

Alma Mater Studiorum
Università di Bologna

DOTTORATO DI RICERCA IN
INFORMATICA GIURIDICA
E DIRITTO DELL'INFORMATICA

Ciclo XXII - IUS/20

Andrea Resmini

Information Architecture Modeling for Historical and Juridical Manuscript Collections

Relatore

Prof. Monica Palmirani

Coordinatore Dottorato

Prof. Carla Faralli

Coordinatore di Indirizzo

Prof. Giovanni Sartor

Esame Finale Anno 2010

Table of Contents

Introduction.....	7
Research Question.....	9
Structure.....	9
Definitions.....	10
Information.....	11
Information Architecture.....	11
Innerio Archives R1 and R2.....	11
Information Architecture.....	13
From Human-Computer Interaction to Human-Information Interaction.....	13
A Brief History of Information Architecture.....	17
From Web-design to Design.....	23
Approaches to Information Architecture.....	23
Information Design.....	25
Information Science.....	26
Information as a Resource.....	26
Architecture of Information.....	27
A Note on Big IA vs Little IA.....	29
A Definition of Information Architecture.....	31
Interaction Design, User Experience, and IA.....	36
User Experience Design.....	41
Interaction design.....	41

Understanding Interaction Design.....	42
Usability Engineering.....	43
Evaluating Information Architecture.....	44
Types of Evaluation.....	46
Evaluation Process.....	47
Heuristics in the Evaluation Process.....	48
A Final Consideration.....	49
Digital Libraries and Digital Archives.....	51
What is a Digital Library.....	51
Information Architecture and Digital Libraries.....	61
Usability and User Experience for Digital Libraries.....	64
IA for Historical and Juridical Digital Archives.....	69
Case-studies.....	70
A Methodology for IA Evaluation of Digital Libraries.....	70
Early Manuscripts at Oxford University.....	74
Analysis.....	76
Laurentius.....	78
Analysis.....	79
The British Library Catalogue of Illuminated Manuscripts	82
Analysis.....	83
Manuscriptorium.....	87
Analysis.....	88
An Issue of Focus and Purpose.....	92

The Innerio Archives.....	95
An Overview.....	95
Accessing the Archives.....	98
Specific Information Architecture Issues.....	98
R1: Initial Vision.....	99
General Issues.....	100
Content Inventory.....	103
Homepage.....	103
About (Presentazione).....	103
Authors (Lista Autori).....	105
Author (Autore).....	107
Codices (Lista codici).....	108
Codex (Codice).....	109
Search (Ricerca).....	110
Register (Come abbonarsi).....	111
Login (Area abbonati).....	112
My Page (Pagina utente).....	112
Help (Guida al servizio).....	113
Terms of Use (Note legali – Informativa privacy).....	114
Partners (Enti partecipanti).....	114
Comparing Innerio with the Case Studies.....	115
Making the Transition.....	117
R2: Second Iteration.....	121

Authors.....	124
Codices.....	126
The Annotation Engine.....	131
Searching the Innerio Archives.....	136
Synopsis of List Pages from R1.....	139
Modeling Complex Historical and Juridical Conversations.....	141
Current Challenges.....	141
The Authenticum Manuscript.....	146
The Montecassino 266 Manuscript	147
Early Findings and Four Basic Goals.....	148
Dynamism.....	151
A Brief Introduction to Standards.....	151
TEI.....	153
TEI / MASTER.....	154
TEI P5 ENRICH.....	156
FRBR.....	158
Complexity.....	162
Correlation.....	165
Relationships in FRBR.....	166
Persistence.....	167
Identifiers as a Social Construct.....	168
Persistence in the Scientific Internet.....	170
URLs, URIs, and URNs.....	172

Table of Contents

PURLs.....	176
Practicalities of Persistence.....	177
Open Issues.....	179
Shortcomings of FRBR.....	180
Conceptual issues with FRBR and URIs.....	181
DOIs and PURLs.....	185
A Few Final Notes on the DOI.....	186
Developing TEI Further.....	187
Integrating Approaches.....	188
Conclusions.....	191
Final Considerations.....	197
Bibliography.....	199

Foreword

Introduction

In recent years, quite some attention has been devoted to the issues brought along by the maturing of the Web as a platform and by the globalization of information: what has been brilliantly dubbed “global information architecture” (Van Dijck, 2007) has been openly explored and discussed. “Think global, act local”, Google’s motto, has become a well-known and well-accepted meme.

We know of the issues involved with complex interchanges among cultures, we know the Maori do not fit that well with the Dewey system we use in libraries all over the world (Van Dijck, 2008) and that Ranganathan’s faceted classification (Ranganathan, 1964) purports a very different stance on mankind and not simply a different way to organize books. We even have come to terms with the fact that women, fire and dangerous things might have something in common in certain circumstances or certain cultures (Lakoff, 1990), but are not we just leaving something out? Should not this “think global” approach reach not only through space, around the globe and among contemporary cultures, but also through time, across the centuries and into our very cultural heritage?

Current numbers are sketchy at the best, but an estimated 30 billion pages were supposed to be available on the World Wide Web at the end of 2007. Most of this is content that normally does not go back more than two-three years, and thought to be ancient when it is from 1996. That is but twelve years: What about the rest? What about the primary sources mankind produced before that, and before the Web?

That content is there, in parts. Not everything of course, but a fair deal. But this is by far and large the world of digital libraries and digital archives, large and small specialized web sites or networked applications from which we can access electronic renditions of manuscripts, incunables, old books, early movies, and daguerreotypes. Digital libraries have been around for quite a few years: still, they are somewhat not part of the mainstream

landscape of the World Wide Web, maybe because they are still perceived as academic self-sustaining efforts (Levy, 2000) with little business or entertainment value attached, or maybe because they .

Either way, through the past fifteen years digital libraries have matured mostly inside the womb of library and information science, with a little contribution, and most of this eminently technical, coming from the field of computer science. Whatever the reasons behind this, digital libraries have basically perpetuated on the World Wide Web the ordered, static model of the traditional bibliographical record.

Unfortunately, this is not always the case. The Web is a different medium: hyper-linking and the absence of shelves and walls (Shirky, 2005) allow unprecedented freedom. Moreover, because of its very nature, not everything inside a digital library can be described as a book: more than often it can not. Physical properties, language, concepts such as authorship, can come to mean different things. Sometimes, as this research will detail, a digital library holds conversations with binding historical and juridical constraints, and library science is definitely not enough. Competencies have to be drawn in from other disciplines and fields.

This research presents the results of the experimental work conducted by the author¹ on the development of the Iernerio Archives² for CIRSIFID, Centro Interdipartimentale di Ricerca in Storia del Diritto, Filosofia e Sociologia del Diritto e Informatica Giuridica, Alma Mater Studiorum University of Bologna, between 2007 and 2009, under the supervision of Monica Palmirani, and the consequent reflections on ways to improve the modeling of complex historical and juridical medieval manuscripts on-line.

As such, this research work draws from three different fields: information architecture, digital libraries or digital archives, and legal informatics, and pays large tribute to the

1 With Régis Riveret and Luca Cervone.

2 The Iernerio Archives are an on-line digital archive containing the digitized versions of 288 medieval codices, amounting to more than 135000 single pages, from the vast collection of legal, theological and philosophical texts in the rare-books collection of the Royal College of Spain in Bologna, Italy. The Archives are named after the Italian jurist Inerius, the first illuminator of legal science in Bologna.

research conducted by Monica Palmirani and her group on the rendering of valid, enforceable, complex legal information in Web environments.

Research Question

How to preserve and render the chain of causality and the complexity of manuscript documents of juridical and historical value on-line by modeling a coherent information architecture and allowing for increased user interaction and a successful user experience?

Structure

This dissertation is composed of this brief introduction and of five distinct sections or chapters:

1. Information Architecture
2. Digital Libraries and Digital Archives
3. The Innerio Archives
4. Modeling Complex Historical and Juridical Conversations
5. Conclusions

Chapter 1 deals with the history and ground definitions of information architecture and why this research considers information architecture a relevant piece of the overall strategy in moving forward from the current static digital library paradigm to a new, dynamic model, which is necessary to capture the complexity of legal reasoning and research in the field of digital documents of historical value. It also introduces the heuristic information architecture evaluation model that was used during the development of the second release of the Innerio Archives R2 and that was used during competitive analysis of other digital libraries.

Chapter 2 deals with digital libraries and digital archives, providing an overview of the field and introducing some of the issues and challenges for digital libraries access and consumption. It details what a digital library is thought to be currently, and how digital libraries related to fields such as usability, user experience, and information architecture. It finally introduces five case-studies, in which digital libraries are assessed based on the heuristic evaluation methodology outlined in Chapter 1.

Chapter 3 describes the initial (R1) and current (R2) versions of the Innerio on-line archives, a collection of digitized medieval codices from the Royal College of Spain in Bologna, Italy, framing them around the methodological and conceptual aspects of information architecture and in the ongoing digital libraries and digital archives debate. The section provides a detailed description of the most relevant technological and user-experience-related aspects of the two installations, together with an account of the rationale behind the introduction of an Annotation Engine for scholars in R2 and its importance in laying the basis for a new iteration.

Chapter 4 introduces the concept of metadata as artifact, and how, for historical and juridical on-line archives, a step forward from the information architecture model implemented in R2 is necessary as we move past the traditional static rendition of a bibliographical record to the actual preservation of the historical and juridical conversation happening through time and space of which the collection and its items are part. This section is heavily based on the work and findings of Monica Palmirani on FRBR in the context of juridical digital documents.

Chapter 5 recomposes the picture and draws the conclusions.

Definitions

Definitions will be introduced in the main body of text when necessary, but given the broadness of meaning that has come to be attached to the one word through the years (information), and the fuzziness and ongoing debate that largely still seem to surround thje

other one (information architecture), I offer brief definitions for these two central terms here.

Information

For the scope of this research, information will be used as a generic term for contextualized, ordered, transmitted data, primarily following Lowe and Hall (1999), who concisely define it as “*(t)he interpretation of data within a context set by a priori knowledge and the current environment*”. This is not dissimilar from how information is framed in information systems research and Langefors (1995), who assumes that “*information is a product of an interpretation process operating on data but also on the pre-knowledge of the interpreter*”.

Information Architecture

In this research, the phrasing information architecture both refers to the field and to the actual implementation of an information architecture strategy and outcome in a given project. In this sense, I will explicitly talk of the information architecture for the Innerio archives. As information architecture is a multi-disciplinary field which touches and draws upon a number of bordering disciplines such as architecture and industrial design, media and communications science, information systems, human-computer interaction, computer science, psychology, library science, borrowed concepts will be now and then introduced and discussed. These will be defined when necessary and framed in the respective theoretical backgrounds.

Innerio Archives R1 and R2

The Innerio Archives went through two major cycles of development: the first one started in 2000 and ended in 2002, the second one started in 2007 and ended between 2008 and 2009. For the sake of simplicity I will refer throughout this research to that first version as release 1 or in short R1, and to the second version as release 2 or R2.

Chapter I

Information Architecture

From Human-Computer Interaction to Human-Information Interaction

“The design of good houses requires an understanding of both the construction materials and the behavior of real humans.”

– P. Morville

In the history of design, the two separated concepts of information retrieval (IR) and user interface are the product of a time when computers were limited, slow, and barely able to interact with humans. It simply made sense to consider operating them and retrieving information from them as two different, separated activities. Modern-time IR, defined as the science of searching for documents or data, for metadata about documents, or for information within documents, has its roots in the 1950s and in Luhn’s work with word indexes and word overlap (Singhal, 2001); industrial design (ID) has been working with what can be considered the modern concept of an interface since the late 1880s, although admittedly mainly in terms of body-tool and product-production relationships (Gregotti, 1986). Both were significant when all interaction was in absolute propositions and conveyed through a stand-alone computer screen: the rigorous use of analysis to

“break down the whole problem into components and first focus on the components that promise to yield”

which is typical of IR (Marchionini, 2004) shows a a fundamental concern with information objects and a certain distance from the people who work with those objects. Intentional distance, it seems, as

“people are less predictable and more difficult and expensive to manipulate experimentally, IR research logically focused on the information objects first.”

Conversely, in the field of design a significant shift towards the idea of interfaces as artifacts was to come about only in the 1980s (Bonsiepe, 1995), when human-computer interaction (HCI) was already an acknowledged if specialty area in computer science (Carroll, 2009). Bonsiepe, once a teacher at the post-Bauhaus Hochschule für Gestaltung in Ulm, Germany, was among the first to foresee that

“si pone la questione di come possano essere portati a unità tre ambiti così eterogenei: il corpo umano, lo scopo di un’azione, un artefatto o un’informazione nell’ambito dell’agire comunicativo. Il collegamento attraverso questi tre campi avviene tramite un’interfaccia. È necessario pensare che l’interfaccia non è un oggetto, ma uno spazio in cui si articola l’interazione tra corpo umano, utensile (artefatto, inteso sia come artefatto oggettuale sia come artefatto comunicativo) e scopo dell’azione.”¹

Even so, and even when very lucidly forecasting the possibility of issues of information overload (Bonsiepe, 1995, 67), Bonsiepe was thinking inside the box of industrial design. His views of an impending rise of what he called *infodesign* was somehow limited to the then relatively new field of user interfaces for software programs and to traditional broadcast media like television and movies. In a way, he viewed upcoming changes inside the frame of a static, desktop computer-driven future.

This is not the case anymore: fast, connected micro-computers are everywhere. They are inside mobile phones, automatic teller machines, low-budget cameras, home appliances, ticket machines. Interaction is not limited to precise queries run through specialized interfaces in a controlled environment or to dealing with a few well-known software programs for office automation: interactions happen everywhere (Greenfield, 2006).

1 “Here we have an issue: how can we bring together such heterogeneous parties as the human body, the goal of a given action, an artifact or a piece of information in the context of communication. What binds these three fields is the interface. An interface is not an item, but a space in which the interaction between the human body, the tool (the artifact, regardless of it being a factual object or a communication object), and the goal is expressed.” Translated from the Italian by the author.

Not surprisingly then, in the scientific area which can be broadly if albeit fuzzily described as the larger field of human-computer interaction (HCI), the focus has been shifting constantly from retrieval and interfaces to loose interactions and user experience since the late 1990s (Cooper, 1999). As a consequence, a number of new, or reformulated, practices and disciplines have emerged which focus more on user experience and social communication (Resmini & Rosati, 2008) and which are now coalescing around actual bodies of knowledge and grounding conceptual frameworks (Madsen, 2009). Information architecture is one of them.

It is interesting to note how some authors (Streitz, 2005) interpret this movement towards an holistic view of what Bonsiepe early identified as the user-goal-tool relationship as a pendulum swinging back. Streitz observes how initially the introduction of computing technology and computers caused a move away from real objects (like books or persons) as sources of information, in favor of desktop computers which have become the interfaces of choice to gain access to information. Streitz argues for

“returning to the real world as the starting point for designing information and communication environments. Our approach is to design environments that exploit the affordances of real world objects and at the same time use the potential of computer-based support”

in a shift from traditional human-computer interaction to what he calls human-artifact interaction (Streitz, 2005).

I argued (Resmini, 2009) instead that the real push is towards human-information interaction with an increased “attention to human participation in the (IR) process” (Marchionini, 2004)². In a sort of reverse McLuhan meme, I maintain that in a scenario

2 “What is more significant (...) is the trend toward all objects becoming more dynamic and less static and dependable for IR purposes. For example, an active blog is an object that is continually changing and its representations must likewise be continually updated as well. This change emanates from new capabilities within objects and from new capabilities in the external environment that contains them. Internally, electronic objects increasingly are designed to exhibit behavior—to ‘act’ according to external conditions. Hypertext is of course the classic example; recommender systems illustrate more contemporary examples; and context-aware sensor-program devices illustrate the latest trend. In addition to the increasingly sophisticated behavior inherent to information objects is the trend toward the global information space (cyber-infrastructure) to store and use context. For example, a retrieval system may

where fast replication and forwarding of pieces of information are the rule, from mobile to desktop, from central to personal, from Twitter³ to Facebook⁴ to e-mail, and where mash-ups and remediation are common, messages have no medium and the message is once again just the message.

In the design of these new bridge-experiences⁵ brought on by the reification of the message interface design is not enough (Grossman, 2006): human-information interaction issues have to be addressed throughout the whole process, incorporating the apparent quirks of human behavior into information seeking strategies (Kaltenbacher, 2009). This marks a shift from the traditional current view on multi-channel strategies, where more than one channel is used simultaneously and alternatively (Rosati, 2007), as the focus gets moved from optimized, data-driven information flows to stable human-information interaction models.

This is also why I believe that IA has to move to a process-oriented approach (Resmini & Rosati, 2009b). The reason for this is that while interfaces are bound to a specific single artifact, interaction models can span a whole process involving a varying number of artifacts. A very good example of this is the way Apple designed a common experience throughout the iPod / iTunes line of products and is now actively bringing it to their stores (Potente & Salvini, 2009). Since the process itself is a continuum, designing the interaction model globally allows to build global, constant interaction patterns: users are not forced to learn or re-learn diverging behaviors even if individual interfaces differ.

not only represent web-page content but continuously update access times and referral pages. Additionally, the system may save increasingly detailed traces of fleeting ephemeral states arising in online transactions—perhaps as extreme as client-side mouse movements as well as clicks. Thus, our objects acquire histories, annotations, and linkages that may strongly influence retrieval and use.” (Marchionini, 2004, 3).

- 3 Twitter (www.twitter.com) is a free social networking service that enables its users to send and read from a number of different platforms text-based messages of up to 140 characters called tweets.
- 4 Facebook is a social networking service which enables people to share text and multimedia content with friends. As of Dec 4 2009, Facebook has more than 350 million users worldwide (Source: Facebook statistics, <http://www.facebook.com/press/info.php?statistics>).
- 5 First defined by Joel Grossman in an article for UX Matters, a bridge-experience is “one in which the user experience spans multiple communications channels, document genres, or media formats for a specific, tactical purpose”. See References.

A Brief History of Information Architecture

“The metaphors we use constantly in our everyday language profoundly influence what we do because they shape our understanding. They help us describe and explore new ideas in terms and concepts found in more familiar domains.”⁶

Information architecture (IA) is an emerging discipline and community of practice focused on bringing principles of, among others, library science, design and architecture to information space (Rosenfeld & Morville, 2002; Resmini & Rosati, 2009b).

As many other contemporary technology-concerned and technology-aware fields such as interaction design, usability engineering, or user experience design, IA is currently eminently practice-led (Fenn, Hobbs & Resmini, 2010) and can be certainly characterized as being a production activity (Francke, 2008). For very similar reasons Haverty (2002) characterizes information architecture as an inductive process: as a new field, it lacks its own body of theory, and cannot be yet considered a discipline. Haverty provides a thoroughly description of her view:

“A formal discipline (e.g., biology) is a branch of knowledge and teaching grounded in a deep theoretic basis; theory guides the activities of workers in the community. A field is an area of academic interest or specialization that draws on theory, but usually from other related disciplines. Lack of internal theory leaves workers in a field to start from scratch with each design problem. As a field matures and workers gain experience, useful solutions may be reused, but confidence in the ability of that solution to fit the given context and fulfill the problem is still tentative and perhaps even unreliable without formal validation.”

6 Morrogh, E. (2003).

In recent years, though, a number of higher education and research initiatives⁷ are coalescing into solid foundations for wide-spread academic interest and acceptance (Madsen, 2009).

Even though its modern use, related to the design of information, goes back no further than the mid 1970s and Richard Saul Wurman's famous address at the American Institute of Architecture conference of 1976⁸, the compound term *information architecture* has been around for a long time: after all, in the English language the use of the term architect to address a creative person or role is rather common (Ronda Leon, 2008), and combining it with information in an age of computers must have seemed rather straightforward.

Conceptually, a turning point can be traced back to an IBM research paper written some twelve years before Wurman's Information Architecture conference, in 1964, and titled "Architecture of the IBM System/360"⁹, which defines architecture as

the conceptual structure and functional behavior, distinguishing the organization of data flows and controls, logical design, and physical implementation.

7 These include higher education programmes at the bachelor, master, and PhD level in both Europe and the United States, and scientific peer-reviewed journals like the Journal of Information Architecture (<http://journalofia.org/>).

8 Wurman wrote also an article with Joel Katz titled "Beyond Graphics: The Architecture of Information", which was published by the AIA Journal in 1975. In an interview with Dirk Knemeyer in 2004, Wurman says about the rationale that brought him to choose the term: "The common term then was 'information design'. What got confusing was information design and interior design and industrial design, at that moment and still today in many and most people's minds, are about making something look good. Interior designers make your place look better, industrial designers were engineers doing something that usually went to an engineer to put a package around it. Information design was epitomized by which map looked the best - not which took care of a lot of parallel systemic parts. That is what I thought 'architecture' did and was a clearer word that had to do with systems that worked and performed. Thought 'architecture' was a better way of describing what I thought was the direction that more people should look into for information, and I thought the explosion of data needed an architecture, needed a series of systems, needed systemic design, a series of performance criteria to measure it. There are thousands of people using the term ('information architecture'), and they have no idea where the term came from, and 90 percent of them aren't doing what I think they should be doing anyway." Knemeyer, D. (2004). Richard Saul Wurman: The InfoDesign interview. Available online at http://www.informationdesign.org/special/wurman_interview.htm. Accessed January 6 2010.

9 Amdahl, G. M., Blaauw, G. A., & Brooks Jr., F. P. (1964). Architecture of the IBM System/360. IBM Journal for Research and Development. April 1964. Cited in Ronda Leon (2008) and available at <http://www.research.ibm.com/journal/rd/082/ibmrd0802C.pdf>.

Although undoubtedly referring to computer architectures, this paper marks an important landmark in the way it conceptualized the term in connection with structure and behavior and not just physical layouts, abstracting it and laying the basis for the subsequent extension of its use to other areas of computing (Ronda Leon, 2008).

A few years later, in 1970, at the Xerox Palo Alto Research Center (PARC) a group of people specialized in information sciences was assembled and then given the charter to develop technology which could support the “architecture of information” (Pake, 1985). As Hearst (1996) recalls,

“Many important contributions resulted from this call, including the first personal computer with a user-friendly interface and bitmapped display, the first WYSIWYG text editor, laser printing, and the ethernet network that could flexibly connect workstations, file servers, and printers in order to provide communication among many workers. Perhaps because of the social nature of information creation and use, much of the technical research at PARC has emphasized the human-computer interaction and social aspects of computing.”

Ronda Leon (2008) correctly evidences how this first documentary evidence of the use of the term information architecture joins specialists in information science and development focused on the users, a trait that will be somehow brought on to the real wave of IA in the 1990s through a seemingly dormant stretch of years all through the 1980s, when the idea of information architecture as the design of dynamically changing information seems to be lost to a view much more akin to that of information systems. Articles written in those years refer to IA as a tool for the design and creation of information systems tout court, with an emphasis on the organizational and business aspects of the information networks¹⁰ (Morrogh, 2003).

¹⁰ This angle has not entirely disappeared, but it is now largely minoritarian. Nonetheless, Carter still maintained in 1999 that “(i)nformation architecture as an holistic way of planning which meets the organisation’s information needs and avoids duplication, dispersion, and consolidation issues. The information architecture is the collective term used to describe the various components of the overall information infrastructure which take the business model and the component business processes and deliver information systems that support and deliver it. Prime components are the data architecture, the systems architecture and the computer architecture.” (Carter, 1999).

Weitzman (1995), in a research for the Massachusetts Institute of Technology entitled “The architecture of information: interpretation and presentation of information in dynamic environments”, supports this notion that the term originally came from Xerox Labs¹¹. Quoting Smith and Alexander, he reports that

“Xerox was among the first corporations to address this notion of information structure and use the ‘elegant and inspiring phraseology of information architecture’ to define its new corporate mission”

Weitzman continues:

“The basic purpose of the Xerox Corporation is the best way to bring order and discipline information. Thus, our main effort, our common denominator, has evolved towards establishing leadership in what we call architecture of information. What we seek is to think of it as a natural and undeveloped, which is made more livable for people who live and work within it”¹²

This high-level framing, the necessity for a broader vision, remained one of the core concepts for those many writing and discussing the development of IA up to the mid 1980s. Brancheau and Wetherber (1986) maintain that the IA process

“starts from a high-level conceptual view, then is successively refined to the lowest level at which the physical database can be implemented”

So it is a design process moving from the general to the specific, and IA is part of that stage which is usually called requirements analysis: it is not a deductive, mathematical procedure or process. Furthermore, Brancheau and Wetherber also maintain that important steps to a successful IA are blueprints, requirements, information categories, and guidelines on the underlying business processes. Global corporate needs are also part of

11 Besides providing further documental evidence to support this notion, Weitzman also underlines how Xerox actually contributed vastly to the general view of information architecture as a tool to support the design and presentation of documents, something that is of vital importance in Wurman’s work.

