: An institutional perspective. Law Policy 19:363–414. https://doi.org/10.1111/1467-9930.t01-1-00033

149. Bennett CJ, Raab C (2006) The Governance of Privacy Policy Instruments in Global PerspectivThe Governance of Privacy Policy Instruments in Global Perspective. MIT Press

150. Priest M (1998) The Privatization of Regulation: Five Models of Self-Regulation. Ottawa Law Rev 29:233–307

151. Rubinstein IS (2011) Privacy and Regulatory Innovation: Moving Beyond Voluntary Codes. I/S: a journal of law and policy for the information society 6:

152. Gunningham N, Sinclair D (2002) Leaders and Laggards. Next-Generation Environmental Regulation. Routledge, London

153. Lachaud E (2019) Adhering to GDPR codes of conduct: A possible option for SMEs to GDPR certification. Tilburg Law School Research Paper

154. Lachaud E (2016) Why the certification process defined in the General Data Protection Regulation cannot be successful. Computer Law and Security Review 32:814–826. https://doi.org/10.1016/j.clsr.2016.07.001

155. Senden L (2005) Soft Law, Self-regulation and Co-regulation in European Law: Where Do They Meet? Electronic Journal of Comparative Law 9:

156. Fabbrini F (2018) The Principle of Subsidiarity. In: Oxford Principles Of European Union Law: The European Union Legal Order: Volume I. Oxford University Press

157. Feichtner I (2007) Subsidiarity. Max Plack Encyclopedias of International Law 1–10

158. Tridimas T (2018) The Principle of Proportionality. In: Schütze R, Tridimas T (eds) Oxford Principles Of European Union Law: The European Union Legal Order. Oxford University Press

159. Black J (1996) Constitutionalising Self-Regulation. Modern Law Review 59:24–55

160. Ayres I, Braithwaite J (1992) Responsive Regulation. Trascending the Deregulation Debate. Oxford University Press, New York

161. Latzer M, Price ME, Saurwein F, et al (2007) Comparative Analysis of International Co- and Self-Regulation in Communications Markets. Vienna

162. EDPB (2019) Guidelines1/2019 onCodes of Conduct and MonitoringBodies under Regulation 2016/679

163. Working Party on the Protection of Individuals with regard to the processing of Personal data (1998) Future work on codes of conduct: Working Document on the procedure for the consideration by the Working Party of Community codes of conduct

164. Article 29 Data protection Working Party (2003) Opinion 3/2003 on the European code of conduct of FEDMA for the use of personal data in direct marketing

165. Article 29 Data Protection Working Party (2009) Second opinion 4/2009 on the World Anti-Doping Agency (WADA) International Standard for the Protection of Privacy and Personal Information, on related provisions of the WADA Code and on other privacy issues in the context of the fight against doping in sport by WADA and (national) anti-doping organizations

166. Article 29 Data Protection Working Party (2015) Opinion 02/2015 on C-SIG Code of Conduct on Cloud Computing

167. Bahr A, Schlünder I (2015) Code of practice on secondary use of medical data in European scientific research projects. International Data Privacy Law 5:279–291

168. (2014) Code of Practice on Secondary Use of Medical Data in Scientific Research Projects - 27 Aug 2014 FINAL DRAFT

169. Garante per la Protezione dei Dati Personali (2018) Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica effettuati nell'ambito del Sistema Statistico nazionale pubblicate ai sensi dell'art. 20, comma 4,del d.lgs. 10 agosto 2018, n. 101 - 19 dicembre 2018 [9069677]

170. Garante per la Protezione dei dati Personali (2018) Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica pubblicate ai sensi dell'art. 20, comma 4, del d.lgs. 10 agosto 2018, n. 101 - 19 dicembre 2018 [9069637]

171. Garante per la Protezione dei Dati Personali (2018) Regole deontologiche relative al trattamento di datipersonali nell'esercizio dell'attività giornalistica pubblicateai sensi dell'art. 20, comma 4, del d.lgs. 10 agosto 2018, n.101 - 29 novembre 2018 [9067692]

172. Garante per la Protezione dei Dati Personali (2018) Regole deontologiche per il trattamento a fini di archiviazione nel pubblico interesse o per scopi di ricerca storica pubblicate ai sensi dell'art. 20, comma 4, del d.lgs. 10agosto 2018, n. 101 - 19 dicembre 2018 [9069661]

173. Garante per la Protezione dei Dati Personali (2019) Codice di condotta per i sistemi informativi gestiti da soggetti privati in tema di crediti al consumo, affidabilità e puntualità nei pagamenti [9141941]

174. Garante per la Protezione dei Dati Personali (2019) Deliberazione del 12 giugno 2019-Codice di condotta per il trattamento dei dati personali in materia di informazioni commerciali [9119868]

175. Garante per la Protezione dei Dati Personali (2018) Regole deontologiche relative ai trattamenti di dati personali effettuati per svolgere investigazioni difensive o per fare valere o difendere un diritto in sede giudiziaria pubblicate ai sensi dell'art. 20, comma 4, del d.lgs. 10 agosto 2018, n. 101 -19 dicembre 2018 [9069653]

176. Kamara Irene, Leenes Ronald, Lachaud Eric, et al (2019) Data protection certification mechanisms : study on Articles 42 and 43 of the Regulation (EU) 2016/679 : final report.

177. Kamara I (2020) Article 41. Monitoring of approved codes of conduct. In: Kuner C, Bygrave LA, Docksey C (eds) The EU General Data Protection Regulation (GDPR). Oxford University Press, pp 725–731

178. BS EN ISO/IEC 17000:2020 (2020) Conformity Assessment - Vocabulary and general principles

179. Garante per la Protezione dei Dati Personali (2021) Provvedimento del 14 gennaio 2021-Regione Veneto. Codice di condotta per l'utilizzo di dati sulla salute a fini didattici e di pubblicazione scientifica [9535354]

180. Kamara I, Leenes R, Lachaud E, et al (2019) Data Protection Certification Mechanisms-Study on Articles 42 and 43 of the Regulation (EU) 2016/679

181. Article 29 Data Protection Working Party (2009) The Future of Privacy Joint contribution to the Consultation of the European Commission on the legal framework for the fundamental right to protection of personal data

182. EDPB (2022) Opinion 1/2022 on the draft decision of the Luxembourg Supervisory Authority regarding the GDPR-CARPA certification criteria

183. EDPB (2020) EDPB Document on the procedure for the approval of certification criteria by the EDPB resulting in a common certification, the European Data Protection Seal

184. EDPB (2019) Guidelines 4/2018 on the accreditation of certification bodies under Article 43 of the General Data Protection Regulation (2016/679). Version 3.0

185. Leenes R (2020) Article 43. Certification bodies. In: Kuner C, Bygrave LA, Docksey C (eds) The EU General Data Protection Regulation (GDPR). Oxford University Press, pp 744–754

186. Commission nationale pour la protection des données (2022) Décision N° 15/2022 du 13 mai 2022

187. EDPB (2022) Opinion 28/2022 on the Europrivacy criteria of certification regarding their approval by the Board as European Data Protection Seal pursuant to

188. EDPB (2022) Opinion 25/2022 regarding the European Privacy Seal (EuroPriSe) certification criteria for the certification of processing operations by processors

189. Ang PH (2001) The Role of Self-Regulation of Privacy and the Internet. Journal of Interactive Advertising 1:1–9. https://doi.org/10.1080/15252019.2001.10722046

190. Article 29 Data Protection Working Party (2010) Opinion 3/2010 on the principle of accountability

191. Von Grafenstein M (2021) Co-Regulation and the Competitive Advantage in the GDPR: Data protection certification mechanisms, codes of conduct and the "state of the art" of data protection-by-design. In: Research Handbook on Privacy and Data Protection Law: Values, Norms and Global Politics