12 Xerox: Searching for an Architecture of Information. Address by Peter McColough, President, Xerox Corporation, before the New York Society of Security Analysis on March 3, 1970. Cited in Weitzmann (1995).

the picture. These are all themes that the modern wave of information architects, the one spearheaded by the Rosenfelds and Morvilles and Wodtkes, will ultimately incorporate in their view of IA in the late 1990s.

Ronda Leon maintains that these two early phases, that of information design (1960s-1970s), and that of system (and systemic) design (1980s), were integrated into the modern idea of IA we know today during the 1990s. He provides a graphical chronology of IA events, mainly books, papers, and conferences, which is definitely interesting.

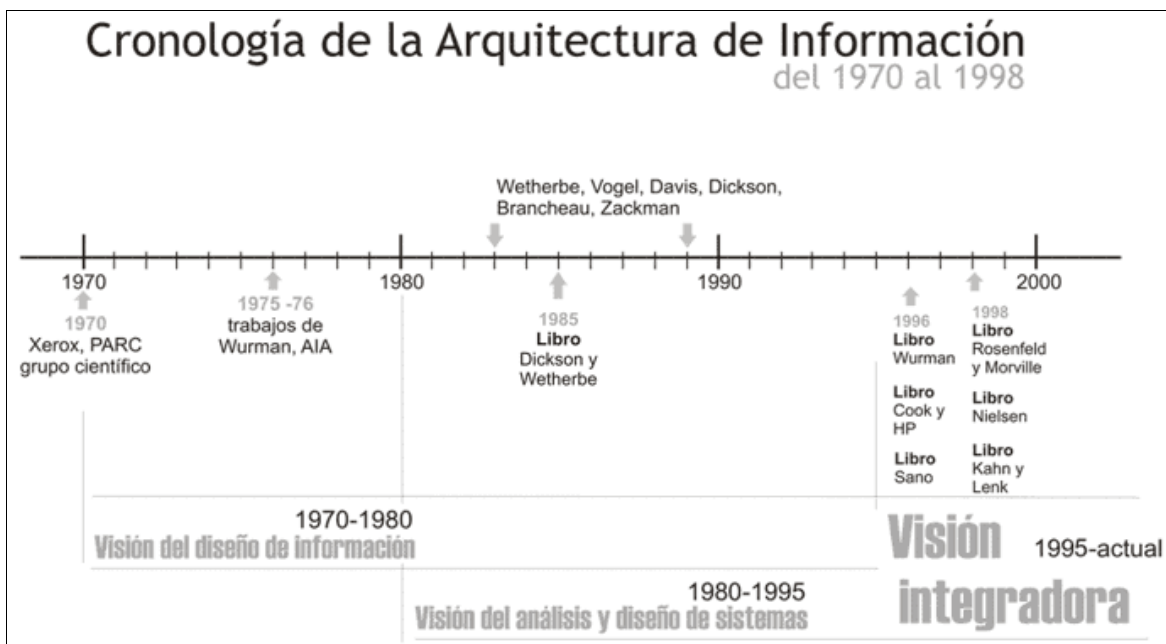


Figure 1: Ronda Leon, R. A chronology of demonstrating the development of the profession of information architect in the 80s and early 90s. References obtained from LISA (1998).

In any case, it is adamant that the early start of the IBM paper and Wurman's initial vision had to wait until a piece that was missing entered the scene. That piece was the possibility for professionals to operate on large amounts of data in a new media, void of preexisting corporate hierarchies. That is why world-wide acknowledgment that something called information architecture was there to redefine first web-design and then the way we design human-information interaction came in 1998 when Louis Rosenfeld and Peter Morville

wrote “Information Architecture for the World Wide Web”¹³, one of the modern cornerstone books in the field.

So-called dot-coms were booming, and at the time their web-design firm, then Argus Associates, originally owned by Rosenfeld and Joseph Janes with Morville an employee¹⁴, was already a significant player on the American market. Interestingly enough, it seems like they came to use the label information architecture without much back-thinking, from a totally naïve get-the-job-done perspective. In the words of Morville they

“(...) found (them)selves using the architecture metaphor with clients to highlight the importance of structure and organization in website design. Lou got a gig writing the Web Architect¹⁵ column for Web Review magazine, and I soon joined in. In 1996, a book titled Information Architects appeared in our offices. We learned that a fellow by the name of Richard Saul Wurman had coined the expression ‘information architect’ in 1975. After reading his book, I remember thinking ‘this is not information architecture, this is information design’.”¹⁶

Rosenfeld and Morville’s initial view was focused on organization, labeling, navigation, and search. In an interview with Scott Hill for O’Reilly in 2000, Rosenfeld explicitly says that these are key points to

“help people find and manage information more successfully. Organization systems are the ways content can be grouped. Labeling systems are essentially what you call those content groups. Navigation systems, like navigation bars and site maps, help you move around and browse through the content. Searching systems help you formulate queries that can be matched with relevant documents.”¹⁷

13 Cited in this research in its 2nd edition, 1998, and now in its 3rd edition, 2006.

14 Morville, P. (2004) A Brief History of Information Architecture. P. xiii. In Gilchrist, A., & Mahon, B. (2004).

15 The "Web Architect" column in the *Web Review* magazine, mostly written by Rosenfeld and Morville, ran from 1995 to 1999 and featured articles exploring information architecture topics and geared towards the practice of IA.

16 Morville, P. (2004) A Brief History of Information Architecture. P. xiii. In Gilchrist, A., & Mahon, B. (2004).

Very famously they remarked that the real difference they could see between their view and Wurman's, *post hoc*, was that for them information architecture was very much the design of what was between the pages of a web site, meaning the links, the structure, the connections, as opposed to the design of the pages themselves. It must also be noted that their book was titled "Information Architecture *for the World Wide Web*" (italics mine), definitely suggesting an aim directed towards a well-defined area, the Web, inside the larger field of information architecture.

From Web-design to Design

Mostly because of the enormous success Rosenfeld and Morville's book, the so-called Polar Bear book because of the Polar Bear on its cover, had, in the late 1990s and early 2000s the practice of IA was usually considered to be tightly integrated to experiencing the World Wide Web (Weinberger, 2002; Gilchrist & Mahon, 2004). That started to change with the rise of ubiquitous computing and the idea of an Internet of things (Greenfield, 2006). Even though a persistent thread keeps anchoring not only IA but most of these digital design disciplines to the creation of Web-only content (Francke, 2008), IA is now seen as a boundary field whose contributions are crucial where complexity, unfamiliarity and information overload stand in the way of the user, regardless of the very nature of the environment being designed, whether be it digital, physical, or a combination of the two (Resmini & Rosati, 2009a). It can be safely said that contemporary culture shows similar convergence patterns (Jenkins, 2006), and this evolution is but a strategic part of the general shift from the design of objects to the design of processes (Sterling, 2005).

Approaches to Information Architecture

"That's why I've chosen to call myself an Information Architect. I don't mean a bricks and mortar architect. I mean architect as used in the words architect of foreign policy. I mean architect as in the creating of systemic,

17 Hill, S. (2000). An Interview with Louis Rosenfeld and Peter Morville.
http://www.oreillynet.com/pub/a/oreilly/web/news/infoarch_0100.html. Accessed Oct 16 2009.

structural, and orderly principles to make something work--the thoughtful making of either artifact, or idea, or policy that informs because it is clear. I use the word information in its truest sense. Most of the word information contains the word inform, so I call things information only if they inform me, not if they are just collections of data, of stuff.¹⁸

As I said before, information architecture has its roots in a rather large number of different disciplines. As opposed to Environmental Psychology, for example, which was born out of the necessity to specialize inside the already established field of psychology (Baroni, 1998), information architecture emerged as a way to tackle issues which seemed to be totally new, unforeseen, and requiring pioneer thinking (Rosenfeld, 2009).

“Pioneers are required to open up any new domain. But then they quickly lose relevance and get pushed aside by homesteaders. Us pioneer types were never ones to fit in anyway, so we look for new frontiers”

As of today, there has been no comprehensive effort to push forward an effective critical analysis of the field. As this research is not directly focused in a critical assessment of the cultural or methodological history of the field, but rather in highlighting the diverse mindsets, methodologies, and outcomes in the light of organizing better historical and juridical on-line collections, following White (2004) we could say simply that IA can be traced back to three broad, different approaches, characterized by the way they work with information:

- ◆ the information design approach
- ◆ the information science approach
- ◆ the information as a resource approach

and add to these one fourth final approach, that of the “architecture of information”, a stance which is becoming more and more important as on-line interactions become richer and more complex.

18 Wurman, R. S. Information Architects. 1996

Information Design

This is where Wurman's contribution and initial vision stems from. For Wurman, design and architecture are the basis for a science and art of creating "*instruction(s) for organized space*" (Wurman, 1997). Wurman recognizes gathering, organizing, and presenting information as tasks analogous to those an architect faces in designing a building. If the architect must:

"ascertain those needs (i.e., must gather information about the needs); organize the needs into a coherent pattern that clarifies their nature and interactions, and; design a building that will - by means of its rooms, fixtures, machines, and layout, i.e., flow of people and materials - meet the occupants' needs"

then the information architect has a definitely similar goal in information space. Presenting information for a purpose is an architectural task (Wyllys, 2000). In 1976 Wurman chaired the national conference of the American Institute of Architects (AIA) and he chose "The Architecture of Information" as conference theme. According to what he said to Dan Klyn in a recent set of interviews¹⁹, he had no plan in mind, he was just trying to "find patterns for himself". Wurman developed this definition of information architect for the conference:

"1) the individual who organizes the patterns inherent in data, making the complex clear. 2) a person who creates the structure or map of information which allows others to find their personal paths to knowledge. 3) the emerging 21st century professional occupation addressing the needs of the age focused upon clarity, human understanding, and the science of the organization of information."

Wurman was never interested in disseminating his ideas to a new audience, nor in creating a new field or profession, and was actually quite surprised when he finally did find out what an echo his pattern-finding activities had managed to stir up.

¹⁹ Klyn, Dan. (2009). Interview with Richard Saul Wurman. Audio available at <http://danklyn.com/blog/?p=397>. Accessed Dec 2 2009. Klyn: "Did you intend to create a movement within the field of architecture to focus on information display and organization and such things?" Wurman: "No".

Information Science

This is where Rosenfeld and Morville initially came from. In the above mentioned interview with Scott Hill for O'Reilly²⁰, they stated:

“In 1994, before the Web took the world by storm, we were teaching some of the first academic and commercial courses about the Internet. We both believed the Internet would become an important medium and that librarians had a great deal to offer this brave new world of networked information environments.”

Rosenfeld and Morville, and those following along their initial LIS view, must be credited for bringing in many of the core methodologies used for navigation, labeling, and site-structure. It must be also said that they offered the blooming community of practice an extremely empirical and practical approach, and they single-handedly brought user-research and usability engineering into the core of mainstream IA tools. As Rosenfeld is fond of saying, they “certainly embraced other disciplines”. So far, Rosenfeld and Morville’s view of IA represents the mainstream and most accredited view, especially for those outside the field.

Information as a Resource

A third approach to information architecture is taken by Roger and Elaine Evernden in their book “Information First” (Evernden & Evernden, 2003), in which they define IA as

“a foundation discipline describing the theory, principles, guidelines, standards conventions and factors for managing information as a resource”

This view is somewhat more tightly connected to the logic of information systems than to the logic of user experience, and as of today it is a minority stance. It is rather plain to see that even those looking critically on the field see no particular interest in pursuing this

20 Hill, S. (2000). An Interview with Louis Rosenfeld and Peter Morville.
http://www.oreillynet.com/pub/a/oreilly/web/news/infoarch_0100.html. Accessed Oct 16 2009.

different point of view. While commenting positively on the views pushed forward by the Everndens, White personally maintains that

“I suspect, and indeed hope, that information architecture will be rather like information science, a set of tools and approaches that can be used by professionals from a wide range of backgrounds to solve an equally wide range of information management problems.”²¹

Architecture of Information

A number of leading practitioners have compared the practice of information architecture to what brick-and-mortar architects do, as both “design spaces for human beings to live, work, and play in” with the primary differences being the materials they work with (Wodtke, 2001). Information architects are then very much like architects in that they too are concerned with spatial relationships concerning navigation and boundaries and with “setting structure to an element to be built that combines components that are grouped together based on users’ understandings and expectations” (Vanderwal, 2001).

Much has been discussed in favor or against this approach, but three facts need to be considered: ubiquitous computing and ambient intelligence, with information moving into small appliances, mobile phones, and intelligent displays; social networks; and the increasing capabilities of network-enabled or web-based applications. All these relatively new phenomena have clearly moved the bar from what could be thought to be appropriate of simple hypertext systems in the late 1990s.

Morrogh (2003) has been the first to clearly articulate brick-and-mortar architecture principles and views as the center of a theoretical discussion on IA, and connecting them to concepts of information-seeking behavior such as “wayfinding” and “navigating”. Research in architecture and in cognitive psychology lends to the idea that there is an eminently demonstrable basis to spatial reasoning and spatial behaviors even when dealing with information space (Munro, Höök, & Benyon, 1999). For example, Johnson (1987)

²¹ White concluded his paper on a positive note, saying: “All we need now is a journal of information architecture!”.

argues that more abstract reasoning is always shaped by such underlying spatial patterns. He notes that the logic of containment is not just a matter of being simply in or out of the container, but it is a graded concept: if someone is in a deep depression, we know it is likely to be a long time before they are well. The deeper the trajectory travels in the container, the longer it will take for the trajector to get out of it²².

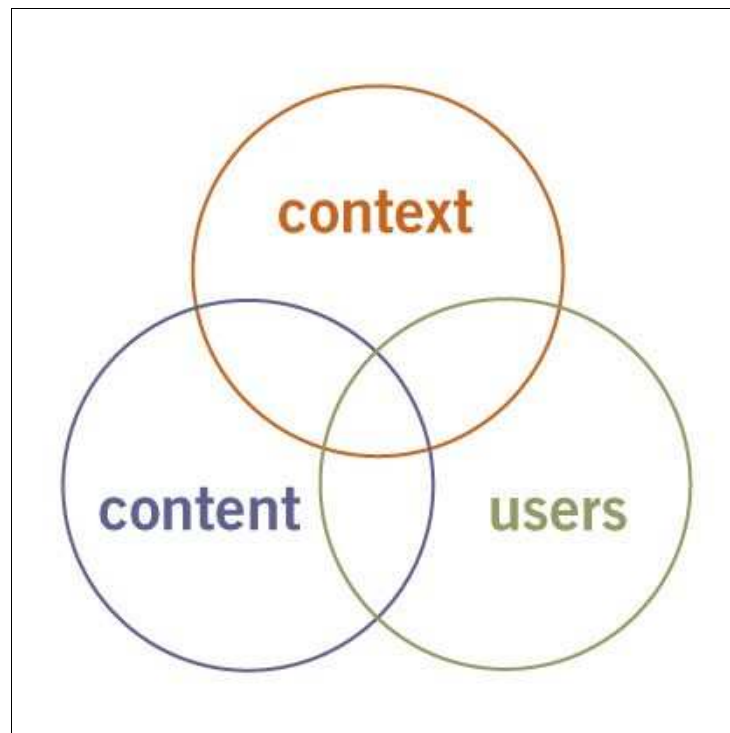


Figure 2: *The three circles*, from Morville, P. (2004). *User experience*.
<http://semanticstudios.com/publications/semantics/000029.php>

In this light, the single most important concept that the architectural approach brings to the discussion table is that of context, or, architecturally speaking, place²³. Context is one of the elements that Rosenfeld and Morville describe in their famous three-part Venn diagram. They make it an essential element in the building of a successful information architecture together with users and content, but in their representation context has very

22 Similarly, Johnson argues that transitivity and the law of the excluded middle in logic are underlaid by preconceptual embodied experiences of the containment schema.

23 For a discussion of space and place in connection with information architecture, see Resmini & Rosati, 2009b.

few of the spatial connotations that Morrogh for example attributes it: context is simply the business or “political” goals and constraints of a given project. As Morrogh explains, the underlying spatial metaphor²⁴ can be stretched further:

“It is easy to imagine occupants of an information space needing to have a sense of place in order to remain oriented; a sense of space so as to know where it is possible to go; and navigation devices commonly seen in physical environments such as maps, signs, paths, and landmarks for navigation. Information systems have even been referred to metaphorically as information cities, and, of course, we’ve all heard the infrastructure of the Internet referred to as the Information Superhighway. These are all spatial metaphors used to assist in the visualization of technologies and professions that are too new or complex for us to understand easily.”

In a similar vein, and with even larger cultural repercussions, the theoretical works of Alexander (1979) have been recently brought back to the attention of the IA community. Alexander presents an architecture-driven pattern model which can be used also for the design and building of usable virtual spaces constructed using an information architecture. Alexander’s ideas have a vast, if controversial, following.

Furthermore, the emergence of a disembodied self (Inalhan & Finch, 2004) living in a world where the boundaries are not physical anymore, but relational, is undermining our ability to form attachments with people, places, and companies. Nonetheless, “people goes where people is” (Inalhan & Finch, 2004), as the expansion of social networks clearly shows: one of the unintended effects of this is that it has strengthened the value of place and aroused a longing for community, which in turn make the case for an architectural approach stronger.

A Note on Big IA vs Little IA

On the side of the general discussion on identity, methodologies, and focus, a scope issue was almost immediately part of the core topics which catered for the attention of the IA

²⁴ Metaphor here is to be intended in the way Lakoff and Johnson describe it (1980): a fundamental mechanisms in our mental processes.

community in the early 2000s. The first complete description of this can be found in an article titled “Big Architect, Little Architect” Morville wrote for the column “Strange Connections” and published on the Argus web site (2000). In this short piece, Morville stated that while there is certainly a core practice of IA that involves structuring and organizing information systems for intuitive access to content, the

“(...) interpretations of the role of the information architect vary depending upon the organizations, the projects, and the people involved. At one end of the spectrum, the little information architect may focus solely on bottom-up tasks such as the definition of metadata fields and controlled vocabularies. At the other end, the Big Information Architect may play the role of ‘an orchestra conductor or film director, conceiving a vision and moving the team forward’, as described by Gayle Curtis, Creative Director at vivid studio”

This vision was to prove one of the most controversial and also of consequences. Jesse James Garrett (2002) connects these two opposing views to a choice of subject: Big IA, that Garrett sees as “a definition encompassing a broad range of responsibilities, including business strategy, information design, user research, interaction design, requirements gathering”, is based on the role, and as such tends to grow towards broadness; Little IA is based “on the discipline”, and tends towards narrowness: “(i)n order to speak meaningfully about information architecture problems and their solutions, we must define the scope of those problems in very concrete ways”. Garrett sees much of the confusion surrounding information architecture as something arising from the incorrect application of one to the other, in two opposite detrimental directions:

“when this definition (intended for the discipline) is applied to the role, it creates for some the fear of being ‘boxed in’, trapped in a role so narrowly defined that many of the elements essential to the success of any given architecture are outside the control or influence of the architect. The expansion of the role of the information architect may well be working to the benefit of the individual filling that role (...) but it is almost certainly working to the detriment of the discipline as a whole. Citing the holistic nature of information architecture work, some people clearly won’t be satisfied until they have direct control over every aspect of the business that

might affect the architecture. This way of thinking (...) undermines every effort to convince businesses of the value of the discipline.”

Although largely aimed at the professional community (*convince businesses of the value*), and clearly not interested in the academic distinctions between field and discipline, much of what Garrett says is an accurate reflection of the issues connected to a maturation of IA into a full-fledged discipline. This will be further expanded in the paragraph “Interaction Design, User Experience, and IA”.

A Definition of Information Architecture

“The common thought is that architects build buildings. No, architects make instructions for having someone else build them. So basically architects, if you’re talking about architects... they give instructions.”²⁵

As with many other maturing fields, the larger IA community seems actually to spend a lot of time trying to frame a perfect, static definition of the discipline itself to the point of having a terribly successful acronym to describe the process, DTDI, for “defining the damn thing” (Resmini, Madsen, & Byström, 2009). This is not uncommon at all in the social sciences (Baroni, 1998) and, if the pulse of the specific IA community has anything to do with the results, absolutely hopeless²⁶. The debate on what IA is and how should it be properly defined is now almost twenty years old, and beginning to rival the one still enveloping “information science” and dating back to the mid-1950s (White, 2004).

Even though a definition is somewhat inescapable if anything has to be communicated, opinions on how to reach a common understanding and what this understanding should be have been maddeningly oscillating through the years. In “IA Growing Roots” (Resmini, Madsen, & Byström, 1999), I wrote that

25 Klyn, Dan. (2009). Interview with Richard Saul Wurman. Audio available at <http://danklyn.com/blog/?p=443>. Accessed Dec 3 2009.

26 In a private conversation via e-mail with the author in the Autumn of 2009, Andrew Hinton, one of the founders of the Information Architecture Institute and then one of the Directors, pointed out that the “only communities which do not quarrel over such things are dead communities”.

“(w)ords represent our view of the world, but even without taking a full leap into cognitive psychology and linguistics, there are simpler, common sense considerations to be taken into account. First, this “define craze” that regularly seems to seize IAs is somewhat a sign of the times and actually fairly common, and it’s a consequence of two different conditions, one internal and the other external: (first), The community is young and somewhat necessarily shallow, and (second) we live in very fast times.”

Then, there is the DTD syndrome. Andrew Dillon, one of the first academics to take a scientific interest in IA, made his opinion clear (2005):

“We don’t have a definition for IA and we don’t need one.”

Dillon maintained that from the academic perspective, because of field shifts, he even foresaw the possible demise of computer science as a discipline tout-court. To IA, this would be a *“very real and present danger”*: as he put it, once the sleeping bear of shifting fields has awakened, it would start *“taking up IA work”* (Dillon, 2005).

Dillon definitely viewed the practicing or researching IA as a maker, a crafter, and recognized the inherent problem with the art and craft approach, first of all the incapability to guarantee consistent reproductions. Doers often cannot articulate their recipes very well²⁷. In one of his favorite examples, Dillon maintained that

*“it wouldn’t matter what Picasso would say about his process: we still wouldn’t be able to paint a Picasso painting. Indeed, crafting is not about the execution of sterile, calculated steps, but rather an intuitive response to a problem.”*²⁸

Clerwall (2003) asserts that information architecture is about constructing some kind of *“building of information”*, or to construct a building constituted of information. If

²⁷ To this extent, practice-led research seems to be a promising methodology. For more about the issues related to scientific discussion of arts and crafts disciplines, see the entry on practice-led research at the Centre for Research and Development, <http://artsresearch.brighton.ac.uk/links/practice-led> and Resmini, A. A Practice-led Qualitative Approach to a Theory of IA (work in progress). These won’t be discussed here.

²⁸ Dillon, A. Opening Keynote at EuroIA 2005. Cited in Gover (2006).

information is defined according to Langefors' theories then, IA is about constructing something which is meant to convey a message, with data that needs to be decoded by the recipient as correctly as possible. Clerwall finally defines IA as

“the art and science of designing a cohesive / interrelated information space, which is acceptable to the intended stakeholders”

It must be noted that Clerwall explicitly talks about presentation space:

“(...) my perspective focuses on what can be called ‘the presentation level’ of an information application. This means that when I write about the structuring of information, I refer to what is shown to the user. This means that I do not include the actual storage of data in my use of the term – although this is a very important part of the total architecture of interactive media.”

This is an extremely important reduction, and one that was not clearly there at the beginning. I will constantly refer to this reduced view when dealing with the Innerio Archives and with case-studies from other digital archives.

Francke (2008) works from a library science point of view, and considers information architecture to be largely about building functionalities for “large web sites”²⁹, while Morrogh (2003) maintains that information architecture is “primarily about the design of information environments and the management of an information environment design process”.