192. EDPB-EDPS (2019) EDPB-EDPS Joint Opinion 1/2019 on the processing of patients' data and the role of the European Commission within the eHealth Digital Service Infrastructure (eHDSI). 1–9

193. European Commission (2019) Taking Stock of the Commission's Agenda Better Regulation

194. European Commission (2019) Better regulation: taking stock and sustaining our commitment

195. European Commission (2019) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Better regulation: taking stock and sustaining our commitment. COM(2019)178 final. 1–14

196. Auffray C, Balling R, Barroso I, et al (2016) Making sense of big data in health research: Towards an EU action plan. Genome Med 8:. https://doi.org/10.1186/s13073-016-0323-y

197. Woolf SH (2008) The Meaning of Translational Research and Why It Matters. JAMA 299:211–213

198. Custer B, Vrabec HU, Friedewald M (2019) Assessing the Legal and Ethical Impact of Data Reuse : Developing a Tool for Data Reuse Impact Assessments (DRIA). European Data Protection Law Review 3:317–337. https://doi.org/10.21552/edpl/2019/3/7

199. Auld G, Gulbrandsen LH, McDermott CL (2008) Certification schemes and the impacts on forests and forestry. Annu Rev Environ Resour 33:187–211. https://doi.org/10.1146/annurev.environ.33.013007.103754

200. Bernstein S, Cashore B (2012) Complex global governance and domestic policies: four pathways of influence. Int Aff 88:585–604

201. Barry M, Cashore B, Clay J, et al (2012) Toward sustainability: The roles and limitations of certification. Washington

202. Powles J, Hodson H (2017) Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 7:351–367. https://doi.org/10.1007/s12553-017-0179-1

203. King D, Karthikesalingam A, Hughes C, et al (2017) Letter in response to Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl) 7:351–367. https://doi.org/10.1007/s12553-017-0179-1

204. Article 29 Data Protection Working Party (2017) Guidelines on the application and setting of administrative fines for the purposes of the Regulation 2016/679

205. ICO (2021) Introduction to anonymisation Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance

206. ICO (2021) Chapter 2: How do we ensure anonymisation is effective?

207. AEPD (2020) Misunderstanding Related to Anonymisation. Nature Publishing Group

208. European Union Agency for Cybersecurity (2019) Pseudonymisation techniques and best practices

209. Marcus JS, Martens B, Carugati C, et al (2022) The European Health Data Space. Publication for the committee on Industry, Research and Energy (ITRE)

210. Corrales Compagnucci M, Meszaros J, Minssen T, et al (2019) Homomorphic Encryption: The 'Holy Grail' for Big Data Analytics and Legal Compliance in the Pharmaceutical and Healthcare Sector? European Pharmaceutical Law Review 3:144–155. https://doi.org/10.21552/eplr/2019/4/5

211. Article 29 Data Protection Working Party (2007) Opinion 4/2007 on the concept of personal data

212. Mourby M, Mackey E, Elliot M, et al (2018) Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK. Computer Law and Security Review 34:222–233. https://doi.org/10.1016/j.clsr.2018.01.002

213. Emam K El (2013) Guide to the De-Identification of Personal Health Information

214. El Emam K, Jonker E, Arbuckle L, Malin B (2011) A systematic review of re-identification attacks on health data. PLoS One 6

215. Spindler G, Schmechel P (2016) Personal Data and Encryption in the European General Data Protection Regulation. Journal of Intellectual Property, Information Technology and Electronic Commerce Law 7:163–177

216. Groos D, van Veen E (2020) Anonymised Data and the Rule of Law. European Data Protection Law Review 4:498–508. https://doi.org/10.21552/edpl/2020/4/6

217. Article 29 Data Protection Working Party (2013) Opinion 03/2013 on purpose limitation

218. Kohane IS, Altman RB (2005) Health-Information Altruists-A Potentially Critical Resource. N Engl J Med 353:2074–2077