Regardless of these differences and disagreements, there seems to be some general consensus over a three-part definition pushed forth by the second edition of “Information Architecture for the World Wide Web” (Rosenfeld & Morville, 2002) and adopted by the Information Architecture Institute (IAI), an international professional organization dedicated to advancing the state of information architecture through research, education,

²⁹ Although it must be noted that she concedes it to be some kind of useful metaphor for dealing with other media, very much akin to document architecture. See Francke, 2008, p. 147.

advocacy and community service³⁰ (2005). According to this canon, information architecture is:

1. the structural design of shared information environments.
2. the art and science of organizing and labeling web sites, intranets, online communities and software to support usability and findability.
3. an emerging community of practice focused on bringing principles of design and architecture to the digital landscape.

While definition 2 seems to bind IA to the confines of World Wide Web-related design, a fairly large number of IA practitioners and scholars (Garrett, 2002; Rosati, 2007) maintained and maintain that IA as both a practice and a future discipline has more to it than the simple art of labeling and organizing of on-line content. Furthermore, the Information Architecture Institute explicitly encourages a constant reworking of this canon:

“Are these definitions definitive? Absolutely not. Our craft is new and still taking shape. We’re clear on the center but fuzzy at the boundaries. This inherent ambiguity challenges us to think deeply and seek diverse perspectives.”

As I said, the term “information architecture” itself has the more design-related part of its roots in the seminal work of architect R. S. Wurman, who re-coined or re-invented the term information architect in the 1970s and wrote an eponymous book in the mid 1990s (Wurman, 1997). The book dealt with the increasing difficulty Wurman was experiencing in communicating rising amounts of information. The primary task of the information architect was to make this information simpler, more direct, more comprehensible: the information architect was expected to draw lines and put some order in these oceans of data for better understanding and enjoyment. This concept was to be later expanded in a vastly more successful book, “Information Anxiety”. Wurman is the primary source and

30 The author operates as a Director for the Information Architecture Institute at the time of this writing.

the original instigator of the information design approach, as I outlined, and even though he was possibly mainly concerned with the static design of large quantities of visual information, his contribution was undoubtedly a major source of inspiration in the initial modern reframing of the field (Wodtke, 2002)³¹.

More recently, Morville has broadened the horizons of the field to encompass environmental design (Morville, 2005). For Morville, information architecture resides at the intersection of way-finding, social software, information retrieval, decision trees, self-organization, evolutionary psychology, librarianship, and authority. Morville maintains that where the Internet meets ubiquitous computing, the histories of navigation, communication, commerce, and information seeking converge to produce needs of findability, those very needs information architecture seeks to answer.

This shift is reflected in definitions 1 and 3 in the list given above, suggesting that this change or broadening of perspective has been well received by the larger IA community and is now part of the canon (Resmini & Rosati, 2009a).

Since information architecture relies on principles of human-information interaction which are largely independent from any specific medium or practice, it provides a flexible but solid conceptual model for the design of cross-context interaction models which span different media and environments, and hence can provide a constant cognitive framework throughout the whole interaction process.

To do this, information architecture has to undergo two different evolutionary changes:

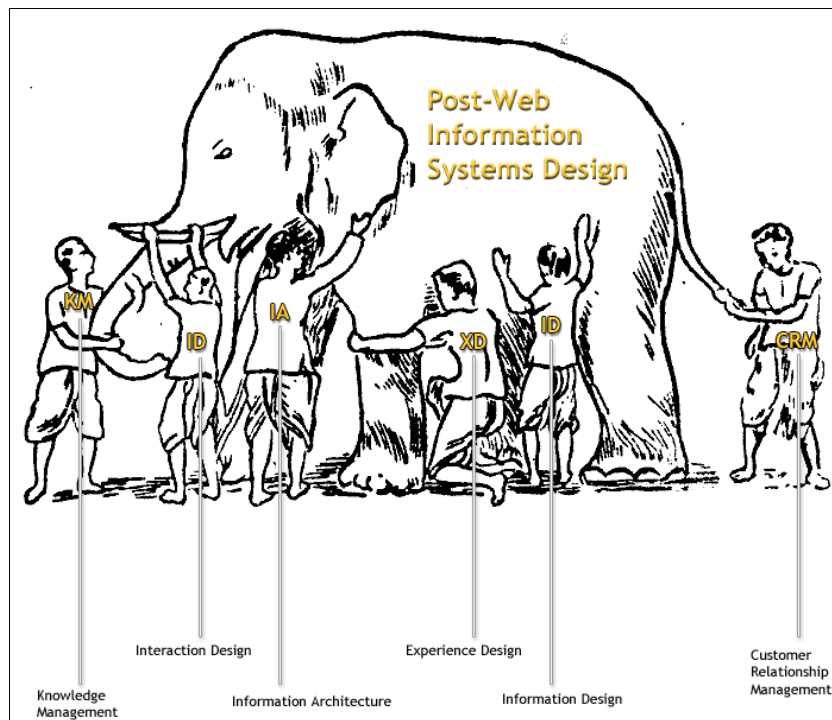
1. IA is not be limited anymore to taxonomies and Web design. As information moves to physical spaces, IA is used to design the entire range of shared informational spaces, places, services, and processes;
2. IA becomes the connector between different media and different contexts, providing a logic and experiential continuity to products and services which more and more span the digital and physical environments.

³¹ See also Dan Klyn's interviews with R. S. Wurman at <http://danklyn.com/blog/> for a better framing of Wurman's pioneering ideas in the context of modern information architecture.

While as we have seen the first point has been matter of discussion inside the professional and scholarly IA community through the years and has entered the canon, the second change is still to come, and although pivotal to the evolution of IA, this broader view has no direct connection with the core of this research and as such will not be dealt with.

Interaction Design, User Experience, and IA

*“So oft in theologic wars,
The disputants, I ween,
Rail on in utter ignorance
Of what each other mean,
And prate about an Elephant
Not one of them has seen”³²*



32 Saxe, J. G. (1873). The Blind Men and The Elephant. The Poems of John Godfrey Saxe. Complete edition. James R. Osgood and Company. Pp. 77-78. Available online at http://rack1.ul.cs.cmu.edu/is/saxe/doc.scn?fr=0&rp=_n.

Tracing down genealogies or assessing the right of the firstborn is definitely not what this research is about. Nonetheless, in the ever-stirring landscape of the fields and disciplines that deal with the design of information-rich environments, a few words on their history and relationships might help better understand why information architecture is considered central, as opposed to say interaction design or usability engineering, as far as the goals of this research are concerned.

A very well known traditional Indian story recounts the tale of how a group of blind men made their first acquaintance with an elephant and became enlightened on the ways truth can be elusive. As with all traditional stories, it has variations, and local versions. Although at least in the English-speaking world Saxe's poem is possibly the best known rendition, what I present here is the Jain version of the story. Jainism is a Dharmic religion, professing non-violence as a means to elevate one's spirit³³, and the story of the Elephant and the Blind Men is meant to illustrate how it can be possible to live in harmony with people who have different beliefs, and how truth can be presented in different, non antithetic ways. It goes like this:

Once upon a time, there lived six blind men in a village. One day the villagers told them, "Hey, there is an elephant in the village today". They had no idea what an elephant is. They decided, "Even though we would not be able to see it, let us go and feel it anyway." All of them went where the elephant was. Everyone of them touched the elephant.

"Hey, the elephant is a pillar," said the first man who touched his leg.

"Oh, no! it is like a rope," said the second man who touched the tail.

"Oh, no! it is like a thick branch of a tree," said the third man who touched the trunk of the elephant.

"It is like a big hand fan" said the fourth man who touched the ear of the elephant.

"It is like a huge wall," said the fifth man who touched the belly of the elephant.

"It is like a solid pipe," Said the sixth man who touched the tusk of the elephant.

33 For further reference on Jainism, see Gopalan, S. (1973). Outlines of Jainism. Halsted Press.

They began to argue about the elephant and everyone of them insisted that he was right. It looked like they were getting agitated. A wise man was passing by and he saw this. He stopped and asked them, "What is the matter?" They said, "We cannot agree to what the elephant is like." Each one of them told what he thought the elephant was like. The wise man calmly explained to them, "All of you are right. The reason every one of you is telling it differently because each one of you touched the different part of the elephant. So, actually the elephant has all those features what you all said."

"Oh!" everyone said. There was no more fight. They felt happy that they were all right.

The moral of the story is that there may be some truth to what someone says. Sometimes we can see that truth and sometimes not because they may have different perspective which we may not agree too. So, rather than arguing like the blind men, we should say, "Maybe you have your reasons." This way we don't get in arguments. In Jainism, it is explained that truth can be stated in seven different ways. So, you can see how broad our religion is. It teaches us to be tolerant towards others for their viewpoints. This allows us to live in harmony with the people of different thinking. This is known as the Syadvada, Anekantvad, or the theory of Manifold Predictions.³⁴

This story and its morals resemble in their development a hundred discussions touching upon the design of the digital and could be actually transposed as is. In the most accredited view, interaction design, usability engineering, information visualization, graphic design, and information architecture are parts of the elephant: user experience is the elephant³⁵. All field-fencing and name-calling by the most vocal practitioners³⁶ notwithstanding, this view purports a certain equilibrium and perfectly satisfies the general problems of addressing and identifying fields, and hence contributions, which are connected with the goals of this

34 Jainism Global Resource Center. Elephant and the Blind Men. Available at <http://www.jainworld.com/education/stories25.asp>. Accessed Oct 22 2009.

35 The most momentous embodiment of this vision being certainly Jesse James Garrett closing plenary at the ASIS&T Information Architecture Summit in Memphis. A transcript of the Memphis Plenary is available at <http://www.jjg.net/ia/memphis/>.

36 See for example Kolko, J. (2009). Our Misguided Focus On Brand and User Experience. <http://johnnyholland.org/2009/12/01/our-misguided-focus-on-brand-and-user-experience-how-a-pursuit-of-a-total-user-experience-has-derailed-the-creative-pursuits-of-the-fortune-500/>. Accessed Dec 15 2009.

research. We already discussed the history of IA, and provided some canonic definitions. We will briefly introduce the other related fields and disciplines to help better frame IA itself and its role in a positive change in the way we structure complex information and specifically complex information in the historical and juridical domains. As with IA itself, many of these definitions come from the work and discussions fostered and fueled by practitioners, as much of the self-awareness in the process still rests on their side. To better illustrate the reciprocal connections, I will refer to the so-called T-Model (Boersma, 2004).

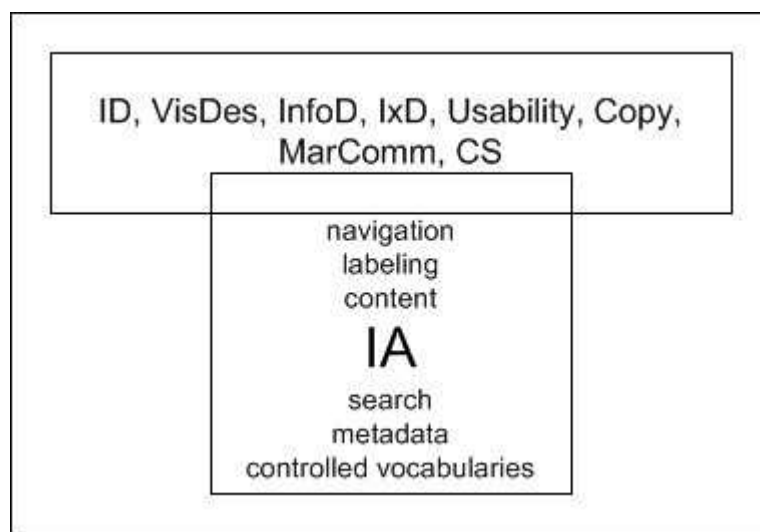


Figure 3: Peter Boersma, the T-Model, Morville's way

In discussing with Eric Reiss, then a Board member of the Asilomar Institute for Information Architecture³⁷, Boersma felt that there was a need to re-label the so-called Big IA³⁸ I briefly discussed earlier on into user experience. He proceeded to sketch his model in the shape of a capital “T”

“with the vertical line representing the field of IA with varying degrees of depth, while the horizontal line represented the width of related fields around (IA). The depth of IA ranges from shallow subjects that have clear overlap with the other fields to deep subjects that other fields hardly touch upon. Shallow subjects are navigation, labeling, and content that overlap

37 Later to become the Information Architecture Institute. Eric Reiss himself would move on to become a two-terms president of the IAI.

38 See par. A note on Big IA vs Little IA.

with interaction design, marketing and copywriting for example. Deep subjects would be search, metadata, and controlled vocabularies. Peter Morville’s little IA’s live here, and each would have his own strengths.”

In Boersma’s initial draft, the one explaining Morville’s view on Big and Little IA, related fields lie on the horizontal block and partly overlap in the area where the shallow subjects are. This seems unsatisfactory: Boersma elaborates that every field in the area of digital design has actually their own variant of the T-model, where their vertical line stands next to IA’s, and they *de facto* share the overlapping horizontal line. For Boersma

“the horizontal overlap is the place where User Experience (UX) practitioners operate. They are likely to have a background in one of the fields (their private vertical line) but in their work they focus on the horizontal line, orchestrating specialists who operate in their vertical.”

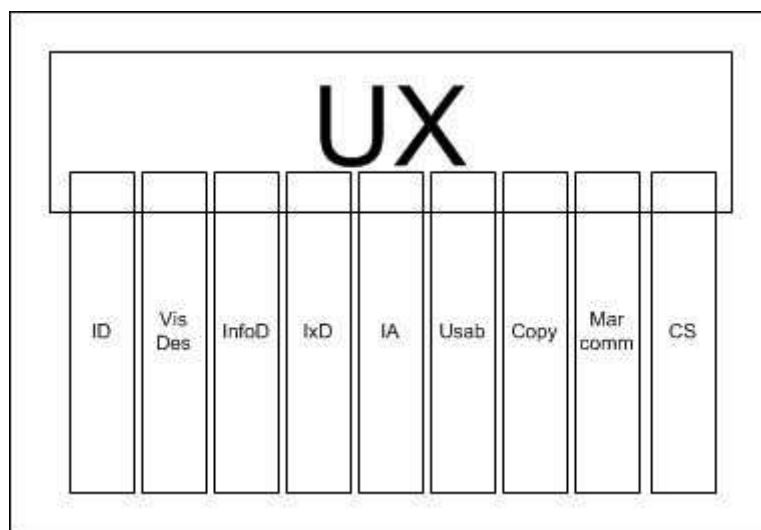


Figure 4: Peter Boersma, the T-Model expanded to visualize all UX fields

Boersma felt that this was a preferable and better approach, not being comfortable with Morville’s vision of the information architect claiming the “Big” label, as this meant “effectively placing the related fields *below* IA instead of at its *side*”.

Boersma later modified this initial scheme to take into account suggestions coming from the project management side of the discussion³⁹. Nonetheless, I will simply refer to the he first iteration of the model itself described here when necessary, since Boersma's second iteration simply adds a business layer to the top of the "T" which is of little interest to this discussion.

User Experience Design

Boersma defines user experience as occupying the horizontal level of his T-model, the place where shallow skills from all digital design disciplines overlap. In general, user experience design (UX or UXD) can be said to be the general design of the architecture and interaction models that impact a user's perception of a device or system. The scope of the field is directed at affecting "all aspects of the user's interaction with the product: how it is perceived, learned, and used" (Norman, 1999). This is important, and marks a shift from traditional human-factors, ergonomics, or usability engineering views. User experience design can be said to incorporate most of the disciplines Boersma mentions in his T-model (see Fig. 4) in a holistic approach to positively impact the overall experience a person has with a particular interactive system.

Interaction design

Interaction design, often abbreviated IxD to avoid it being confused with industrial design (ID), is a rather recent and successful field of practice in the arena of digital design. As usual, the term goes back a little longer than its actual current use: interaction design was first adopted by Bill Verplank and Bill Moggridge in the second half of the 1980s (Moggridge & Verplank, 1987). To Verplank, it was an adaptation of the computer science term user interface design to the industrial design profession.[5] To Moggridge, it was an improvement over soft-face, which he had coined in 1984 to refer to the application of industrial design to products containing software (Moggridge, 2006). In August 2003,

³⁹ Boersma, P. (2005). Shoulder IA: t-model expanded with business layer.
[<http://www.peterboersma.com/blog/2005/03/shoulder-ia-t-model-extended-with.html>]

following up long months of debate, Bruce Tognazzini, then a principal at the Nielsen Norman Group, wrote a post on his web site called “It’s time we got respect”⁴⁰. In this piece he maintained that the times were mature for the recognition of what got to be named interaction design following a community effort⁴¹. Shortly after that, the Interaction Design Association (IXDA), a volunteer non-profit organization, was founded with the goal to promote the practice and theory of interaction design. According to the IXDA

*“(i)interaction design (IXD) is a professional discipline that illuminates the relationship between people and the interactive products they use. While interaction design has a firm foundation in the theory, practice, and methodology of traditional design, its focus is on defining the complex dialogues that occur between people and interactive devices of many types — from computers to mobile communications devices to appliances ”*⁴²

Understanding Interaction Design

The primary focus of interaction design is the definition of interactivity in high technology products or services. The IXDA grounds their definition of what interaction design is in three principles: the creation of “useful and usable products and services”; the adherence to the “fundamental tenets of user-centered design”; the balancing of these with “business goals and technological capabilities”. The interaction designer should then be able to provide “solutions to complex design challenges, and define new and evolving interactive products and services”. Good interaction design should manage to:

40 Tognazzini, B. (2003). It's time we got respect.

[<http://www.asktog.com/columns/057ItsTimeWeGotRespect.html>]. Accessed Oct 03 2009.

41 Interestingly enough, Tognazzini's initial position was that “designer” was not the right word to use, because of associated “soft” characteristics. Tognazzini was a supporter of the wording “interaction architect”.

42 It might be interesting to highlight how a previous version of this page (retrieved Aug 2008) openly mentioned user experience design and actually framed interaction design as a part of its larger whole. The opening paragraph read: “Interaction design (IXD) is the branch of user experience design that illuminates the relationship between people and the interactive products they use. While interaction design has a firm foundation in the theory, practice, and methodology of traditional user interface design, its focus is on defining the complex dialogues that occur between people and interactive devices of many types—from computers to mobile communications devices to appliances. “ As of Jan 2010, that framing is no more.

- ◆ effectively communicate a system's interactivity and functionality;
- ◆ define behaviors that communicate a system's responses to user interactions;
- ◆ reveal both simple and complex workflows;
- ◆ inform users about system state changes;
- ◆ prevent user error.⁴³

In recent times interaction design has gained considerable traction and visibility inside the academia.

Usability Engineering

Usability engineering is a prescriptive field generally concerned with making human-computer interfaces more usable and user-friendly⁴⁴: assessing the usability of an interface and illustrating ways to improve it⁴⁵. Usability engineering has roots in disciplines as diverse as psychology, ergonomics, and cognitive science: as opposed to IA, IxD, and UX, usability is a *post-mortem* set of procedures which analyses, usually quantitatively and statistically, the way an already designed system interacts with its users. Because of its strong ties with human-computer interaction and conversely with academic discourse, usability engineering has had a certain role in the assessment of acceptable user standards in digital libraries and digital archives, and as such will be discussed in the chapter on Digital Libraries. It might just be worth mentioning though that user experience design and usability have an uneasy relationship. Bevan (2009) offers a straightforward if simple distinction in that where usability methods and usability assessments have the “objective of improving human performance”, user experience methodologies have the “objective of

43 IxDA. About Interaction Design. [http://www.ixda.org/about_interaction.php]. Accessed Dec 4 2009.

44 For the scope of this research, I define an interface as user-friendly when it allows users to effectively and efficiently accomplish the tasks for which it was designed and I will strictly refer to interfaces in the World Wide Web domain.

45 For a different stance on how usability relates to the improvement of interfaces, see Cooper, 2004.

improving user satisfaction with achieving both pragmatic and hedonic goals”. This is not far from Kaltenbacher’s view on UX being more attentive to the emotial aspects of interaction and it also resonates positively with long standing assumptions in the professional community⁴⁶.

Evaluating Information Architecture

This research will primarily focus on those user- and design-related aspects of IA that we identified within the “information design” and “architecture of information” approaches, giving precedence to problems of navigation, categorization, and general findability. In this respect, this research will also push the idea that context, or place-making, is of the utmost importance for users to successfully understand and use a digital library or digital archive (Resmini & Rosati, 2009a). A basic methodology has been used to heuristically evaluate the IA of the system during the development of the second iteration of the Irnerio Archives: I will briefly introduce it here, as this will allow me to better frame the competitive analysis in Chapter II, Digital Libraries and Digital Archives.

Being IA currently practice-led, practical, professional methodologies for IAs design abound, but very few general-scope evaluation methodologies are to be found. One of the most complete is the one written by Steve Toub (2000), then an Argus Associates employee, for the ACIA. The methodology is rooted in Argus practices of the time, with a certain slant, towards business processes and “IA as information science”⁴⁷, and that has to be considered. I will describe it, comment it, and then describe the updated, or reduced, depending on the point of view, heuristic evaluation that was developed during the design of the Irnerio Archives.

46 See for example Peter Morville's UX honeycomb. Morville, P. (2004). User Experience Design. <http://semanticstudios.com/publications/semantics/000029.php>. Accessed Oct 16 2009.

47 Toub defines IA as “the art and science of structuring and organizing information environments to help people effectively fulfill their information needs.”

It must be remembered that the guidelines were meant both as an internal tool for information architects to evaluate the general quality of an artifact, possibly theirs, and to prove that IA had business values to be exploited. Toub stresses that his white paper

“(p)rovide(s) background on why evaluating IA is important; Place(s) IA evaluation in context and provide an overview of methodological issues; Describe(s) techniques for evaluating aspects of an IA, particularly structure and grouping.”

And Toub is obviously keen in presenting a case for why his first point above is important:

“An evaluation can identify opportunities to improve a web site, maximizing profit and increasing productivity. (...) Evaluation is often necessary to help justify the value of the investments in IA.”

Once firmly stated the business significance of IA, Toub moves on to the methodology proper, and identifies site organization as the core of a specific IA evaluation: this involves the *“structure, grouping, and labeling of site content”* (Toub, 2000). Elements considered essential but possibly covered by specially adapted usability techniques include navigation, page design, and indexing, and as such they are not covered by Toub as his paper *“focuses on evaluation techniques unique to web site organization”*.

While indexing is undoubtedly a technical element, I disagree with the idea that page design and navigation can be successfully covered by usability metrics alone. Page design is a fundamental piece of the structuring of content, and cannot be evaded while doing competitive analysis. As such, it embeds and embodies a good part of the initial design vision. Navigation is tightly coupled with way-finding and context-awareness, and it is an important component of place-making⁴⁸. Place-making is defined by Resmini and Rosati (2009) as

⁴⁸ It is perhaps worth noting that at the time Toub writes, Andrew Dillon, as reported by Toub himself, had just called for a new set of specific IA metrics based on aesthetics, cognitive overload, perceived usability, perception of information shapes, acceptance level, and self-efficacy (Dillon, 2000).

“the capability (of an information architecture) to help users reduce disorientation and increase legibility and way-finding in digital, physical, and hybrid environments”

For similar if opposite reasons, information retrieval techniques which have a rather high business value like log analytics and that Toub recommends, were not used and are not part of my updated model. While log analysis was used during the design phase to attest how the Innerio Archives were performing in terms of “which search terms yield zero results or ‘too many’ results” (Toub, 2000) and to verify search-engine placement, it was not an integral part of the information architecture strategy and as such will not be discussed.

Types of Evaluation

Evaluations occur at any phase of design and development. They can measure the performance of a whole information architecture at a given moment in time, or they can:

- ◆ Compare your site to a competitor’s site ;
- ◆ Compare the new version to previous version(s) of your site ;
- ◆ Compare several different design options for the same IA design issue: for example, comparing a broad and shallow hierarchy versus a narrow and deep hierarchy for the same site

Furthermore, evaluations can be quantitative, systematically and empirically investigating properties and phenomena to employ and produce a mathematical model, or qualitative, aimed at an in-depth understanding of human behavior and the reasons that govern it. Because of the general constraints of the project and of this research, and because of the inherent nature of IA as a practice-led, arts and crafts, design-driven field, all evaluations conducted, and hence the final model, were qualitative.

Evaluation Process

Toub (2000) outlines a three-part process which evaluates structure, grouping, and labeling. This in turn is aligned with a general sequential scheme which identifies five different, loosely labeled, stages of development: identifying requirements, research, early design, later design, post-launch. Heuristic evaluations can happen at any of these stages, or not at all.

Toub presents the general scheme in table format in his white paper:

Phase	Types of Evaluations
Identifying requirements	<ul style="list-style-type: none"> • Obtain base-line measurements that can be used for later comparisons (Qualitative or Quantitative) • Validate existing IA designs with users (Qualitative) • Compare your IA to the IA of a competitor's site (Qualitative or Quantitative)
Research	<ul style="list-style-type: none"> • Explore users' grouping/ labeling behavior (Qualitative)
Early design	<ul style="list-style-type: none"> • Validate IA designs with users (Qualitative) • Compare this IA to a previous IA for the same site (Qualitative or Quantitative)
Later design	<ul style="list-style-type: none"> • Refine/ validate an IA design (Qualitative)
Post-launch	<ul style="list-style-type: none"> • Compare this IA to a previous IA for the same site (Qualitative or Quantitative) • Compare this IA to the IA of a competitor's site (Qualitative or Quantitative)

Figure 5: Steve Toub. (2000). Evaluation phases and types of evaluation.

Because of the way the Innerio Archives were designed, and because of concerns with the applicability of the methodology *per se*, for example the requirements were already largely set by contract and could not be negotiated nor re-assessed, Toub procedures were largely remodeled into a new scheme:

- ◆ Early design: (B) compare this IA to a previous IA for the same site (qualitative);
- ◆ Later design: refine / validate an IA design (qualitative);
- ◆ Post-launch: (A) compare this IA to a previous IA for the same site (qualitative).

The comparison with IAs of competitors was not carried out in post-launch as in Toub’s evaluation heuristics, but during early and later design. Place-making assessments were similarly carried out. The final process can be schematized as follows:

Early design	competitive analysis
	compare this IA to a previous IA for the same site
	place-making assessments
Later design	refine / validate an IA design
	finer competitive analysis
	place-making assessments
Post-launch	compare this IA to a previous IA for the same site

Table 1: Evaluation of the Information Architecture for the Irnerio Archives

Heuristics in the Evaluation Process

As I mentioned above, Toub identifies three general areas of assessment: structure, grouping, and labeling. These are reported here for completeness. Structure is the

“completeness (i.e., the inclusion or exclusion of content and functionality) and form (e.g., hierarchy vs. database) of the site as a whole.”

At this stage responsiveness to user’s needs, structural design, hierarchies, and completeness and form of content objects, such as applications, documents, or pages can be assessed. This is usually heuristic-based, as experts evaluate the design and compare it to “agreed-upon design principles”.

Grouping is defined as the

“coherence and integrity of categories or groups of content objects.”

Categories, their overlapping, representativity of elements inside categories, can be assessed at this stage. This is often done with methods such as card sorting or freelisting. None of these has been extensively tested on the Innerio Archives.

Finally, labeling refers to the

“name or icon of a content object, such as the title of a page, or the title of a category or heading.”

It generally includes all issues pertaining to navigation, way-finding, and place-making. This was investigated thoroughly during development of the Innerio Archives.

A Final Consideration

Toub briefly introduces the idea that in using an information architecture known-item finding⁴⁹ tasks, tasks performed by users that already know what they are looking for,

“do not reflect the entire range and complexity of real-world tasks. Finding, especially on the Web, tends to involve open-ended exploration (“gathering”), a shifting of goals and priorities (“berry-picking”) or comparisons. It also includes switching quickly between each of these different modes of finding in the same task. Also, finding tasks are not ends in themselves; they are the stepping-stones to larger tasks and goals.”

This is a simple yet often overlooked point in the design of shared information spaces. This concern, to account for gathering and berry-picking and allow users to step from smaller tasks to larger tasks as easily as possible, was also reflected in the set of open

49 For more on this, see Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. [<http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>]. Accessed 5 November 2009. Or Bates, M. (2002). Toward an integrated model of information seeking and searching. Keynote Address, 4th International Conference on Information Needs, Seeking and Use in Different Contexts. [http://www.gseis.ucla.edu/faculty/bates/articles/info_SeekSearch-i-030329.html]. Accessed 5 November 2009.

information architecture questions that were formalized during the initial stages of redesign and that were utilized during competitive analysis⁵⁰. The most important ones being connected with place-making and with how to deal with a different navigation paradigm for a supposedly well know item such as a book, when it is in fact not a book? How to deal with missing data or metadata in order not to impair the user experience? How to deal with content which is largely, from a cultural point of view, either inaccessible or scarcely comprehensible?

These questions will be discussed in detail and documented in their resolution in Chapter II, Digital Libraries and Digital Archives, when illustrating the case studies.

⁵⁰ Please note that these are strictly IA questions, and have possibly no direct implications to historical or juridical treatment of the archives.

Chapter II

Digital Libraries and Digital Archives

What is a Digital Library

In a way, the term digital library (DL) is simply the most recent way to describe an idea, that of the computerized library, that has been around almost as long as computers themselves. Licklider envisioned in 1965 a computer-based library he called the “library of the future”, and through the years a number of different terms such as electronic library, virtual library, and library without walls, have been used¹. All of these different wordings described one single vision, a non-physical library “that would supplement, add functionality, and even replace traditional libraries” (Harter, 1996).

The use of “digital library” is directly connected to the birth of the Digital Libraries Initiative funded by the National Science Foundation, the Advanced Research Projects Agency, and the National Aeronautics and Space Administration in the United States (Harter, 1996). In 1994 these agencies granted substantial amounts of money to a number of US universities to be employed in “digital library research”. Hence, even if the term itself has possibly no more than 15 years of history behind its back, “work in bringing digitized information resources to libraries has a history spanning several decades” (Harter, 1996).

“With few exceptions, formal published papers on aspects of digital libraries delivered at conferences or published in scholarly journals tend to focus on technical topics and are written by computer scientists or librarians with a technical bent. With all this activity, little attention has been paid to what is being built -- how digital libraries actually are libraries in the traditional sense of a library -- and how they are not. As we

¹ For more historical information on the process, see Drabenstott, K. (1993). Analytical Review of the Library of the Future. Council on Library Resources.

build new digital libraries we add functionality -- capabilities that have never been present in traditional libraries.”

Stephen Harter addressed these words to the audience of the International Conference on Digital Libraries and Information Services for the 21st Century in Seoul, Korea, in 1996. His paper considers the nature of the digital library oscillating from a narrow profile, strictly based on the properties of a traditional library, to a broad profile on the basis of which the Internet can be considered, if uselessly, one single digital library. This huge variance is a problem to be confronted and Harter states very clearly that

“the idea of digitization is perhaps the only characteristic of a digital library on which there is universal agreement.”

This comes as no surprise though, as digital libraries are by their nature a multi-disciplinary effort. As such, not only digital libraries have been variously defined in accordance with the leading perspective of the moment, but the digital library systems to handle them have been developed pragmatically as well, borrowing and adapting techniques from other disciplines as it seemed fit. This has made having a common conversation harder, technical interoperability difficult (Candela et al, 2007), and the development of a standard information architecture for DLs either impossible or seemingly unnecessary.

Harter summarized this variance in a table presenting three columns². Column 1, labeled Narrow View, lists essential characteristics we can find in traditional libraries. From the point of view of a digital library, the extent to which any of these is part of the design defines the library itself. For example,

“a digital library may be organized by human specialists (indexed, classified, cataloged) or it may be entirely unorganized, using free text searching for providing some or all access to the objects in the library.”

2 It could be argued that a digital library might also have properties that a traditional library does not have, but these are not part of Harter's recognition work.

Similarly, a traditional library has a physical location (item 3), and most of the objects therein preserved are information resources of some kind (item 1), according to the type of library.

	Narrow View	Broader View	Broadest View
1	objects are information resources	most of the objects are information resources	objects can be anything at all
2	objects are selected on the basis of quality	some of the objects are selected on the basis of quality	no quality control; no entry barriers
3	objects are located in a physical place	objects are located in a logical place (may be distributed)	objects are not located in a physical or logical place
4	objects are organized		no organization
5	objects are subjected to authority control	some aspects of authority control are present	no authority control
6	objects are fixed (do not change)	objects change in a standardized way	objects are fluid (can change and mutate at any time)
7	objects are permanent (do not disappear)	disappearance of objects is controlled	objects are transient (can disappear at any time)
8	authorship is an important concept	concept of author is weakened	no concept of author
9	access to objects is limited to specific classes of users	access to some objects is limited to specific classes of users	access to everything by everyone
10	services such as reference assistance are offered		the only services are those performed by computer software (AI)
11	human specialists (called librarians, etc.) can be found		there are no librarians
12	there exist well-defined user groups	some classes of objects have associated user groups	there are no defined user groups (or, alternatively, infinitely many of them)

Table 1: Harter, S. (1996). *Properties of a Digital Library*

Item 5 and 8 in the list, authority control and authorship, are a key feature of a traditional library: names of authors, variants of works (editions), and subject headings or descriptors are all verified and controlled. It can be certainly said that authorship should be a key feature in digital libraries as well, and decidedly forms a core requirement for the design of a digital library that wants to preserve the chain of causality and ownership in documents of legal relevance:

“The importance of authorship and ownership in the traditional library reflects the importance of these ideas in traditional scholarly communication, where scholars and scientists cite in reference lists the authors and works from whom they have borrowed ideas, words, or facts. This is how (...) original authorship is acknowledged. Formal legal rights of ownership are defined by national and international copyright law.”
(Harter, 1996)

Then items in a traditional library obey certain rules. These have to do with their physical nature, or status (Items 6 and 7):

“First, they are fixed - they do not normally change, or if they do, various editions are identified and considered to be different from one another. Objects are also permanent - they do not normally disappear from a collection.” (Harter, 1996)

And finally, services are usually available by means of specialized personnel, and acted upon the content and the library itself, with some limitations (Items 9-12). The library offers

“help with searching information resources, reference and research services, readers advisory services, and others. A traditional library typically offers only limited access to materials and services; access is restricted to certain classes of potential users.”

Harter maintains that among these properties “physical location is the least likely to survive in a digital library”, with resources being effectively distributed among different places, but he also prospects that all of these properties are jeopardized by the very

concept of the digital library. In which way these can be challenged is what column 2 and column 3 are all about: by acting upon on the core set of properties in column 1, either broadening them a little or broadening them a lot, we run the full range from physical, controlled, and serviced, to digital, uncontrolled, and automated³. In other words, the broadest view prospects a digital library where the properties of the traditional library have largely disappeared:

“(q)uestions then immediately arise concerning science, scholarship, teaching, and learning. What will happen to the concept of authorship and the notion of fixed, permanent documents? How will these changes affect scholarship and research? These are crucial social questions that are extremely important to contemplate⁴.”

This led Harter to a list of questions that I reproduce here entirely in Table 2⁵. It is worth noting that his final set of issues presents again some of the problems that are at the core of the research questions this work tries to frame and at least partly address.

- ◆ How can we establish and control the currency, accuracy, and integrity of information sources (quality problem)
- ◆ How can we maintain the data and intellectual integrity of IRs? (authority control problem)
- ◆ What can be done to provide intellectual access to IRs? (organizational problem)
- ◆ How can we recognize different versions of the same IR? (fluidity problem)

3 And although some individual items are debatable, it is certainly be said that in this light the World Wide Web might as well fit into the description provided in column 3 of the table. Harter notes that “(...) the Internet itself as it exists today (...) has essentially none of the properties of the traditional library listed in (the) table (...). The Internet is anarchic and individualistic. It is not a collection of information resources selected on the basis of their quality, organized by subject, etc. There are real problems with the concept of author on the Internet. The concept of control is almost entirely absent. Many of the objects on the Internet are here today and gone tomorrow. Those that remain are in a constant state of change. There are very few services offered by human beings (as opposed to computer software). The metaphor of the traditional library is no longer applicable to such a library.” (Harter, 1996).

4 Harter also interestingly comments that “if supporting scholarly communication and research is a goal (...) then there are several kinds of information resources not collected by the traditional library that might be considered. These artifacts of the Internet are products of informal communication among scientists and scholars that are produced in digital form and are easily captured and saved.”

5 Hartes passingly mentions that “users and their environments have been little discussed in the literature on digital libraries”, but then he neglects somewhat to address this issue in his questions. See Table 2.

- ◆ How can we address the issue of transient IRs? (preservation problem)
- ◆ How can we preserve the concept of authorship?
- ◆ How can copyright laws for IRs be observed? (legal problem)
- ◆ Will access to some IRs be limited to some classes of users? (political problem)
- ◆ What services, if any, should be offered by the digital library?
- ◆ Should digital libraries be integrated into traditional libraries? If so, how can this be accomplished?
- ◆ Does a digital library have librarians? If so, what do they do?
- ◆ Does a digital library have well-defined classes of users?

Table 2: Harter, S. (1996) Questions and Issues Related to Information Resources (IRs) in the Digital Library

The National Science Foundation (NSF), as I said earlier on one of the turn-key agencies in the DL landscape, provides this definition of what a digital library is on its web site for the Digital Library Initiative Projects (NSF, 1999):

Digital Libraries basically store materials in electronic format and manipulate large collections of those materials effectively.

It is a loose and simple definition, which identifies a digital library according to its content and its successful handling of it. It does not say anything in respect to what channels a digital library should or might use, and it leaves any hint of network or Internet access out of the picture: the basic tenet is that a DL simply stores materials digitally, and allows effective maintenance and access. According to this definition, the original scannings prepared by CIRSIFID from the codices in the Royal College of Spain collection, being definitely ordered and made manipulable, that is searchable and accessible, on DVD, qualify as a digital library. Nonetheless, the NSF believes that research into DL is

“(...) research into network information systems, concentrating on how to develop the necessary infrastructure to effectively mass-manipulate the information on the Net”

So, even if the Net (sic) is currently not part of what a DL is, research should be carried out to exploit the mass-manipulation capabilities of the World Wide Web. Seadle and Greifeneder (2007) provide a more complex and layered definition:

“A digital library is fundamentally a resource that reconstructs the intellectual substance and services of a traditional library in digital form. Digital libraries consist of digital contents (which are sometimes but not necessarily text-based), interconnections (which may be simple links or complex metadata or query-based relationships), and software (which may be simple pages in HTML or complex database management systems). A single, simple, stand-alone web page is probably not a digital library in any meaningful sense, any more than a single page or a single book is a traditional library. A mass of raw data such as comes from the Hubble telescope is probably also not a digital library, though its contents arguably belongs in one. Digital libraries are not replacements for traditional libraries. They are rather the future of traditional libraries, much as medieval manuscript libraries simply became a specialized and much revered part of the larger print-based libraries that we have today.”

The first two lines of this latter definition are extremely interesting. A DL is a *resource*⁶, something that can be used as support and help, which *reconstructs* the *intellectual substance* and *services* of a traditional library in digital form. I understand this as saying that a traditional library is the *generic model* a DL tries to replicate in information space, and not that, more strictly, there is no digital library without a physical library behind it. Further confirmation to this interpretation seems to come from the Hubble example, where the point in case for it not being a DL as is amounts to the fact that it is raw data, definitely belonging to a DL but not constituting one *per se* for lack of some essential characteristics.

These characteristics, these salient components, are identified as: digital content, interconnections, and software. Digital content as these DL may contain text, but also multimedial artifacts such as sound, videos, or, as it is the case with the Imerio Archives, pictures; interconnections, as DL may link, have complex metadata overlays, or resort to

6 The Merriam-Webster dictionary reports that a resource is: a) a source of supply or support; an available means —usually used in plural; or b) a natural source of wealth or revenue —often used in plural; and e) a source of information or expertise.

IR-type queries, but they definitely correlate elements in the collection; software, as this is the layer which manipulates the element in order to allow user interaction. This covers the gamut from simple XHTML pages to full-blown stand-alone programs.

Miksa and Doty (1994) argue that no discussion can be carried out without considering that the naming presents interesting aspects that usually go unobserved and that have implications in the way we approach the phenomenon:

“(w)hy should a digital library, or an electronic library, or a virtual library - for the purposes of the remarks here these three terms will be considered synonymous - why should such a phenomenon be called a "library"? A digital library might well be called something else - a digital information system, or a digital publishing system, to name two possible alternatives.”

On this basis, they proceed to define a digital library as a collection of information sources in a place, either a physical or a logical place, as a broader definition would imply that the artifact is something else from “what is normally understood to be a library”.

From a more technical standpoint, Candela, Castelli and Pagano (2007) argue that a digital library is actually one component of a base information system architecture which also comprises a digital library system (DLS), and a digital library management system (DLMS). The digital library is “the abstract system as perceived by end-users”, while the DLS is the software “providing all the functionality that is required by a particular DL”. The DLMS is then simply the software that allows administration of the DLS and DL (Candela et al, 2007).

Finally, according to the Digital Library Federation (2002), a digital library is:

(an) organization(s) that provide(s) the resources, including the specialised staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily available for use by a defined community or set of communities.

It is interesting to note how this definition provides a framing for institutions, or digital libraries as working environments, but not for digital libraries as artifacts and collections.

For the scope of this research, where the digital library or digital archive is simply the means to an end, that of improving the overall information architecture and scientific fruition of complex information in the historical and juridical domains on-line, I will define the digital library as a World Wide Web-enabled abstract artifact which renders organized, correlated digital content pertaining to one or more physical or logical collections accessible and findable by users.

Much more important in connection with the general goals of this research, it seems that is possible to reuse or reinterpret Harter's original table of properties in the light of the requirements and inner workings of a digital library for the handling of complex and dynamic information like this:

Narrow View	Broader View	Broadest View
objects are information resources		
	some of the objects are selected on the basis of quality	
	objects are located in a logical place (may be distributed)	
objects are organized		
objects are subjected to authority control		
		objects are fluid (can change and mutate at any time)
objects are permanent (do not disappear)		
authorship is an important concept		
	access to some objects is limited to specific classes of	

	users	
		the only services are those performed by computer software (AI)
		there are no librarians
	some classes of objects have associated user groups	

Table 3: A compound table of properties for digital libraries handling complex and dynamic historical and juridical information.

This compound table will be recalled when dealing with case-studies and with the Innerio Archives.

One final note on the terminological dichotomy between digital libraries and digital archives. Without any doubts, traditional brick-and-mortar archives are different from libraries: an archive usually contains primary sources⁷ of information such as letters, paintings, documents pertaining to a certain subject, individual, or organization; a library contains secondary sources, such as books.

Furthermore, traditional archives usually contain unique content, such as the manuscript collection conserved at the Royal College of Spain for example, as opposed to the copies maintained by libraries. Digital archives of course break down this distinction: while they still generally contain copies of single primary sources, these can be reproduced by third parties and may actually have been reproduced from other sources on the Web.

For this reason in the course of this research, while I will refer to the Innerio digital collection as the Innerio Archives for historical reasons, the term digital library and digital archive will be used interchangeably and often shortened to DL. Finally, the term will be used to identify the artifacts and collections, and not to any organizational superstructure.

⁷ I follow the historiographic lesson that identifies primary sources as documents or recordings of any kind created at the time being investigated, and secondary sources as sources that build upon the former by citing, quoting or otherwise commenting and speculating on them.

Information Architecture and Digital Libraries

In their thorough report for the Digital Library Federation, Greenstein and Thorin (2003) underline how, after a questionnaire was circulated to academic libraries connected to the DLF in 2001, the data they received was still undoubtedly reflecting an unsettled scenario: different approaches, different programs, heavy ties to campus personalities, circumstances and needs (Greenstein, Thorin, & McKinney, 2001). In this light they traced a development path, a coming of age, leading from a young digital library to a maturing digital library and finally to an adult digital library. The most interesting remarks concern the future library, a new holistic and cross-medial organism:

“(i)t is becoming apparent that the adult digital library program will no longer be organizationally or functionally distinct from the library as whole. When digital libraries mature, we will talk about libraries in a manner that assumes electronic information, computer technologies, and networked collections and services are as much a part of the infrastructure as book stacks and catalogs were for the traditional library.”

This is an important part of the ongoing discussion, and an important part of the normalization spirit that lies at the base of this research trying to bring together what should be a mere technical artifact, “*as much a part of the infrastructure as book stacks and catalogs*”, an applied art perspective on the way this should be made useful and intelligible to intended audiences, information architecture, and the need to preserve juridical meanings and status that legal informatics is bringing to the table. This is still a design perspective, and Bonsiepe is not far away: nonetheless, information architecture has historically being absent in the larger conversation on the design of digital libraries.

Why is that, or what reasons might lie behind an apparent dismissal of reciprocal attentions is hard to say without proper investigations, which again lie outside the scope of this research. Nonetheless, even a brief nominal check illustrates perfectly the lack of common research across the two fields. At the ASIS&T IA Summit⁸, now in its 11th year

8 IA Summit, <http://www.iasummit.org/>

and the largest international IA-related event, we find only two accepted submissions, one in 2004 and one in 2006, that relate to DL or DA. Both of these seems to be somewhat tangential to the core nature of the relationship between information architecture and digital libraries.

In 2004 Gary Geisler⁹ presented a talk titled “A Case Study of Redesigning a Digital Video Digital Library”¹⁰. The talk described the Open Video Project, a “*digital library of contributor-donated digital video and associated metadata, used by a diverse set of users with wide variety of interests and goals*”. The project was based at the University of North Carolina, and began in 1999 with a simple web site and less than 200 videos. Through the years, and up until 2004, it grew up to contain roughly 2000 videos.

The then current iteration of the web site was said to be built around information architecture principles, and Geisler reported that the major challenges were connected to the fact that

“users vary widely in their individual characteristics (e.g., demographics, technology experience, video experience), their goals (e.g., learning, work, research, entertainment), and their specific tasks (e.g., finding a known video, finding a specific type of scene or visual style, browsing for something fun to watch), while the characteristics of our content--digital video--also vary greatly in terms of genre, topic, and style”.

In 2006, at the 7th ASIS&T IA Summit in Vancouver, Canada, Ramesh Srinivasan hosted a panel on “Reinvoking Culture and Context in Digital Libraries and Museums”.¹¹ Its abstract maintains that “(Srinivasan) will argue for the creation of digital library, museum, and other collections with an acknowledgment of the local context and culture of objects. This can be achieved without sacrificing interoperability issues. The approach in this panel

9 While commentaries and the official web site for the IA Summit 2004 report Anthony Hugues as one of the authors, the paper itself does not. I will follow the paper.

10 The <http://open-video.org> web site is still online today and has topped at 4200 videos.

11 Unfortunately, although mentioned on the ASIS&T IA Summit web site and on numerous personal pages Ramesh Srinivasan hosts at the University of California at Los Angeles, the paper seems not to be available anymore and the author failed to retrieve it.

will describe how to begin to achieve such a harmony between local knowledge of an object and standards, information sharing, and information access”. Even though touching upon central IA themes as those connecting local issues and international standards, the panel still does not seem to touch core DL-related IA questions such as those presented in Chapter I.

The ASIS&T European IA Summit (EuroIA)¹², as of this writing in its 5th year, has never presented any digital library-related papers so far.

A similar literature review on the opposite side, that pertaining to digital library conferences and initiatives, shows a confirming pattern of scarce or inexistent cross-pollination. Nonetheless DELOS¹³, a network for the advancement of digital libraries partially funded by the European Commission in the frame of the Information Society Technologies Programme, report on their web site that one of the core issues of their WP1 is to enable “the coordinated development of information architectures by an adoption of a set of common standards and protocols”¹⁴. This obviously recalls a definition of information architecture possibly more akin to what we called the 1980s information systems IA than to what the IA community and this research mean, but a mention is most decidedly there.

Even more interesting, the DELOS vision was that “Digital Libraries will become the universal knowledge repositories and communication conduits for the future, common vehicles by which everyone will access, analyze, evaluate, enhance, and exchange all forms of information”. These digital libraries will be accessible “at any time and from anywhere, and will offer a friendly, multi-modal, efficient, and effective interaction and exploration environment”.

12 EuroIA, <http://www.euroia.org/>

13 DELOS, Network of Excellence in Digital Libraries. The program started on January 1st 2004 and lasted 48 months.

14 DELOS. Network of Excellence in Digital Libraries. Digital Library Architecture (WP1). http://www.delos.info/index.php?option=com_content&task=view&id=20&Itemid=46. Accessed Jan 11 2010.