219. Garben S (2019) Article 168 TFEU. In: The EU Treaties and the Charter of Fundamental Rights: A Commentary. pp 1445–1455

220. Panagopoulos A, Minssen T, Sideri K, et al (2022) Incentivizing the sharing of healthcare data in the AI Era. Computer Law and Security Review 45:. https://doi.org/10.1016/j.clsr.2022.105670

221. Roski BJ, Bo-linn GW, Andrews TA (2014) Creating Value In Health Care Through Big Data: Opportunities And Policy Implications. Health Aff 7:1115–1122. https://doi.org/10.1377/hlthaff.2014.0147

222. O'Sullivan S, Nevejans N, Allen C, et al (2019) Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery. The International Journal of Medical Robotics and Computer Assisted Surgery 15:e1968. https://doi.org/10.1002/rcs.1968

223. Briganti G, le Moine O (2020) Artificial Intelligence in Medicine: Today and Tomorrow. Front Med (Lausanne) 7:. https://doi.org/10.3389/fmed.2020.00027

224. Bernier A, Knoppers BM (2020) Pandemics, privacy, and public health research. Canadian Journal of Public Health 111:454–457. https://doi.org/10.17269/s41997-020-00368-5

225. Fabbrini F (2018) The Principle of Subsidiarity. In: Schütze R, Tridimas T (eds) Oxford Principles Of European Union Law: The European Union Legal Order. Oxford University Press

226. OECD. (2014) OECD Regulatory Compliance Cost Assessment Guidance. OECD Publishing

227. Al-rawashdeh M, Keikhosrokiani P, Belaton B, et al (2022) IoT Adoption and Application for Smart Healthcare: A Systematic Review. Sensors 22

228. Information Comissioners Office (2012) Anonymisation: managing data protection risk code of practice

229. European Parliament (2022) Artificial Intelligence in Healthcare. Applications, Risks, and Ethical and Societal Impacts.

230. Kessler J (2019) Data Protection in the Wake of the GDPR: California's Solution for Protecting "the World's Most Valuable Resource." South Calif Law Rev 93:99–128

231. Erickson A (2019) Comparative Analysis of the EU's GDPR and Brazil's LGPD: Enforcement Challenges with the LGPD. Brooklyn J Int Law 44:859–888

232. Farmaindustria (2022) Code of Conduct Regulating the Processing of Personal Data in Clinical Trials and Other Clinical Research and Pharmacovigilance Activities

233. Scope Europe (2020) EU Cloud Code of Conduct EU Data Protection Code of Conduct for Cloud Service Providers

234. CISPE.cloud (2021) Data Protection Code of Conduct for Cloud Infrastructure Service Providers

List of abbreviations

AEPD – Agencia Española Protección Datos

AI – Artificial Intelligence

AIA – Artificial Intelligence Act

AI-HLEG – AI High Level Expert Group

CM – Certification Mechanism

CNIL - Commission Nationale de l'Informatique et des Libertés

CNPD – Commission Nationale pour la Protection des Données

CoC – Code of Conduct

CompSA – Competent Supervisory Authority

DA – Data Act

DGA – Data Governance Act

D.lgs – Decreto Legislativo

DMA – Digital Markets Act

DPA – Data Protection Authority

DSA – Digital Services Act

DPD – Data Protection Directive

DPIA – Data Protection Impact Assessment

EC – European Commission

EDIB – European Data Innovation Board

EDPB – European Data Protection Body

EDPS – European Data Protection Supervisor

EHDS – European Health Data Space

EHR – Electronic Health Record

EP – European Parliament

EU – European Union

GDPR – General Data Protection Regulation

GPDP – Garante per la Protezione dei Dati Personali

ICO – Information Commissioners Officer

ISO – International Organisation for Standardisation

MS – Member State

NAB – National Accreditation Body

NLF – New Legislative Framework

PSB – Public Sector Body

WP29 – Article 29 Working Party

•