To push forth this vision, DELOS has been trying to bridge the gap between what DLs are and what DLs should be by researching into a number of critical aspects of digital libraries. DELOS made a point of their multi-disciplinary approach:

“Digital Libraries represent the meeting point of a large number of technical areas within the field of informatics, i.e., data management, information retrieval, document management, information systems, the web, image processing, artificial intelligence, human-computer interaction, mass-storage systems, and others. Moreover, Digital Libraries draw upon several other disciplines and fields beyond informatics, such as library sciences, museum sciences, archives, sociology, psychology, etc.”

There is informatics, and then there are other fields and disciplines. And in the vast sea of informatics there is a passing reference to human-computer interaction, but there is no apparent concern for any user-centered digital design practice, let alone information architecture. On the other hand, even though it was not something DELOS focused on, usability and to a lesser extent user experience have been part of the digital library landscape for quite a few years now.

Usability and User Experience for Digital Libraries

Usability is covered by two different ISO standards: ISO/TR 16982:2002, “Ergonomics of human-system interaction - Usability methods supporting human-centered design”, provides information on user-centered usability methods which can be used for design and evaluation. It basically details advantages and disadvantages of the different approaches and it lists other factors that might impact the use of any of the methods; ISO 9241 (1998) is a multi-part standard that covers a number of aspects for people who work with computers. It was originally called “Ergonomic requirements for office work with visual display terminals (VDTs)” and was later retitled to a more generic “Ergonomics of Human System Interaction”¹⁵. After a general introduction in Part 1, task design in Part 2, and

¹⁵ Some renumbering and reordering is being performed as well to allow the standard to include more topics. The first section to be renumbered was section 10 (now section 110).

physical characteristics of computers in Parts 3-9, Part 110 (ex 10) and 11-19 deal with usability in software: general usability heuristics for the design of different types of dialogue, and general guidance on the specification and measurement of usability. This is considered the core documentation.

An important, related and often quoted standard is ISO 13407:1999, “Human-centred design processes for interactive systems”¹⁶. While not dealing directly with usability standards, it outlines a design procedure for user-centered systems and applications.

Beginning in the early 2000s, researchers have attempted to evaluate different aspects of digital libraries. This is reflected in some of the DELOS-related literature, for example, and in the proceedings of conferences such as JCDL and ECDL. These studies have been largely influenced by a human information behavior approach though, and usability has a different meaning according to the discipline from which evaluations originate: for example,

“librarians perceive the usability of an information service in terms of efficient and effective access to information.” (Chowdhury, 2004b)

while the human-computer interaction community usually defines usability in respect to user interfaces (Nielsen and Lavy, 1994) and their effectiveness, efficiency or user satisfaction (Nielsen, 1993).

In the context of digital libraries, usability is usually defined as how easily and effectively users can find information, with an increasing emphasis being placed on the user (Dillon, 1994), while Van House et al. (1996) maintain that the usability of digital libraries depends on three key components: content, functionality and the user interface.

A more quantitative approach guidebook was written based on a research with 24 DLF member libraries (Covey, 2002). A distinguished DLF fellow, Covey conducted interviews, surveys, focus groups with personnel, and performed transaction log analysis

16 Now ISO FDIS 9241-210 (2009), Human-centred Design Process for Interactive Systems. For a full list of all ISO publications concerning usability engineering, see the Bluffer's Guide to ISO 9241, <http://www.userfocus.co.uk/resources/iso9241/>, and check the Introduction.

in order to document, among other indicators, “why digital libraries assessed the use and usability of their on-line collections and services ”, and “what aspects of those collections and services they were most interested in assessing ” (Troll Covey, 2002). She finally tried to evaluate the methodologies involved and how, or if, their assessments had any real impact on the libraries. She remarked that five key challenges had to be addressed for this to happen:

1. Gathering meaningful, purposeful, comparable data
2. Acquiring methodological guidance and the requisite skills to plan and conduct assessments
3. Managing assessment data
4. Organizing assessment as a core activity
5. Interpreting library trend data in the larger environmental context of user behaviors and constraints (Covey, 2002).

Covey concludes that dissemination of “a limited subset of digital library statistics and performance measures to facilitate gathering baseline data and enable comparisons ” would go a long way, as much as white papers, manuals and workshops covering

“the popular assessment methods (surveys, focus groups, and user protocols) and the less well-known but powerful and cost-effective discount usability testing methods (heuristic evaluations and paper prototypes and scenarios)”

would definitely steer the process in the direction of better guidance. Saracevic (2004) maintains that a usability evaluation of digital libraries, while possible, configures as a difficult task to accomplish because of a number of reasons:

“complexity – digital libraries are highly complex and therefore are hard to evaluate; pre-maturity – digital libraries are in an early stage of development; interest levels – interest in evaluation has been limited; funding – limited funds have been made available for evaluation; culture – evaluation is not a part of the culture of operating digital libraries; and cynicism – who wants to know about or demonstrate performance? ”

Usability is concerned with requirements and metrics to evaluate results and grade them on a given scale. User experience on the other hand, while broadening the spectrum to include

“(t)he range of human responses that would be measured to include pleasure; (t)he circumstances in which they would be measured to include anticipated use and reflection on use” (Bevan, 2009)

has somewhat skipped some stones, as a concern for standard metrics does not mean a concern for requirements:

“(u)ser experience seems to be following in the footsteps of other fields where a focus on evaluation has preceded a concern with establishing criteria for what would be acceptable results of evaluation.” (Bevan, 2009)

It is not difficult to see that information architecture easily fits as one of these “other fields”. And certainly there are similarities and analogies in the development of the fields, as I highlighted in Chapter I, with one important difference: the debate about the scope and definition of user experience has in this context of measurability some clear landmarks in the definitions provided by ISO FDIS 9241-210: user experience is

“(a) person's perceptions and responses that result from the use and/or anticipated use of a product, system or service. ”

Nonetheless, this definition is in contrast with the revised definition of usability in the same standard, where usability is defined as the

“(e)xtent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.”

Not only that, but the definition of user experience in ISO FDIS 9241-210 states that:

“(u)ser experience includes all the users’ emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviours and accomplishments that occur before, during and after use.”

Now, if user experience includes all behaviors, it presumably includes effectiveness and efficiency, and this seems consistent with uses among the practitioners, “who appear to have subsumed usability within user experience” (Bevan, 2009), as opposed to uses among the researchers, who “consider user experience to be entirely subjective”.

Bevan conceptualizes the relationship as three possible different approaches:

1. An elaboration of the satisfaction component of usability;
2. Distinct from usability, which has a historical emphasis on user performance;
3. An umbrella term for all the user’s perceptions and responses, whether measured subjectively or objectively.

Given the scope of this research, I will characterize globally user experience as the umbrella term, as this is the stance the profession has on it (Garrett, 2002b), but it must be noted that, terminology notwithstanding, the optimization of human performance is not the same as the optimization of user satisfaction, the “achieving (of) both pragmatic and hedonic goals” (Bevan, 2009). The former largely connotes usability, the latter user experience.

Furthermore, whatever the methodologies and approaches involved, usability is largely a *post mortem* procedure. It assesses if a given application, in the broader possible sense, complies with the various ergonomic standards set forth in the ISO documents. As such, many different layers of user interaction and user expectation are left behind or difficult to account for.

“the difference in emphasis between task performance and pleasure leads to different concerns during development.” (Bevan, 2009)

User experience, on the other hand, is an evolving design procedure or methodology being applied during and on the development of an artifact (Garrett, 2002b).

Kaltenbacher (2009) maintains that one of the reasons behind this is that the traditional usability model was developed in the context of software development. Although the roots of usability also lie in the vast literature and studies on ergonomics which were a large part of design in the 1950s-1970s, Kaltenbacher correctly points out that this needs to change and that it bears relevance for the general development on information architecture for at least two reasons:

“(...) on the Internet information design and retrieval (IR) benefits from its application just as much as software development did, due its vast user base. Secondly, large parts of the Internet are application or software driven by now. At the same time, the interplay of information and applications on the Internet has produced new ways of interaction, and new demands towards the quality of interaction. Consequently, the traditional usability model needs to be expanded beyond an entirely functional focus, to accommodate the richer notion of the user experience.”

But if usability has been on the map of digital libraries for some time now, research trying to assess the overall user experience by means of a standard methodology has been scarce, and accounts basically to a single paper by Toms, Dufour, and Hesemeier (2004): I will discuss this latter and in more detail when reporting on the case-studies.

IA for Historical and Juridical Digital Archives

If user experience has been touched upon but not really embraced, and information architecture has somewhat remained outside the digital library camp for reasons that remain to be investigated, the more specialized field of historical and juridical on-line archives, the subject of this research proper, is even more scarce in terms of contributions, either requested or offered.

Partly to solve an issue of framing, and partly because it was interesting to assess advancements in connection with similar efforts, a number of digital libraries and digital archives freely available on the Web were accessed, studied, and assessed during the development of the second iteration of the Innerio Archives and during the post-launch

phase¹⁷. Some of the results of these observations are relevant to this research and will be presented here in the form of four short case-studies out of the sixteen case-studies originally evaluated at different degrees of detail during development.

Case-studies

A Methodology for IA Evaluation of Digital Libraries

I described the methodology used during the development of the Innerio Archives and largely derived from Toub (2000) in Chapter I. I will briefly recall the schematics aspects here, as these were used during the evaluation of the case-studies. The assessments were done as part of a three-steps procedure:

Early design	competitive analysis
	compare this IA to a previous IA for the same site
	place-making assessments
Later design	refine / validate an IA design
	finer competitive analysis
	place-making assessments
Post-launch	compare this IA to a previous IA for the same site

Table 4: Evaluation of the Information Architecture for the Innerio Archives

This is the general cycle to be carried out during the lifetime of a project: highlighted in light grey the competitive analysis steps in which the case-studies themselves were assessed. In each phase or step of this cycle structure, grouping, and labeling, the base

¹⁷ A comparative study was initially carried out in the Spring and Summer of 2008, during a research stay at the Department of Informatics of the Jönköping International Business School. Later parts and refinements were finalized during a three-month research period at Copenhagen Business School in the Summer of 2009.

heuristics of the information architecture evaluation carried out, can be checked and verified. In table format:

Structure	completeness, inclusion or exclusion of content and functionality, and form, including hierarchy and correlations	response to user's needs, structural design, hierarchies, and completeness and form of content objects, such as applications, documents, or pages
Grouping	coherence and integrity of categories or groups of content objects	categories, their overlapping, representativity of elements inside categories
Labeling	name or icon of a content object, such as the title of a page, or the title of a category or heading	all issues pertaining to navigation, way-finding, and place-making

Table 5: Overview of the information architecture evaluation heuristics

The methodology was used as-is, nor was it expanded later, as it fulfilled its instrumental role perfectly and was not the main focus or goal of this research. And even though they were not in the end integrated in any way, the scientific methodology for a user experience evaluation of digital libraries described by Toms, Dufour and Hesemeier (2004) was thoroughly examined for indicators. Toms' approach is decidedly broader:

“When one thinks of evaluating digital libraries, one conjures up the usual recall and precision metrics, or some notion of relevance, or the efficiency and effectiveness of the search process, or perhaps even the usability of the system. Yet, users of digital libraries much like users of physical libraries do more than look for information; they browse and meander through the virtual stacks, becoming immersed in the material, moving along a continuum from information to entertainment.”

Toms also makes the case for a clearer distinction between what is usability and what is user experience along the lines of the performance versus satisfaction indicator I cited earlier on:

“Assessing user experience is popularly equated with usability. Yet usability deals with the functionality of the system (...) and whether the system is humanly usable, rather than assessing the playfulness, enjoyment and pleasure that interacting with such a system may bring.”

This is correlated by Toms with the fact that strangely enough, we normally place an emphasis on reading for pleasure, or learning through pleasure rather than obligation, but these qualities are neither examined nor discussed in the context of digital libraries. But “not all interactivity need be utilitarian and functional” (Toms et al, 2004).

Toms, Dufour and Hesemeier found some interesting approaches to the evaluation of pleasure in marketing research and especially in the Experiential Value Scale (EVS), based on qualitative research and validated using 302 catalog and Web shoppers, and adapted it for information-rich environments. They reports that:

“(n)ot all dimensions were appropriate; we removed efficiency and economic value as both were less concerned with digital library use. Other dimensions confirmed by previous research were added. Novelty for example was found to be a key motivation in browsing. In addition, the e-commerce terminology was modified to reflect an information-rich environment.”

The resulting scale, dubbed Digital Library User Experience Scale (DLues), includes nine indicators of user experience and 22 statements measured using a seven point scale from -3 to +3 to indicate the degree of agreement with each statement. (Toms et al, 2004). The statements are:

Visual Appeal	Entertainment	Escapism
Intrinsic Enjoyment	Excellence	Site Preference
Novelty	Future Patronage Intent	Overall Recommendation.

I will not use the indicators directly, but some of the observations carried out during the assessments have clear implications that connect them to the DLues, such as layout and

structuring with visual appeal and ease of navigation with intrinsic enjoyment, and that suggest that further investigations in this direction might be necessary.

	A - Narrow View	B - Broader View	C - Broadest View
1	objects are information resources		
2		some of the objects are selected on the basis of quality	
3		objects are located in a logical place (may be distributed)	
4	objects are organized		
5	objects are subjected to authority control		
6			objects are fluid (can change and mutate at any time)
7	objects are permanent (do not disappear)		
8	authorship is an important concept		
9		access to some objects is limited to specific classes of users	
10			the only services are those performed by computer software (AI)
11			there are no librarians
12		some classes of objects have associated user groups	

Table 6: The compound table of properties for digital libraries handling complex and dynamic historical and juridical information derived from Harter (1996).

Finally, the compound version of the narrow-broader-broadest schema proposed by Harter was used as a reference to measure empirically how the digital libraries being assessed

adhere to the general model for juridical and historical DL that was shaping up: numbers and letters from 1 to 12 and from A to C, corresponding to the different properties and views, have been added to make it possible to refer to the single items in matrix like syntax.

For example, A1 is short for “objects are information resources”. In the analytical overview for the various case studies a positive match will be acknowledged by reporting the short matrix label. Negatives will be simply skipped.

Early Manuscripts at Oxford University

This web site provides access to over 80 early manuscripts in the care of institutions associated with the University of Oxford. Between 1995 and 2000 the Early Manuscripts Imaging Project created high resolution digital images from manuscripts which were selected as major treasures from their respective libraries, to create wider availability for originals which may otherwise be too fragile for handling. Some of the manuscripts are incomplete as omissions were only found out late in the acquisition process.

The screenshot shows a web page layout for a manuscript. On the left is a dark sidebar containing the text 'Early Manuscripts at Oxford University', 'Bodleian Library MS. Auct. F. 2.13', and links for 'About this manuscript' and 'Copyright Notice'. Below these are two thumbnail images labeled 'inside front board' and 'fol. i (recto)', with 'View All' and 'Next' buttons. The main content area has a white background with black text and is divided into sections by black headers: 'Bodleian Library MS. Auct. F. 2.13', 'COPYRIGHT NOTICE' (with text about copyright), 'VIEWING OF IMAGES' (with text about viewing permissions), and 'COPYING OF IMAGES FOR PERSONAL USE ONLY' (with text about downloading and copying rules).

Figure 1: Early Manuscripts DL, the Bodleian Library: manuscript landing page

In 2001 interim responsibility was taken up by the Libraries Automation Service's R&D section and the site is currently maintained by the Oxford Digital Library: the original static (X)HTML pages have been replaced with a Java Servlet Architecture which generates all web pages in real time from a single XML file using XSLT.

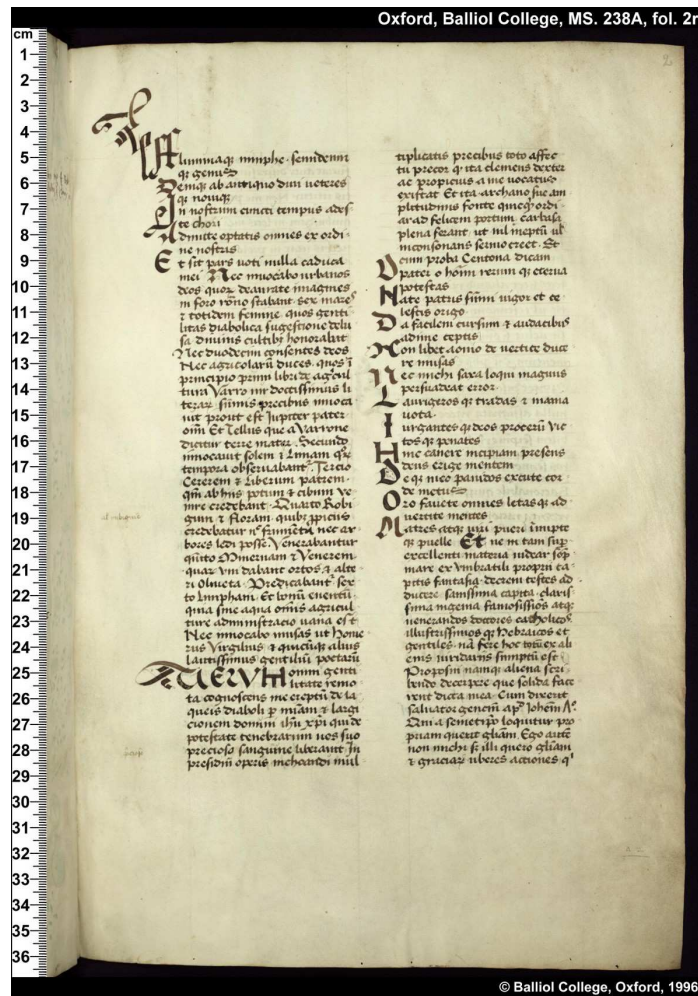


Figure 2: A sample image from the Early Manuscripts Collection. Note the measured reference on the left to assess the dimensions of the sheets on-line.

Initially, the project was part of the “Specialised Research Collections in the Humanities” initiative supported by the Higher Education Funding bodies of England, Scotland, Wales and Northern Ireland under the direction of the Non-Formula Funding Committee.

High resolution images were digitized from primary sources and then stored on the Oxford University Hierarchical File Server for general availability. The main rationale behind the choice of manuscripts to be digitized and made available on-line was the correlation between their importance and their physical fragility.

The first manuscripts to be digitized were Celtic manuscripts preserved in different Oxford libraries: these date from the 9th to the 19th century, and include Irish, Welsh, Cornish and Breton texts. The project was subsequently extended and completed in September 2000. In January 2001, the web site was completely redesigned. The Early Manuscripts at Oxford University digital archive is located at <http://image.ox.ac.uk/>.

Analysis

The Early Manuscripts at Oxford DL is simply structured in a one way tree that moves from the homepage to the manuscript pages through the selection of a particular collection and a manuscript homepage. This latter presents the possibility to navigate between pages and to access manuscript information on the left: the main body contains legal information concerning use and re-use of the images.

Access	Open
Navigation	Strictly hierarchical, from homepage to collection to manuscript to page. No correlations, no horizontal navigation. No going back from manuscript to collection
Labeling	Minimal, extremely technical
Grouping	By collection; by manuscript
Search	No search function available
Images	Available in hi-resolution; measured
Information	No data available for syndication or reuse
User services	None
URL structure	http://image.ox.ac.uk/show?

	collection=bodleian&manuscript=msautcf213
Harter compound	A1; B2; B3; A4; A5; A7; A8; C10; C11

There is no provision to go back to the list of manuscripts in a collection, as that is simply handled by going back to the homepage and by reselecting the chosen resource.

Navigation is basically either horizontal inside the manuscript, in a next-previous fashion, or can use an overview page where all pages are presented together. Users can jump to any page and then resume the next-previous navigation from there. No information apart from physical dimensioning of the digitized folio is presented on the pages.

It is possible to access all manuscript at once, but no help is provided in this case to make

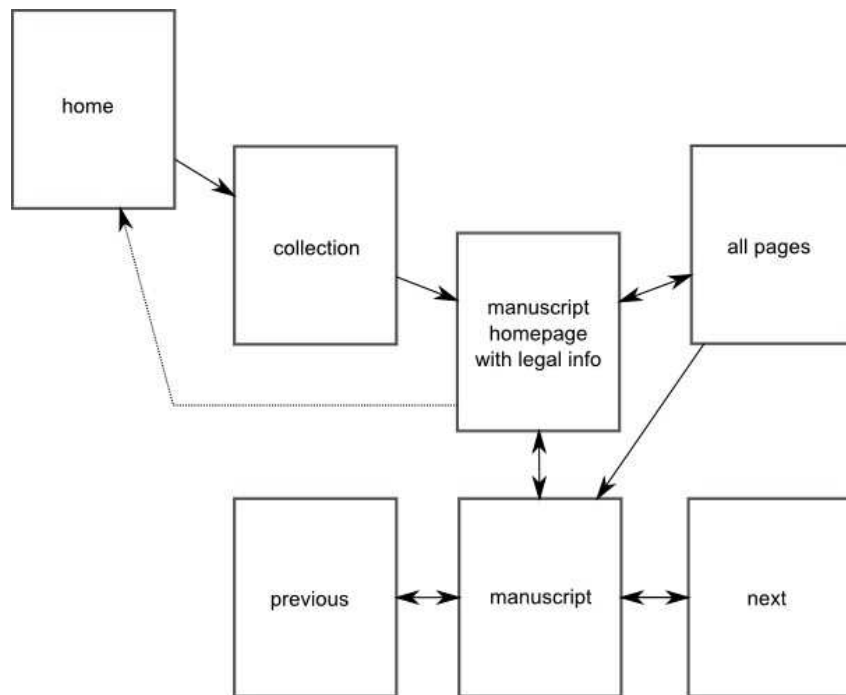


Figure 3: Hierarchy and structural navigation in the Oxford Early Manuscripts digital library

navigation easier apart from some basic labeling and sectioning, and the sheer size of the page makes it cumbersome to navigate and difficult to understand.

Laurentius

The Saint Laurentius on-line collection of medieval manuscripts at the University of Lund, Sweden, consists of 70 volumes. The nucleus of the collection comes from Lund Cathedral. Most of the manuscripts are in Latin but there are texts in Greek, Syriac, Russian, Flemish, German, French, Danish and Swedish.

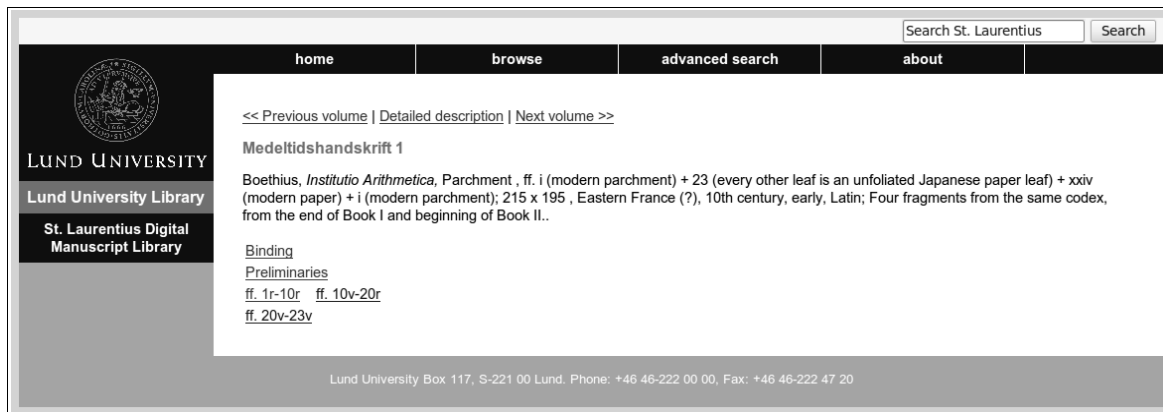


Figure 4: A brief description page for a manuscript at Saint Laurentius digital library

The oldest manuscripts in the collection are about 800 papyri from Egypt, written in Greek possibly in the 1st or 2nd century, while the oldest Nordic manuscript, *Necrologium lundense*, is a 12th century manuscript come into possession of the university library in 1671. The Codex¹⁸ Falkenberg, one of the most important Vadstena manuscripts relating to St. Bridget's revelations, is also part of Saint Laurentius DL.

The Saint Laurentius digital manuscript library can be searched and navigated remotely using the standard SRU search and retrieval protocol. SRU is a way of accessing information that focuses on the content of manuscripts¹⁹, meaning that a remote search for “prayers” will return a list of pages containing prayers besides the usual matches for the

18 A codex is a collection of text in the format used for modern books, with separate pages bound together and given a cover. The term is usually reserved for manuscript books produced from Late Antiquity through the Middle Ages. It will be used as such throughout this research.

19 SRU is an international, standard XML-focused search protocol for Internet search queries, utilizing CQL (Contextual Query Language), a standard syntax for representing queries. For more on SRU, see the Library of Congress at <http://www.loc.gov/standards/sru/>.

string “prayer” itself. The St Laurentius Digital Manuscript Library is located at <http://laurentius.ub.lu.se/>.

Analysis

The Saint Laurentius DL presents a much more complex information architecture than the Early Manuscripts. A constant navigational layer is provided via a main menu, although the *About* link takes users to a page part of the University web structure which differs from the DL in terms of layout and presentation, and that could hence result in a loss of context and confuse users.

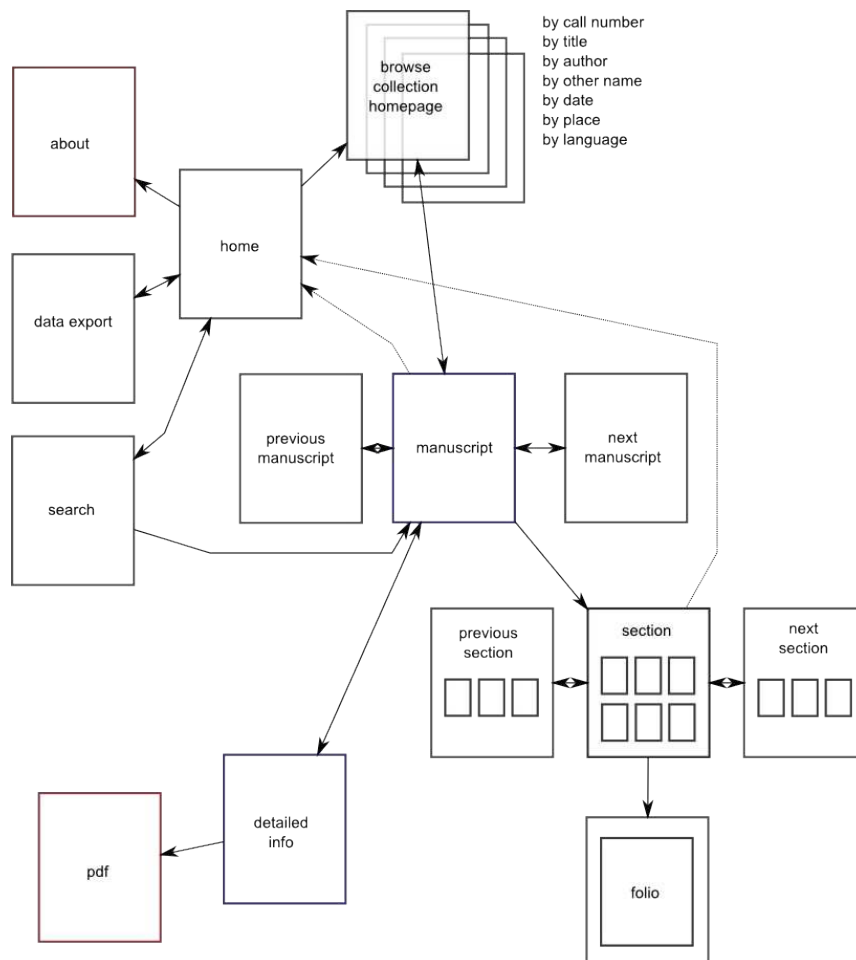



Figure 5: Saint Laurentius Digital Library structural and navigational schema



LUND UNIVERSITY
Lund University Library
St. Laurentius Digital
Manuscript Library

home
browse
advanced search
about

<< Previous | Brief description | PDF version | Next >>

Medeltidshandskrift 1
Lund University Library

Ofm: Bbl. ant. M. 64. 4:o No 11; Bbl. Ms. H. L. a) 4:o-40

Boethius, Institutio Arithmetica. Eastern France (?), 10th century, early, Latin

Contents

1 (ff. 1r-23v) Boethius, Anicius Manlius Torquatus Severinus Institutio Arithmetica
 1.1 (ff. 1r-8v) Begins defectively: "duobus secundis et tercio id est xxv." Ends defectively: "ad hunc modum infinita progressio est consequens ex ordine trianguli aequilateralis procedebuntur." Corresponding to 1.32.19-2.6.1.Ck: Köln Dombibliothek, Codices 83, 165 and 186 (<http://www.oecoc.uni-koeln.de/>).
 1.2 (ff. 9r-11v) Begins defectively: "ab his inter quos bene loquendi ratio non negatur." Ends defectively: "Quod si continuam quis faciat, cunctos huiusmodi numeros in compendioso ordine prosequatur." Corresponding to 1.32.19-2.6.1.Ck: Köln Dombibliothek, Codices 83, 165 and 186 (<http://www.oecoc.uni-koeln.de/>).
 1.3 (ff. 10r-11v) Begins defectively: "Tetragoni." Ends defectively: "Nam si omnes ab unitate impares disponantur, iuncti figuris cybeas explicabunt: I. III. V. VII. VIII. XI. XIII. XV. XVII. XVIII. XXI."
 1.4 (ff. 12r-23v) Begins defectively: "Quae mediata quibus rerum publicanum status comparantur." Ends defectively: "hucus discriptiones subter exemplar adiecimus." Corresponding to 2.45.1-2.54.9; on f. 17r there is an illustration from elsewhere in the manuscript. Following the explicit there are four illustrations.; F. 23v: blank.

Physical description

Support: Parchment

Extent: ff. i (modern parchment) + 23 (every other leaf is an unfoliated Japanese paper leaf) + xxiv (modern paper) + i (modern parchment) 215 x 195

Collation: Original collation disturbed by restoration and rebinding in 1964.

Layout: Ruled with a dry point, no trace of pricking, written in two columns throughout, each of 26 lines.

Script: Caroline minuscule, written in one hand, with the exception of ff. 22-23, rewritten in the 11th-12th century, rubricated.

Decorations

Dialectic diagrams illustrating the text, carefully traced in thin lines of red ink with the use of ruler and compasses, one full page diagram on f. 23r with decorative elements. Retracings and secondary diagrams in brown ink.

Initials in square capitals with some exceptions in uncials, elegantly drawn mostly in brown ink, contemporary with the main portion of the text. A large initial (S) begins Book II on f. 2r. Secondary initials on ff. 22r and 22v show simple floral ornamentation.

Detailed description: f. 23r: Full page decoration with a combination of four separate rubricated diagrams within a common decorated frame, illustration to II.54.9. Above left Geometrica, a diagram with a horizontal band divided into four with two concentric segments of a circle above, the outer of which connects the outermost fields, and the inner, which is divided by a vertical line, connects the two central fields, and below two intersecting segments connecting the outer fields with the central field in the opposite half of the horizontal band. Above right Arithmetica, a diagram with a horizontal band divided into four with two concentric segments of a circle above, the outer of which connects the outermost fields, and the inner, which is divided by a vertical line, connects the two central fields, and below a larger segment connecting the far left field with the central right, and a smaller segment connecting the far right field with the central right. Below left Armonica, a diagram with a horizontal band divided into two bands of which the upper is divided into four fields with two concentric segments of a circle above, the outer of which connects the outermost fields, and the inner, which is divided by a vertical line, connects the two central fields. The lower band is divided into two and two arcs connect the outer parts of the fields with the division line in the middle. Below right Consonantiae, a diagram with a vertical band divided into four with a large segment of a circle to the left connecting the outer fields and three smaller segments connecting the respectively juxtaposed fields. To the right two intersecting segments connect the outer fields with the central field in the opposite half of the vertical band. The diagrams are separated from each other by a cruciform decoration consisting of two parallel lines transgressing the frame and ending with knobbed sprouts which diverge in the vertical axis, and intersect in the horizontal axis, and a secondarily enhanced cross inscribed between these lines with knobs at the ends and in the center. Short knobbed sprouts of the same kind protrude from the corners of the frame.

In comparison with the luxurious Bamberg manuscript of De Arithmetica, Bamberg Staatsbibliothek, Class. 8 (H.J.V.12), there is a marked absence of independent artistry in Medeltidshandskrift 1. The simplicity of the diagrammatic decoration presents us with a neutral complement to the text, common also in astronomical and geometrical treatises throughout the Middle Ages, without any stylistic features connecting the decoration with the art historical currents of eastern France to which the manuscript has been ascribed on palaeographical grounds, or any other region. There is reason to believe that the attitude towards the diagrams as part of the text rather than images justified an exact copying of the exemplar instead of an interpretation in the taste of the day, which makes the only decorated illustration the more interesting.

The full page diagram on f. 23r displays characteristic protrusions in the framework. Though simple in design they are distinct enough to say that they do not convey to the notion of either Franco-Saxon style in particular or late Carolingian style as a whole, but show closer affinities to insular decoration. The spelling and abbreviations in the Lund manuscript noted by Lehmann seem to confirm the textual dependence of an insular prototype. Parallels to the frame in this manuscript are to be found in the evangelist portraits in the Trier Gospels and the Macgregol Gospels. In detail, however, insular book illumination often show a more elaborated interlace pattern, whereas the simpler kind of knobbed sprouts are found in late antique works of art in different media.

In comparison to diagrams and schemas in related works on the liberal arts by Cassiodorus and the Corpus Agrimensorum, with which De Arithmetica was often associated, there is a striking similarity in execution independent of the manuscript's date or place of origin. In the case of the Corpus Agrimensorum, which is preserved in both 6th century manuscripts and later copies, it can be stated that the Carolingian artists followed their exemplars very closely. Suggesting that the same can be suspected of the relation between Medeltidshandskrift 1 and its prototype, it is likely that its decoration represents a faithful copy ultimately derived from a 6th century exemplar without extensive alterations.

The medietary role played by the insular monasteries during the 7th and 8th centuries to promote classical learning is well known, and De Arithmetica was translated by Alfred the Great in the 9th century. Considering the convincing late antique character of the framework on f. 23r it seems plausible that an insular 6th century copy of a late antique prototype was the immediate exemplar of Medeltidshandskrift 1.

The importance of the monastery of Fleury in the dissemination of Boethius texts, and the contacts between Fleury and English monasteries during the 10th century are well documented, and preserved late Carolingian copies from its scriptorium such as e.g. the Codex Schoenbergianus have many features in common with Medeltidshandskrift 1. Hypothetically, thus, Fleury is perhaps the most likely point where different tendencies as the Carolingian script, the classical layout and the insular spelling converge.

The combination of the diagrams on f. 23r form the picturing of the "maxima et perfecta symphonia, quae tribus distenditur intervalis" which is the subject of II.54. The concept, based on Timaeus of Plato and De Caelo of Aristotle, is set in a Christian context by Boethius and becomes a parallel to the concordance of the gospels, the "evangelical harmony". The similarity of the setup of the Boethius page to the cruciform pages with all four evangelist symbols in The Book of Kells, The Trier Gospels and the Macdurnan Gospels cannot be ignored. In the light of the arian conflict at the time of Boethius, the emphasis on symphonia as opposed to differentia gives a profound theological dimension to the mathematical treatise, an idea pronounced by Boethius also in his text De Trinitate.

Binding

Blind-tooled dark brown leather binding, preserved covers northern Europe, likely Denmark or Germany, late 15th or early 16th century. Rebinding signed by Hans Helland Germany, Stuttgart 1964. Size: 232 x 220 x 34 mm.

Dark brown leather binding over thick square edged wooden boards. Rebinding with covers from former binding preserved. One engraved metal hook-clasp fastening at fore-edge, new nails, rivets and strap. Tight back with four double raised bands. Trimmed edges. Single flyleaves and separate pastedowns of parchment. A preserved piece of endpaper with former shelf marks (cf. above) mounted on front pastedown. All along sewing on four double cords laced through the boards. Two thick additional quires of Japanese paper at the end of the textblock. Textblock restored and foliated with Japanese paper. Photographic documentation of former binding and description of the rebinding mounted on pastedown of the lower cover.

The covers are blind-tooled with a triple line fillet border and panel decorated with rosettes, a lozenge and a floral tool. The panel is divided into ogival diapers by a repeated panel-stamped pomegranate ornament.

Foliation

The parchment leaves are foliated in pencil by modern catalogue in upper right corner; the Japanese paper leaves are un-foliated.

History

Written in Eastern France in the early 10th century.

Provenance

The manuscript appears to have been the property of the library of Lund Cathedral (cf. former shelfmarks).

Acquisition

The manuscript came to the library in 1708 from the estate of professor Andreas Röddemarck (1651-1707).

Bibliography

- Boethius: Boece, Institution arithmétique Gullaumin, Jean-Yves (ed.) Paris, 1995, pp. toxi. Medeltidshandskrift 1 is Gullaumin's nr 73.
- Boethius: Anici Manli Severini Boethii De Arithmetica Corpus Christianorum. Series Latina cura et studio Henrici Oosthout et Iohannis Schilling (ed.) Turnhout - Belgium, 1999. 94 A. Anici Manli Severini Boethii Opera. Pars II. p. 103-120 (esp. p. 106).
- Lehmann, P.: Skandinavische Reise Früchte. 1. Nachlese 1-2 Nordisk tidsskrift for bok- och biblioteksväsen 1937, 24 pp. 103-120 (esp. p. 106).
- Paulson, J.: De fragmento Lundenis Boetii de institutione arithmetica librorum Lunds universitets ärsskrift Lundae, 1885. 21 pp. 1-30.
- Pellegrin, E.: Manuscrits d'auteurs latins de l'époque classique conservés dans les bibliothèques publiques de Suède Bulletin d'information de l'Institut de recherches et d'histoire des textes Paris, 1964. 3 pp. 7-32 (esp. pp. 28-31).

Lund University Box 117, S-221 00 Lund, Phone: +46 40-222 00 00, Fax: +46 40-222 47 20

Figure 6: Saint Laurentius, a manuscript's detailed description page

A second-level navigation horizontally moves between manuscripts. A third-level navigation allows users to move between pages pertaining to the same manuscript. Since Laurentius hosts a single collection, there is no added layer in moving from the homepage to the manuscripts.

Users clicking *Browse* are moved to a page where they are provided with a default alphabetical and rather long list of manuscripts, as no paging is provided. By using a number of different criteria, such as author, place, date, or title, users can reorder the manuscripts. It is important to point out though that whatever the reordering, the second-level navigation between manuscripts always follows alphabetical (or call number as it is called) order, which might be confusing.

Although no explicit horizontal structure exists, as it is not possible to correlate authors, or manuscripts, or works, navigation is more structurally sound than the one found in the Early Manuscripts DL. An internal content-oriented (as explained above) search engine is provided that allows simple boolean AND chaining of free-text entries.

Manuscripts provide info via either a terse synthetic view, or an extremely verbose detailed description (Figure x). Access to single pages in a manuscript is provided via a section view that divides the manuscripts in batches (Figure y). Navigation across batches is provided via top-page links (previous / next) and seems to be at times totally arbitrary (as in the Medeltidshandskrift 2 (Lectionarium Lundense I)).

URLs within the digital library are somewhat built in a human-readable format that helps recognition and ties them to the call names of the manuscripts, but all images are served via a cache from a third-party server within the university network resulting in totally different URLs.

Users can both download the complete catalog or the detailed descriptions of single manuscripts as PDF files. Images can be downloaded at medium-high resolution (1200px), but no feature exists to obtain a full manuscript from start to finish with a single download. The full-size images of the codices are neither measured nor color-coded.

Access	Open
Navigation	Via a main menu (browse, search), and a local filtering menu (by call number, name, title, author, place, date, language). No navigation is possible when single images are loaded
Labeling	Technical, not always coherent ('download catalog' or 'PDF' both provide access to downloadable content)
Grouping	By collection; by manuscript; by section (of pages)
Search	Free-text, multi-field; Compound, boolean AND
Images	Available in medium-hi-resolution (1200px)
Information	Syndication and remote reuse via SRU
User services	PDF downloads (catalog, manuscripts)
URL structure	http://laurentius.ub.lu.se/volumes/Mh_1/preliminaries.html
Harter compound	A1; B2; B3; A4; A5; A7; A8; C10; C11

The British Library Catalogue of Illuminated Manuscripts

The Online Catalogue of Illuminated Manuscripts holds a large collection of Western manuscripts from the Middle Ages and the Renaissance and “and aims to provide access to images and information about its manuscripts to students, scholars, and the general public”. The Online Catalogue holds very specific content as well, illuminated manuscripts, or manuscripts in which the text is supplemented by the drawings and decorations, such as glyphs, marginalia, and miniatures²⁰.

The general goal of the Catalogue is to include around 9000 Western manuscripts upon completion of the project, thus including all illuminated British Library codices up to the 17th century, providing “at least one image from each manuscript”.

The Online Catalogue lives at <http://www.bl.uk/catalogues/illuminatedmanuscripts/>.

²⁰ Originally the definition of illuminated manuscript was strictly applied to manuscripts decorated in gold and silver only, but it is today widely used for decorated manuscripts regardless of colors and materials.

Analysis

The Catalogue offers four different searches: a quick search, a simple search, a manuscript search, and an advanced search. Unfortunately, none of these offers any help to the user by means of suggest-while-you-type or type-ahead mechanisms, so that, unless we are in a know-item approach, getting the desired results might be problematic.

BRITISH LIBRARY

CATALOGUE OF ILLUMINATED MANUSCRIPTS

About | Simple search | Manuscript search | Advanced search | Tours | Glossary | Contact us

Use this website to find and view descriptions and images of medieval and Renaissance manuscripts in the British Library, one of the richest collections in the world.

Quick search:

GO

You can

- perform a quick search (this searches for a word or number in all sections of each catalogue entry, including images);
- perform a [simple search](#) using keywords and dates;
- look for information about a [particular manuscript](#) if you know its collection name and manuscript number;
- perform an [advanced search](#) using different types or combinations of information;
- take a thematic [tour](#) of various aspects of the British Library's western illuminated manuscript holdings;
- browse or search an illustrated [glossary](#) of terms; and
- read about the [scope and history](#) of this catalogue.

Updated 28 September 2009. Please note that cataloguing of manuscripts in the Royal collection is in progress, and that manuscripts in the Cotton and Additional collections are not yet included in the Catalogue.

For details of how to use the system, see our [search tips](#).

Supported by grants from the Arts & Humanities Research Council and the Getty Foundation.

[About](#) [Simple search](#) [Manuscript search](#) [Advanced search](#)
[Tours](#) [Glossary](#) [Contact us](#) [Copyright](#) © The British Library

Figure 7: British Library Catalogue of Illuminated Manuscripts, homepage

Both quick search and simple search, for example, return 3725 paged results for “latin” . But even more problematic might be the use of exact matches in search: while looking for a specific manuscript by Attaleiotos, typing “Attaleiotos” correctly returns one manuscript (Egerton 266, Opus de Jure, or Pragmatica), but typing “Attaleio” returns zero.

Nonetheless, the Catalogue is a search-oriented digital library, with scarce cure devoted to navigation, correlation, or horizontal navigation. Although a main menu is always visible,

there is no direct access to manuscripts if not through search, to the point that the only way to browse the content of one collection is to use manuscript search and leave all fields empty. This produces a collection view containing small-scale images and text, from which users can choose to enter manuscript view, which provides them with information on the chosen volume, and clickable previews of the available pages. The Catalogue does not offer horizontal navigation of manuscripts at all: all pages pertaining to a single codex can only be reached from the manuscript homepage.

Correlation between the different manuscripts or collections is likewise totally absent.

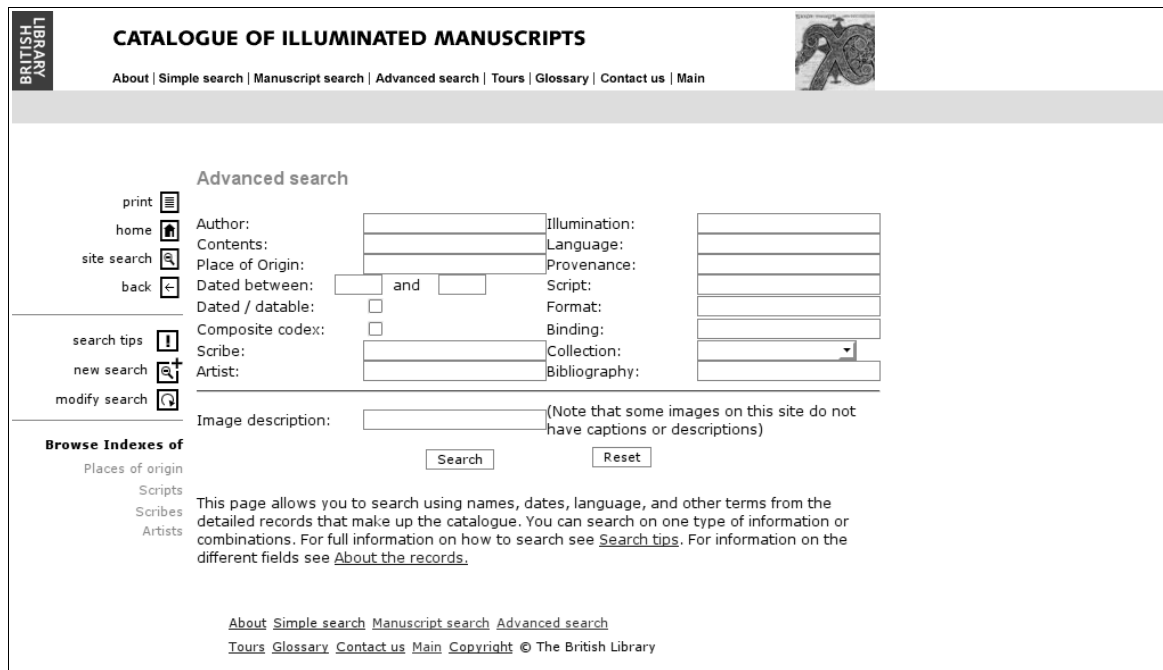


Figure 8: British Library Catalogue of Illuminated Manuscripts, advanced search page

Navigation of the Catalogue is therefore rather difficult and cumbersome to master, and requires a number of movements up and down the single tree that goes from search to manuscript page to image, and back to search.

Furthermore, the navigational buttons visible in Figure 8, advanced search, on the left, are actually part of a super-global navigation that can confuse users: the first four buttons from the top move the user back to or operate the parent web-site, the British Library.

BRITISH LIBRARY

CATALOGUE OF ILLUMINATED MANUSCRIPTS

About | Simple search | Manuscript search | Advanced search | Tours | Glossary | Contact us | Main

print

home

site search

back

search tips

new search

modify search

Latest results list
Manuscripts Online
Catalogue

Detailed record for Arundel 7

Author	Valerius Maximus
Title	Facta et dicta memorabilia, with glosses
Origin	Italy, N. (Lombardy?)
Date	2nd half of the 14th century
Language	Latin
Script	Gothic, written above top line
Decoration	1 large historiated initial, in colours and gold (f. 1). 9 large decorated initials, in colours and gold, at the beginning of the remaining books (ff. 10v, 20, 28v, 37, 46, 53v, 60v, 68v, 77), the first one with a three-sided foliate border. Smaller initials in red with blue pen-flourishing or in blue with red pen-flourishing. Small initials in plain red or blue.
Dimensions in mm	320 x 210 (215 x 110)
Official foliation	ff. 78 (+ 2 modern paper and 1 medieval parchment flyleaves at the beginning, and 1 medieval parchment and 2 modern paper flyleaves at the end)
Form	Parchment codex
Binding	BM/BL in-house.
Provenance	Unidentified owner, last quarter of the 14th century: with his arms azure, two cows proper on a fess or, below three fleurs-de-lys or (f. 1). Willibald Pirckheimer (b. 1470, d. 1530), son of Johann Pirckheimer, Nuremberg patrician, diplomat, and humanist: his armorial bookplate, with a birchtree, the arms of Pirckheimer, and a crowned mermaid with two tails, the arms of his wife, Crescentia Rieter (inside upper cover). ? Thomas Howard (b. 1585, d. 1646), 2nd earl of Arundel, 4th earl of Surrey, and 1st earl of Norfolk, art collector and politician. Henry Howard (b. 1628, d. 1684), 6th duke of Norfolk, presented to the Royal Society in 1667. The Royal Society, London (its ink stamp: 'Soc. Reg. Lond / ex dono HENR. HOWARD / Norfolciensis.', f. 1v; its book-plate, inside upper cover). Purchased by the British Museum from the Royal Society of London together with 549 other Arundel manuscripts in 1831.
Notes	At the end of the manuscript, epigram and the name of M. P(etrus) de Mullo who died in 1382 (f. 78).
Select bibliography	<p><i>Catalogue of Manuscripts in The British Museum, New Series</i>, 1 vol. in 2 parts (London: British Museum, 1834-1840), 1, part 1: <i>The Arundel Manuscripts</i>, p. 2.</p> <p>Alessandro Palma di Cesnole, <i>Catalogo di manoscritti italiani esistenti nel Museo Britannico di Londra</i> (Turin: Roux, 1890), p. 17 no. 247.</p> <p>Rino Avesani, "Epaeneticorum ad Plim II Pont. Max. libri V in <i>Enea Silvio Piccolomini Papa Pio II : Atti del convegno per il quinto centenario della morte e altri scritti raccolti da Domenico Maffei</i> (Siena : Accademia senese degli Intronati, 1968), pp. 15-98 (p. 61 n. 162.)</p> <p>Niklas Holzberg, <i>Willibald Pirckheimer: Griechischer Humanismus in Deutschland</i>, Humanistische Bibliothek, ed. by Ernesto Grassi, series I: <i>Abhandlungen</i>, vol. 41 (Munich: Fink, 1981), p. 39.</p> <p>Dorothy M. Schullian, <i>A revised list of Manuscripts of Valerius Maximus</i> (Padua: Antenore, 1981), 695-728 (p. 708).</p> <p>Paul Oskar Kristeller, <i>Iter Italicum: Accedunt Alia Itinera: A Finding List of Uncatalogued or Incompletely Catalogued Humanistic Manuscripts of the Renaissance in Italian and other Libraries</i> 7 vols (London: Warburg Institute; Leiden: Brill, 1963-1997), IV (1989), 126.</p> <p><i>The Libraries of King Henry VIII</i>, ed. by James P. Carley, <i>Corpus of British Medieval Library Catalogues</i>, 7 (London: British Library, 2000), p. 180 no. [H2]997.</p>

Last revised: 02 December 2005

Images

f. 1
Rome

f. 1
Rome

f. 60v
Decorated initial

f. 60v
Decorated initial

print

home

site search

back

top

[About](#) | [Simple search](#) | [Manuscript search](#) | [Advanced search](#)

[Tours](#) | [Glossary](#) | [Contact us](#) | [Main](#) | [Copyright](#) © The British Library

Figure 9: British Library Catalogue of Illuminated Manuscripts, manuscript's detailed description page, with folios.

85

The catalogue offers a fully-illustrated glossary of terms, ordered and paged alphabetically, with no internal search function. Finally and curiously enough, when staying on the homepage the misnamed link *Main* is correctly inactive and cannot be clicked²¹, but this is the only time this behavior occurs, as all other menu items remain clickable all the time.

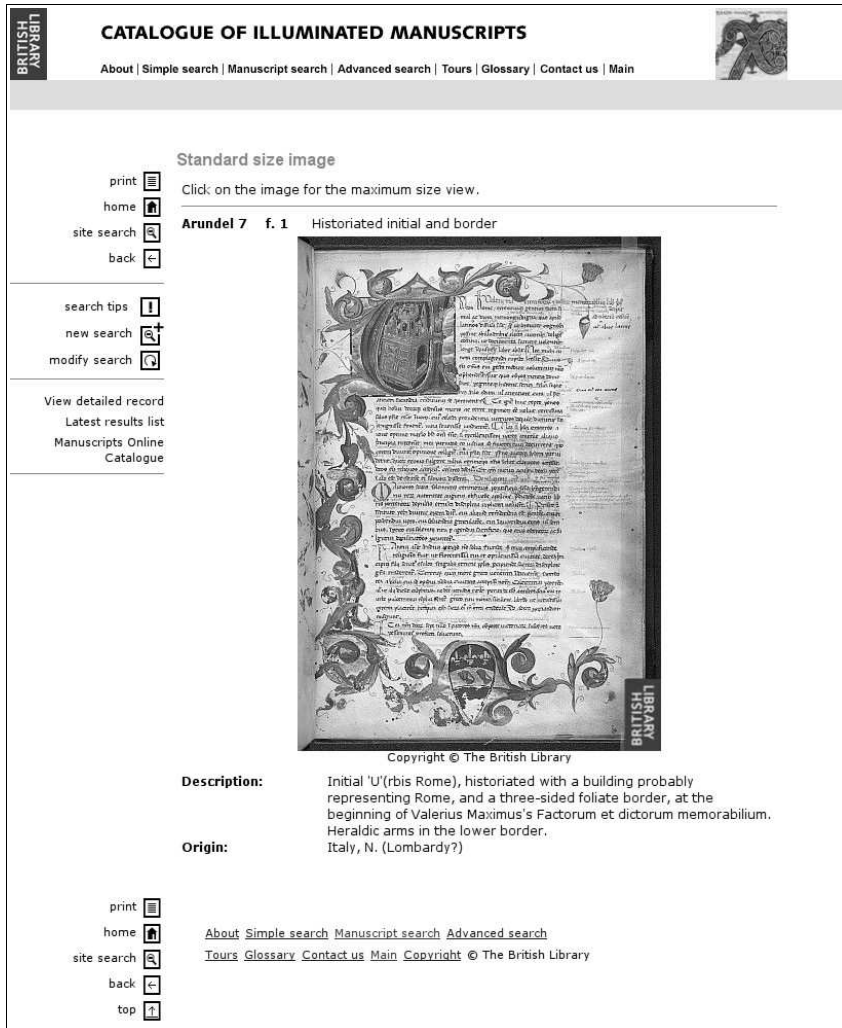


Figure 10: British Library Catalogue of Illuminated Manuscripts, a folio.

Images in the Catalogue are downloadable at medium-high resolution (1600px), and no particular user-related function is in place.

21 Thus obeying to the heuristic principle of offering no confusing link to self (Nielsen, 1993).

Access	Open
Navigation	Based on search. From manuscript to single pages
Labeling	Minimal, sometimes erroneous or misleading (Main for Home)
Grouping	By collection
Search	Free-text, multi-field; compound; no type-ahead
Images	Available in medium-hi-resolution (1600px)
Information	No data export
User services	Limited memory function (Last search)
URL structure	http://www.bl.uk/catalogues/illuminatedmanuscripts/record.asp?MSID=1640&CollID=20&NStart=7
Harter compound	A1; B2; B3; A4; A5; A7; A8; C10; C11

Manuscriptorium

Manuscriptorium is a digital library created within the scope of the ENRICH project²² with the goal to provide “seamless access to information about the vast collections of manuscripts and incunables distributed across major European libraries” (ENRICH, 2007).

Manuscriptorium targets “existing digital documents in the sphere of historic book resources (manuscripts, incunabula, early printed books, maps, charters and other types of documents)” and is particularly important to a discussion of the information architecture of existing implementations as it is directly tied to TEI P5 ENRICH and to the ENRICH group. In a certain sense, it should be a best practice implementation.

Manuscriptorium currently provides seamless access to more than 5 million digital images, and “is designed for easy searching and viewing of documents, enabling the creation of individual collections and virtual documents”.

22 ENRICH and TEI P5 ENRICH will be discussed in detail in Chapter IV.

Analysis

Two main navigational modes are provided, search and index. The latter offers simply a way to add pre-inserted keywords into the search field, and these are of little way outside of a know-item scenario, as the keywords are in such formats as “00”, “000”, “0000”, “000001”, “000002”, and so on.

Results from a search are paginated, but grouping is not easily ascertained as well because of the way results are presented (linearly, without any group description), and because of inconsistencies in the way strings are matched by the search engine. For example, searching for “000005” returns fifty matches, two of which are items from Biblioteca Nacional de España in Madrid labeled “INC/5” while the rest match exactly; searching for “00005” returns also records such as “INC/52” and “000052” but no “000005”; searching for “5” returns more than 80000 records with no further possibility to filter the records. The index function seems to be contextual, as when searching for example for “biblia latina” the listing moves to keywords connected with the root “latin”, but even so, use of the function and consistent results are difficult to obtain.

Access	Requires free registration
Navigation	Based on search. From manuscript to single pages
Labeling	Minimal, extremely technical at the main level
Grouping	By search terms
Search	Free-text; pre-entered keywords
Images	Available in high resolution (2400px)
Information	No data export; TEI P5 based
User services	Login; personal collections
URL structure	http://beta.manuscriptorium.com/apps/main/en/index.php?request=show_tei_digidoc&virtnum=0&client=
Harter compound	A1; B2; B3; A4; A5; A7; A8; C10; C11

As there is no effective navigation between collections or manuscripts, everything flows in a hierarchical way from search to list to manuscript and from here to pages.

Homepage » Catalogue » Search Results
 Search results: 382 records

◀ 1 2 3 4 ... 20 ▶▶

Biblia latina | Biblia latina
 Česká republika; Praha; Knihovna Akademie věd České republiky; 1 TB 3
 Add to Selection | Facsimile | Document description

Biblia latina | Perikopy "de tempore" | (perikopy) In dedicacione Ecclesie ... Epistole et ewangelia de sanctis | perikopy Epistole et ewangelia de Commune sanctorum, et primo Epistole de Appostolis ...
Jeronymus
 Česko; Brno; Moravská zemská knihovna v Brně; Mk 41
 Add to Selection | Facsimile | Document description

Lokální katalog pražské Universitní knihovny, 3. část | Signatura XXIV | Signatura XXV | Signatura XXVI ...
 Česko; Praha; Národní knihovna České republiky; IX.A.12c
 Add to Selection | Facsimile | Document description

Figure 11: A section of the search results page, Manuscriptorium

On the search results page (Figure 11) single manuscripts offer the possibility to be added to a personal selection, to actually go see the manuscript (the oddly named *Facsimile* entry), and to get a lengthy and extremely technical document description (Figure 12). No possibility to export this data is provided along, and no possibility to see related manuscripts or to move back to an upper level exists at this stage.

When accessing the *Facsimile* link, a new window or browser tab is opened²³ and a rich interface is loaded which allows users to choose pages, zoom in and out, and navigate (Figures 13, 14, and 15). While the pages can be examined in stunning detail thanks to the high resolution images provided (2400px), navigation is somewhat limited and hindered by the interface itself.

23 Incidentally, this is something that might confuse users as it breaks the Back button and that is actually considered malpractice in respect to accessibility of on-line content.. For an in-depth analysis of this specific point, see also Pilgrim, M. (2002). Dive Into Accessibility, http://diveintoaccessibility.org/day_16_not_opening_new_windows.html.

The screenshot displays a detailed record for a manuscript collection. At the top, there are navigation links and a search bar. The main content area is divided into several sections:

- General Information:** Includes the title 'Manuscriptorium', a description 'Document 1 / 362 f. folio - description', and a list of authors: 'Biblia Latina', 'Biblia Latina', 'Dietrich von Aino', 'König von Böhmen und Tschechien', '1703', and 'König von Böhmen und Tschechien'.
- Physical Description:** Lists 'Country: Česká republika', 'Settlement: Praha', 'Repository: Archivna Akademie věd České republiky', and 'Edition: 1703'.
- Object Description:** Includes 'Form: other', 'Material: perg', and 'Support: pergament'.
- Notes:** Lists 'Type: dimensions' and '1 x 434 (incl. 435) x 1 f., f., 362 sig'.
- Decorated Description:** A long list of 435 numbered items, such as '90791a: Figurální incipit F, keramičtými písmy Pustulov', '11x11 cm: Figurální incipit L, v ní upravená Jana Nepomácká a švédské slovo', '27x27 cm: Figurální incipit H, v ní Jáchym a Izabela', '40x40 cm: Figurální incipit L, v ní Rogier', '57x57 cm: Figurální incipit H, v ní Rogier a Izabela', '80x80 cm: Omamovací incipit E', '75x75 cm: Omamovací incipit P', '82x82 cm: Omamovací incipit I', '84x84 cm: Omamovací incipit F', '94x94 cm: Omamovací incipit G', '103x103 cm: Omamovací incipit E', '113x113 cm: Omamovací incipit P', '123x123 cm: Omamovací incipit A', '130x130 cm: Omamovací incipit C', '141x141 cm: Omamovací incipit V', '146x146 cm: Omamovací incipit V', '152x152 cm: Omamovací incipit L', '153x153 cm: Omamovací incipit A', '155x155 cm: Omamovací incipit E', '160x160 cm: Omamovací incipit T', '163x163 cm: Omamovací incipit A', '167x167 cm: Omamovací incipit I', '172x172 cm: Omamovací incipit V', '180x180 cm: Omamovací incipit B', '180x180 cm: Omamovací incipit D', '180x180 cm: Omamovací incipit D', '191x191 cm: Omamovací incipit D', '194x194 cm: Omamovací incipit S', '198x198 cm: Omamovací incipit E', '201x201 cm: Omamovací incipit C', '205x205 cm: Omamovací incipit D', '213x213 cm: Omamovací incipit P', '219x219 cm: Omamovací incipit V', '222x222 cm: Omamovací incipit D', '223x223 cm: Omamovací incipit D', '288x288 cm: Omamovací incipit D', '295x295 cm: Omamovací incipit V', '292x292 cm: Omamovací incipit E', '296x296 cm: Omamovací incipit E (2x)', '296x296 cm: Omamovací incipit D', '291x291 cm: Omamovací incipit A', '297x297 cm: Omamovací incipit V', '299x299 cm: Omamovací incipit V', '300x300 cm: Omamovací incipit V', '300x300 cm: Omamovací incipit B', '303x303 cm: Omamovací incipit E', '303x303 cm: Omamovací incipit D', '306x306 cm: Omamovací incipit D', '306x306 cm: Omamovací incipit V', '307x307 cm: Omamovací incipit I', '308x308 cm: Omamovací incipit I', '310x310 cm: Omamovací incipit D', '312x312 cm: Omamovací incipit E', '311x311 cm: Omamovací incipit P', '320x320 cm: Omamovací incipit H', '320x320 cm: Omamovací incipit L', '328x328 cm: Omamovací incipit H', '330x330 cm: Omamovací incipit I', '340x340 cm: Omamovací incipit F', '345x345 cm: Omamovací incipit Q', '349x349 cm: Omamovací incipit F', '350x350 cm: Omamovací incipit I s roztřísnými prvky', '363x363 cm: Omamovací incipit P', '363x363 cm: Omamovací incipit P', '371x371 cm: Omamovací incipit P', '372x372 cm: Omamovací incipit P', '372x372 cm: Omamovací incipit P', '372x372 cm: Omamovací incipit P', '372x372 cm: Omamovací incipit P', '376x376 cm: Omamovací incipit P', '376x376 cm: Omamovací incipit P', '376x376 cm: Omamovací incipit P', '380x380 cm: Omamovací incipit P', '381x381 cm: Omamovací incipit P', '381x381 cm: Omamovací incipit P', '384x384 cm: Omamovací incipit H', '384x384 cm: Omamovací incipit P', '384x384 cm: Omamovací incipit I s roztřísnými prvky', '392x392 cm: Omamovací incipit P', '392x392 cm: Omamovací incipit S', '396x396 cm: Omamovací incipit Q', '397x397 cm: Omamovací incipit S (2x)', '398x398 cm: Omamovací incipit I', '398x398 cm: Omamovací incipit I', '398x398 cm: Omamovací incipit A', '403x403 cm: Omamovací incipit A', '403x403 cm: Omamovací incipit B', '406x406 cm: Omamovací incipit C', '410x410 cm: Omamovací incipit D', '410x410 cm: Omamovací incipit E', '412x412 cm: Omamovací incipit F', '412x412 cm: Omamovací incipit G', '413x413 cm: Omamovací incipit H', '414x414 cm: Omamovací incipit I', '415x415 cm: Omamovací incipit L', '415x415 cm: Omamovací incipit M', '417x417 cm: Omamovací incipit N', '418x418 cm: Omamovací incipit O', '418x418 cm: Omamovací incipit P', '418x418 cm: Omamovací incipit B', '418x418 cm: Omamovací incipit S', '421x421 cm: Omamovací incipit T', '422x422 cm: Omamovací incipit V', '422x422 cm: Omamovací incipit Z'.

At the bottom, there are sections for 'Administrative Information' and 'Acquisition History', with a date 'Popis vytvořen: 12.5.2004'.

Figure 12: The detailed listing for a manuscript, Manuscriptorium. Extremely long and with no direct links to relevant items.

For example, it is not entirely clear that the manuscript in Fig. 13 effectively has more than the six pages currently viewable in the portlet on the right. There is no way to know that until one clicks on any other page preview, as then the interface moves the previews a number of pages further, similarly to a slide-show and opening up new navigation (Fig. 14). No sign or hint is provided to the user that the manuscript contains more pages though before: no arrows, no signs, no paging.

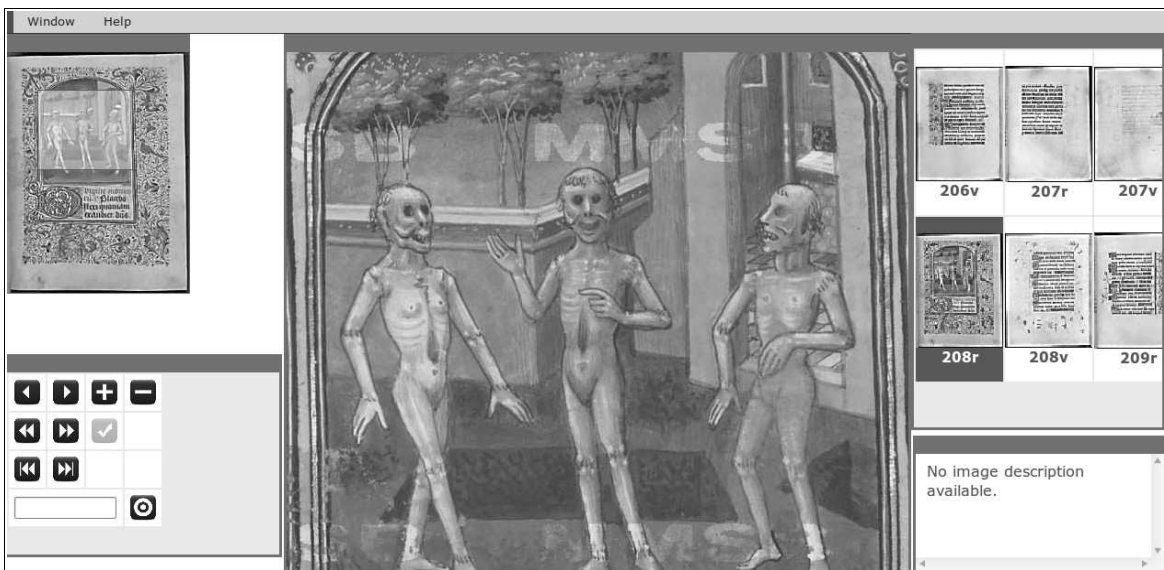


Figure 13: The rich manuscript interface at Manuscriptorium. This manuscript seems to have just six pages available for browsing (right)

Coupled with the fact that any place-making information is lost when navigating the rich interface, as not even the name of the manuscript is visible anywhere except that on the browser's window title, this is potentially extremely confusing for users.

Labels are not provided: each piece of the five-portlet interface needs to be figured out by the users themselves, some of them might be even exceedingly familiar with manuscripts but not with graphical or web interfaces. Furthermore, users who did not notice for any reason, either technical or cognitive, that a new window or tab has been opened with the rich interface view will find themselves closing it over and over again to restart searching from the homepage, a frustrating process.

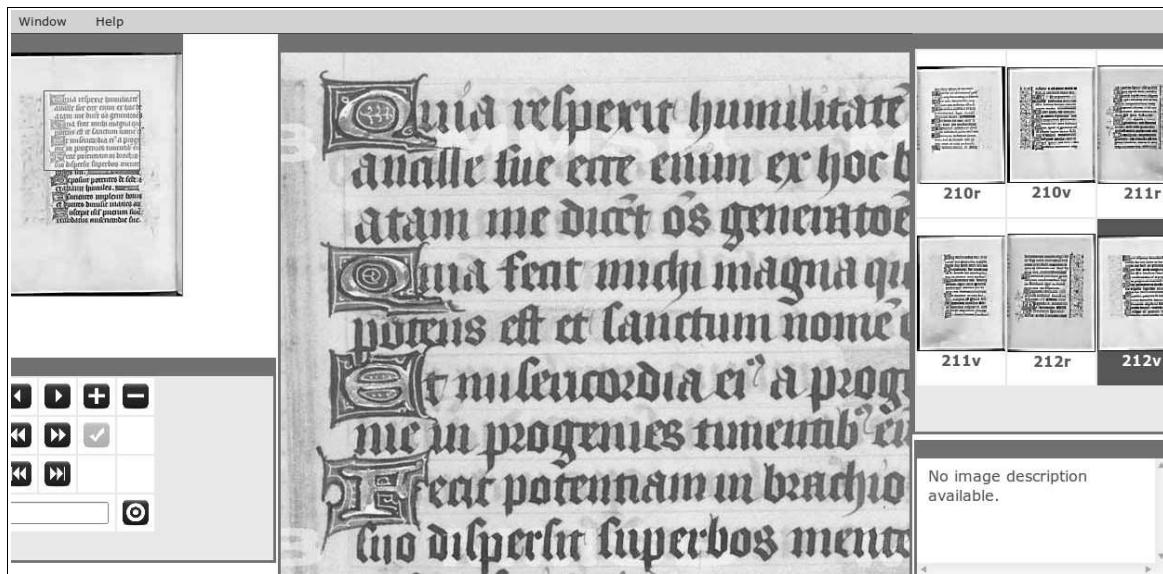


Figure 14: The same manuscript from Figure 13, showing more pages (right)

Download of manuscript images, either in low or high resolution, is possible via the “Save image” browser menu function. No clear legal info is available on the manuscript pages concerning use or re-use of such images.

An Issue of Focus and Purpose

“All the inhabitants of the Earth will be brought into one intellectual neighborhood” - Alonzo Jackman, 1846, praising the telegraph²⁴

Being there is not enough. Connecting, correlating, is not enough. The lesson that emerges from the various digital libraries and digital archives that were briefly assessed in the case studies is that there is no common understanding, besides the very basics, that the digital library is both something different from an on-line storage place and from a traditional library: specific heuristic and metrics should be tested and addressed. And because of the nature of these efforts, if usability concerns are somewhat neglected, information architecture principles such as presenting the users with a coherent mental map of the library, or offering good navigation, are totally absent.

²⁴ Cited in Spigel, L. (2001) Welcome to the Dreamhouse. Duke University Press. P. 34.

Scholars have been questioning the fact that this might mean some lack of purpose (Levy, 2000). While it might seem counterintuitive to think about this possibility when we are still discussing consolidation, when focus seems to be the larger problem at hand, it must be said that this is not a preposterous question.

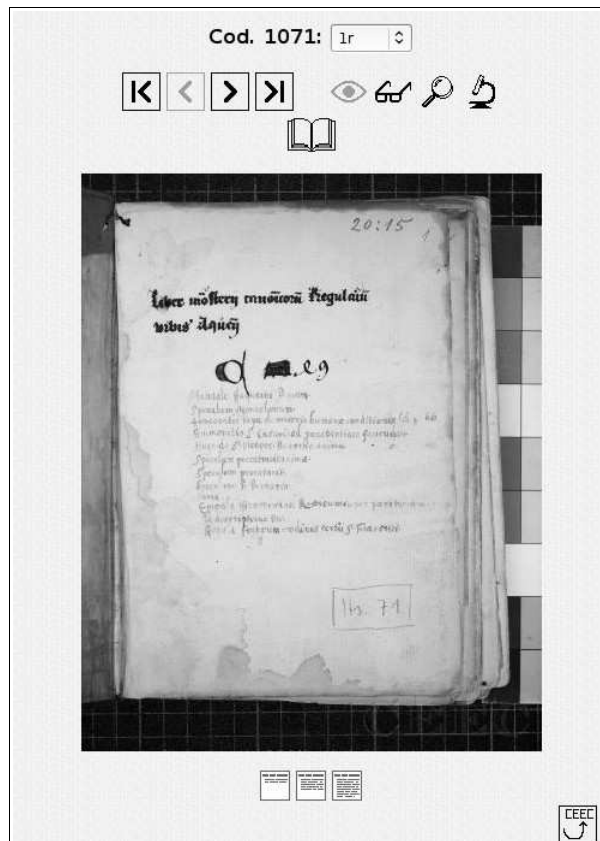


Figure 15: A page with color correspondences from CEEC, the Codices Electronici Ecclesia Coloniensis

It is actually fairly easy to see how some of the properties and issues Harter introduces in his narrow-broader-broadest model mean serious disruption if taken to the extreme consequences. They challenge the very meaning we associate with concepts such as authority, ownership, correlation, and causality. If the World Wide Web is a digital library, nothing is a digital library. Focus is connected to purpose. It is also fairly easy to see that digital libraries are still being defined today, so impacting on their development is still a possibility.

This research tries to address some of the problems connected to purpose and focus in digital libraries by pointing out a few inadequacies in the way historical and juridical information is currently represented on-line and by suggesting a few steps in what seems a reasonable direction with the help of information-savvy practices such as information architecture. Historical and juridical digital libraries might be a narrow subset of the area, but as such allow designers to design effectively: by providing focus, they allow a clearer view on the conceptual or methodological tools that actually make that focus enactable. Levy (2000) maintains that:

“(a)t first, it might seem that purpose is simply irrelevant to digital libraries. Especially if we think of them as enabling technologies with which to create and manage collections of digital materials, then digital libraries would seem to be purpose-neutral. But this simply shifts the problem of purpose from the technologies, per se, to the collections and the institutions which manage them.”

This is the reason why the role of information architecture is pivotal. Digital libraries can not be all things to all people: focus is important, and with focus comes purpose, and this purpose need to be contextual, specific, humane, and not simply driven by technology or development for the sake of development:

“I would argue that researchers and funders currently have a shared purpose, a mission, which stretches from research and development through to the greater good of humankind, and provides psychic energy and justification for the current research agenda. This mission, which might be called “the digital library faith,” is an almost evangelical belief that digital materials are right and good.” (Levy, 2000)

Chapter III

The Innerio Archives

“People are beginning to understand that they can tell their stories on the web. They can self-publish; they can say what's important; they can show pictures of their pets and tell their war stories and remember Grandpa. What's happening is that an alternative historical record is being built up. The ephemera of daily life is being preserved as never before. The Web, I realize, is a deeply conservative medium. It looks to the past, not to the future. It is a natural medium for memory.”

– Jon Carroll¹

“I have always imagined that Paradise will be a kind of library.”

– J. L. Borges

An Overview

The Innerio Archives is an on-line digital archive publishing 289 medieval legal, theological and philosophical manuscript codices part of the rare-book collection of the Reale Collegio di Spagna, the Royal College of Spain², in Bologna, and comprising the most important books of law compiled in Bologna, Italy, during the Middle Ages. As Pieri (2002) reports,

“(t)he work done beginning in the 12th century by the jurists active in Bologna (...), namely, the Roman and Canon law they put together, makes up to this day the foundation of law in Europe, a law imbued with concepts - such as 'aequitas', 'iustitia', 'genus et species', and 'ius' - occurring in philosophical as well as in doctrinal-juridical thought.”

1 Carroll, J. (1997). New Thoughts on the Last New Thing. San Francisco Chronicle.

2 The College of Spain was built in 1364 at the initiative of Cardinal Gil de Albornoz as a dormitory for Spanish students in Bologna, whose university was already a well-know center in the study of canon law, theology, and medicine (Pieri, 2002). It is by far the oldest Spanish institution anywhere, and as of today it still continues to accommodate, free of charge, some fifteen *becarios*, students attending different schools at the university in Bologna and who in Spain are traditionally referred to as *los bolonios*.

The project, named after the *primus illuminator* in Bolognese legal science, the jurist Irnerius, was headed by CIRSIFID, the Interdepartmental Center for History of Law, Philosophy and Sociology of Law, and Legal Informatics at the University of Bologna in 1999. Fundings came initially from the Cassa di Risparmio Foundation in Bologna and from the Italian Ministry of Higher Education and Scientific Research, out of consideration of the extremely high scholarly value of the holdings in the collection:

“witness the epistles of Lucius Annaeus Seneca, found in a codex dating to the first half of the twelfth century, and the Laudi Sanctae Crucis of Rabano Mauro, found in what is thought to be the oldest codex in the entire collection, and the most elegant at that. Scholars in theology can look at the thirteenth-century codex containing Thomas Aquinas's Quaestiones de Potentia Dei. Legal scholars have at their disposal precious copies of Accursius's Magna Glossa and of Gratian's Decretum. There is, too, a late-thirteenth-to-early-fourteenth-century copy of Rolandino de' Passaggieri's Super Arte Notariae (supplemented with certain additiones, of a much later date, by Pietro d'Anzola), and equally worthy of note is codex no. 219, dating to the end of the thirteenth century, and containing, among other things, Gratia Aretino's Summa de Iudiciis - a revised and expanded version of it penned by the author himself. There also reproduced in the collection the entire Corpus Juris Civilis and Corpus Juris Canonici, preciously decorated in several places.” (Pieri, 2002)

An electronic catalog was built at first under the supervision of Domenico Maffei and Andrea Padovani, simultaneously laying the foundations for better handling of the often fragile codices, and further preservation and future study and use. This catalog would also be used as the primary source of metadata for the then still to come digital archives.

Volumes in the Irnerio Archives are of different provenience and were meant for different uses. Thus, we have the illuminated codices and we have another set of codices, plainer and more scholarly, clearly intended *ad usum scholarum*, for the use of scholars. These latter codices usually collect different works, some in the form of summaries, some in the form of anthologies, and they present materials of extreme interest for the historical

inquiry, as when looking to reconstruct the curricula and the educational system of the day.³

A vast work of digitization and classification of the collection began: all of the 289 codices were scanned in high resolution and a no-loss format⁴, from cover to cover, under the editorship of the publishing house CLUEB. In the end more than 135000 images in resolutions up to 3400 pixels were stored, among other supports, on a hundred DVDs and on backup tapes⁵. Medium resolution images suitable for Web use were obtained from these.



Figure 1: Folio 1 recto from the Biblia Vulgata attributed to Hyeronimus, Codex 2

-
- 3 Pieri (2002) details that an early 15th codex of roughly 420 leaves has “close to 520 repetitions, consilia sapientum, quaestiones, and rubricae on topics that make up a coherent group. It is not difficult for us to speculate how this collage came about, with the dominus legens laying out the subject matter and then the judges and the jurists providing examples, quite likely advancing contrasting opinions in the process.”
 - 4 The original scanning are in TIFF 24bit format, uncompressed, and average 3200 pixels on the longest side. The medium resolution images are in JPEG format and were especially edited for use on the World Wide Web.
 - 5 These have now been moved to redundant NAS appliances.

Accessing the Archives

Because of the ample interplay between the different entities involved, the Royal College, CIRSFID, the Cassa di Risparmio Foundation, the discussion on what legal constraints on the use and access to the high-resolution images was a long, difficult process. The model the Innerio Archives were to follow was finally outlined as an open-for-all digital library as far as the medium-resolution images and the content from Maffei's catalog were involved, and a paid-for subscription model for accessing the high-resolution images in full detail. This led to the development of the *Cart* concept in R1, and of the *My Page* concept and administrative interface in R2. I will not document those here, as their assumptions and outcomes largely lie outside the scope of this overview of the Innerio Archives.

Specific Information Architecture Issues

A collection of manuscripts presents unique challenges to an information architecture approach: a codex is not a book the way we are used to consider it today. The physical form can be deceptive: a codex has no cover, no title, no clear global authorship, and it is often either incomplete, missing pieces or pages, a loose collection of unrelated and distant manuscripts, or both. Finally, the language they are written in and their generic cultural allure is mostly alien to us (Diringer, 1982). As such, navigation and place-making are easily the most important issues that need to be addressed. In the Archives, authors, works, and codices are the self-evident primary elements around which a global structure can be built. Together with these, time of writing could be identified as an orthogonal sectioning principle across the whole collection. Finally, since moving around in large collections can be problematic, horizontal correlations had to be exploited whenever possible (Rosati, 2007). This was never totally resolved in R1. The model for global correlation of elements was not there, and navigation was sketchy at best. Long lists were presented to the user unfiltered and unedited, and moving between codices, authors, or works became difficult. To better assess how to solve these, a long and thorough content analysis and inventory was conducted on the initial release of the Innerio Archives.

R1: Initial Vision

At the time the 289 codices started to be digitized and classified, and the digital renditions of every single page stored on disc and tape at CIRSIFID, work began on the development of the web application which had to publish these on-line. I will refer to this application as R1, short for Release 1.⁶

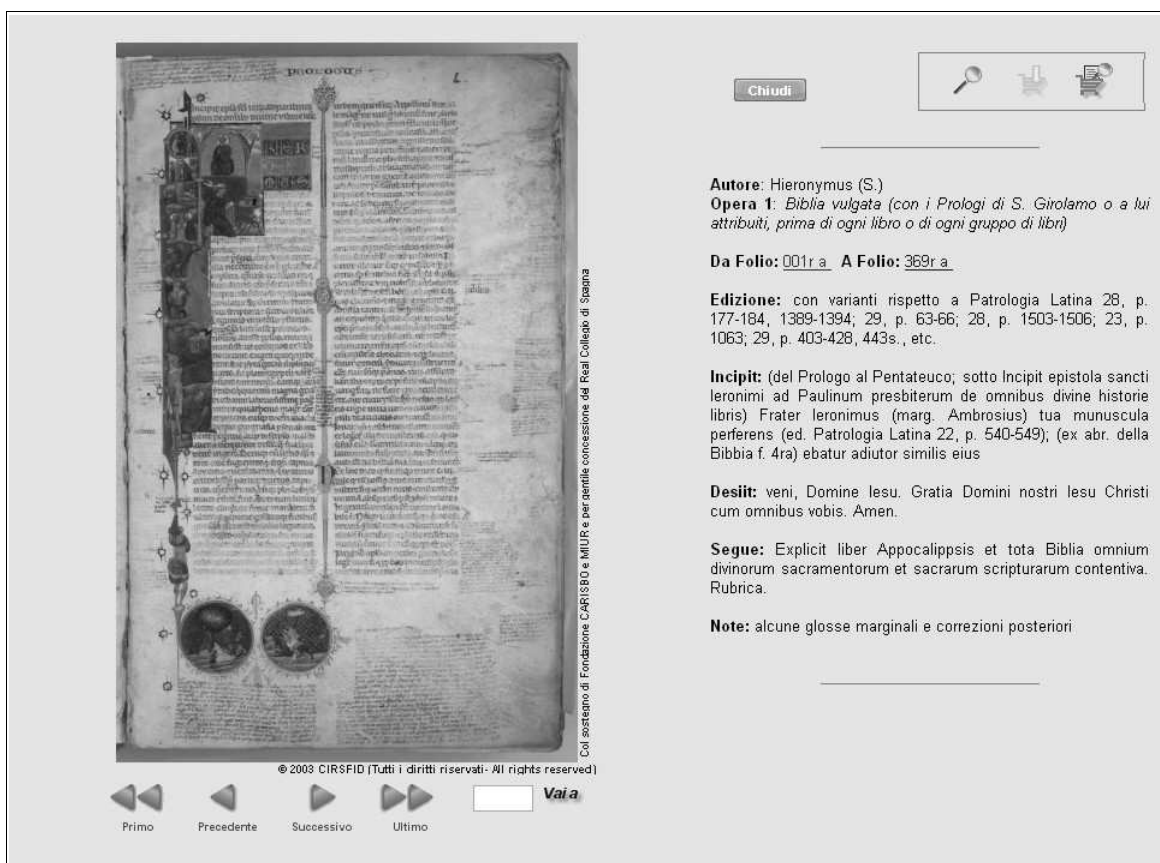


Figure 2: R1, a medium resolution codex page view. Codex navigation is below the image. Access to high-resolution images and user functions is visible top-right.

R1 was developed as a client-server PHP⁷ application incorporating a Java⁸ applet for access to the high-resolution images. PHP allowed for rapid prototyping and sufficient

- 6 Most of the facts reported and concerning R1 I learned directly in either private conversations from some of those involved in the development or in meetings. As such, no references can be provided.
- 7 PHP is an open source scripting language especially developed for the Web. Through the years it matured to become the de facto standard for many applications, including large-scale applications such as the Yahoo! repositories and, just to name a widely known and used social application, Facebook. See also <http://www.php.net/>.

robustness, the Java virtual machine allowed the developers to finely control how the high-resolution images were accessed and what could be done with them. Because of the way the Archives policies and business model were being laid out, for example, no download or local image save could be allowed by the application.

Its general purpose and structure complied largely with the “traditional digital library” model I have analyzed in Chapter II, but a few elements, either technical or content-related, were totally unique to the Innerio Archives:

- ◆ the historical-juridical nature of the texts, seen as objects of study and scientific discussion;
- ◆ the existence of the Maffei catalog as the primary source for metadata;
- ◆ the clear distinction between two separate levels of navigation at two different levels of security or privilege, one open to everyone and free, the other requiring authentication and paid-for;
- ◆ the consequent legal and business issues brought along by the necessity to work out an e-commerce administrative structure to allow users to “buy” codices for perusal at high-resolution.

General Issues

R1 was developed over a short period of time, and with incomplete and somewhat flexibly specifications, corrected and amended as the project moved along. This had reflections especially in the way the code was structured: R1 suffered heavily from fragmentation and redundancy, with many different scripts doing over and over the same piece of code. This is not only difficult to maintain, but presents security risks as well as opening up the application to logic fallacies. For this reason, and since the original developers were not part of the development team anymore, the very first step taken before starting to work on the second iteration, was a thorough code and application audit to understand some of the

8 Java is a general-purpose programming language developed at Sun Microsystems which runs on many different platforms. See <http://www.java.com/>.

inner workings and where the most important strengths and weaknesses lied. Most of these were technical, dealing with the way the programming layer had been structured. Nonetheless, a list of touch-points was elaborated discursively for communication with the project management. It included the following⁹:

- ◆ XHTML compliance in all generated pages was generally absent. The pages did not validate under the W3C Validator¹⁰. The DTD¹¹ was often not declared.
- ◆ No content / layout separation was in place.
- ◆ A proper print CSS was lacking for pages with informative content.
- ◆ Lists pages were generally too long to browse and uninformative. Some of them generated a 24000 px long page).
- ◆ The global menu was imprecise and needed to be reformulated. Items in the menu were not aware of state change (such as login / logout) and some items remained clickable even in circumstances where they should be disabled. For example, there is no need for a *How to register* menu if the user is already logged in.
- ◆ All items in the menu auto-linked and this had to be changed.
- ◆ Some pages (*Come abbonarsi*, *Registrazione – persone fisiche*, *Guida al servizio*, *Note legali*) used characters encoded in the Windows character set directly and not HTML entities. These might present issues when content-negotiating the correct ISO Character Set with browsers and generate validation errors in any case.

9 I deliberately maintain the text in its original immediateness, with very little editing and mostly directed towards making the references understandable in this context. I moved all sentences to the past tense though to improve the style and flow of discourse within the Chapter.

10 The Validator is a freely available W3C tool which analyzes web pages for compliance. Compliance is important to avoid having visual models breaking down in particular browsers. See <http://validator.w3.org>.

11 A Document Type Definition (DTD) is a set of markup declarations that define a document type for the XHTML document to conform to. Every valid compliant XHTML (or HTML) page contains at least the DTD declaration, a head section, and a body section.

- ◆ Some pages used a formal, legal language to describe the services provided by the web site (*Come abbonarsi*): could they be rendered in a more user friendly way (legal text not lost, goes to small prints)?
- ◆ There was no provision for the users to change their data, including their e-mail or username, or ways to retrieve lost information (such as passwords).
- ◆ The site provided English versions of the relevant pages only visible if browsing through a local indirect layer (*English version*), which is a frustrating experience since there is little way for a non-speaker to know that for example, “*Come abbonarsi -> English version*” actually means *Subscribe* in the first place. Language should be a global high-level choice, be separated from content and made modular through specific coding.

Presentazione	yes
Lista autori	no, unnecessary ¹²
Lista codici	no, unnecessary
Ricerca	no, unnecessary
Enti partecipanti	no, possibly necessary ¹³
Come abbonarsi	yes
Area abbonati	no, unnecessary
Guida al servizio	yes
Note legali - privacy	yes

Table 1: The original list of pages with available English versions for R1

With this list of shortcomings and improvements ready, a second phase of content inventory and analysis was started.

¹² The general translation of the user interface and in-page messages was considered a part of the language redesign and hence not connected to content. See Architecture for more on this.

¹³ For more on this see also the paragraph on Content Inventory, Pages.

Content Inventory

The list that follows details the original content inventory and structural, per page, first-view analysis of the pages in R1 that was conducted during the assessment stage for R2, with the exception of the Java Viewer applet and the Cart function which were not considered¹⁴. I left the text largely untouched and just edited when references to materials or documents developed at the time of redesign made it unclear or potentially confusing: as such, it sports a largely action-oriented and prescriptive language.

Also, all following paragraphs report the English title for the page or section and the Italian version used at the time between parenthesis. R1 was developed in Italian and only partially translated to English, so the English titles used here actually refer to the naming that was being studied for R2: since the labels allows for better comprehension, though, I decided to use them here as well. Whenever possible, a screenshot of the R1 page being described is provided for reference, as the application is no longer available on-line.

Homepage

The page only contains the project logo and the main menu. It feels empty, it should provide actual content, and possibly explain what the web site is and what may be done with it. The *Home Page* link label should be corrected to *homepage* and should not link back to itself (when on the homepage, the link should not work).

About (Presentazione)

The page details what the project is about and how it came to be, and it is rather long by web standards. The literature (Nielsen, 2007) suggests to revise the text and split it into coherent parts in different pages, or provide a local table of contents with anchoring to and from the relevant sections. Paragraph titles as well could be useful to improve place-

¹⁴ The reason being that it was a standard Java image viewer with the possibility to zoom in and zoom out of the picture. The Cart was not analyzed as per the specifications it had to be redesigned from scratch.

making and navigation and should be used. Paragraphs 9 and 10 of the original text could be moved out and used in the page for *Partners (Enti partecipanti)*.

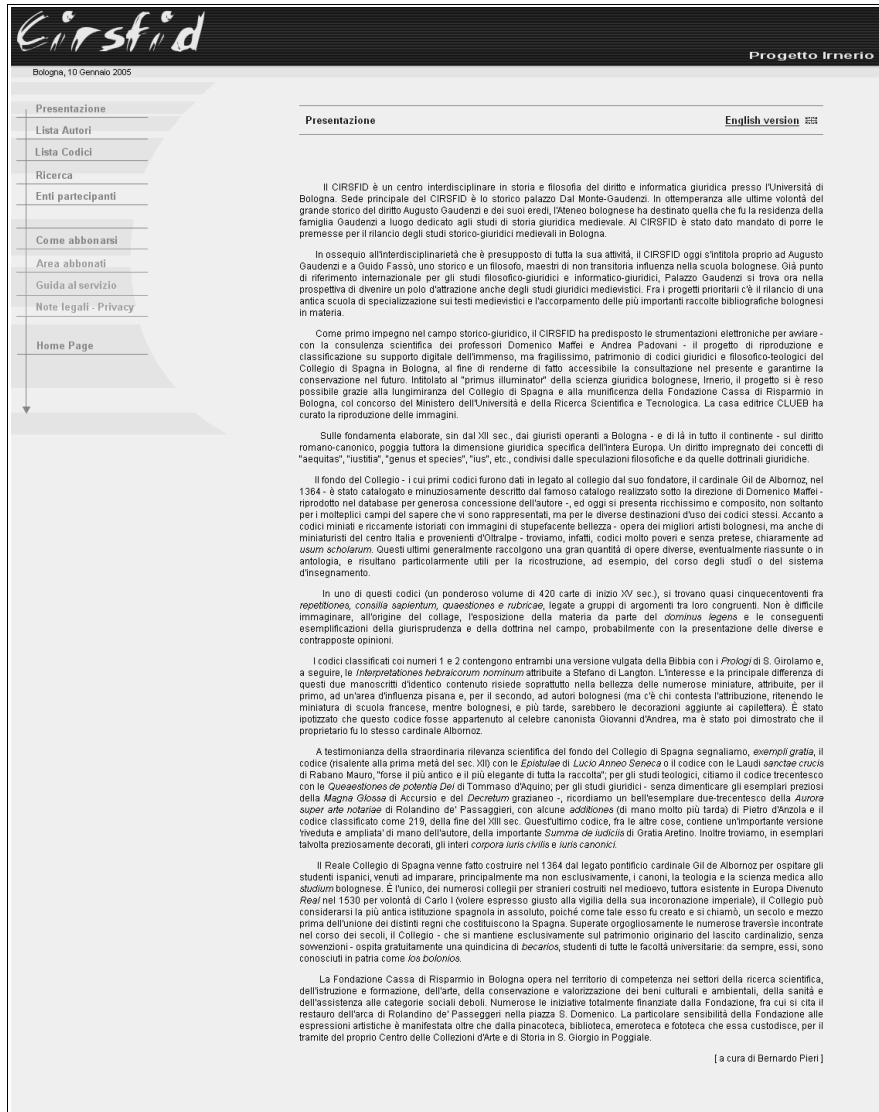


Figure 3: R1, About page (Presentazione)

Having just one single author attribution for this piece raises more issues than it solves and provides a fairly disconnected view on authorship (Who wrote the rest? Why are these lines so important that I am being told who wrote them?). As such, the author tag should

be moved to a proper authors / credits page¹⁵, as attributions are not mentioned anywhere else.

Authors (Lista Autori)

This page presents a simple list of the more than 800 authors who have works represented inside the Archives¹⁶. Technically, it is not rendered as a HTML or XHTML list with proper UL or OL descriptors, so it is semantically incorrect. The usability of this list is highly debatable, and its informational value is very low (Nielsen & Loranger, 2006).

The list could be split into a number of shorter initial listings, for example alphabetical or by date of birth / death, which should be the main *Authors* page. Once a user selects a range, the page is reloaded appropriately, for example showing all authors whose name starts with *A*. Once the user selects a single author, the detailed view for that author is loaded (see *Author*, below). A possible alternative could be to present users an initial textual authors page, with background information (which does not exist as a coherent whole and should have to be written), and have the *authors menu(s)* in a sidebar (Fig. 4).

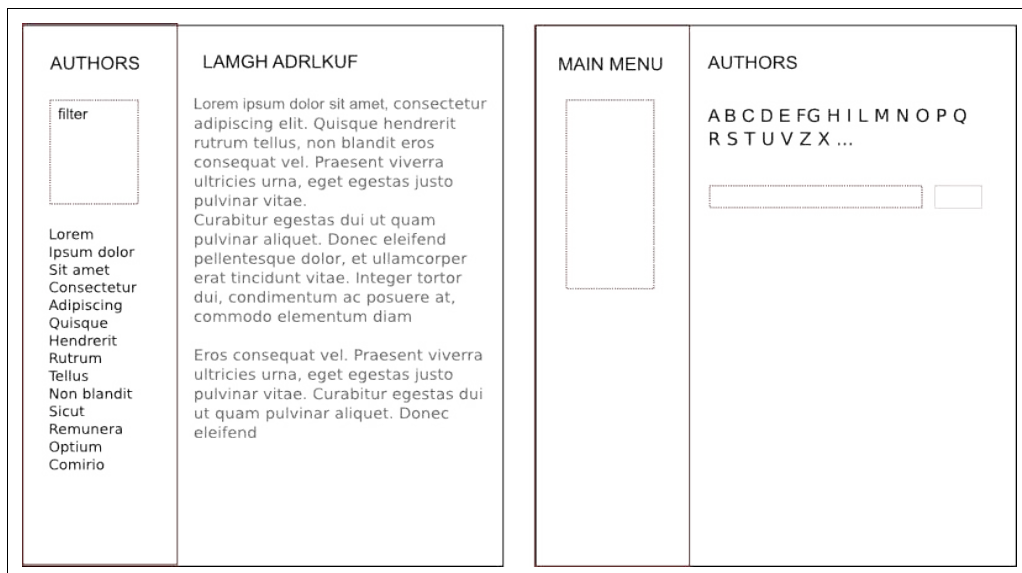


Figure 4: Alternative layouts for the Author listings

16 For a comparative view of these long listings and their actual screen size, see Fig. 5 and end of chapter.

As authors often have name variations, a second name, a nickname (*Sopranome*) is

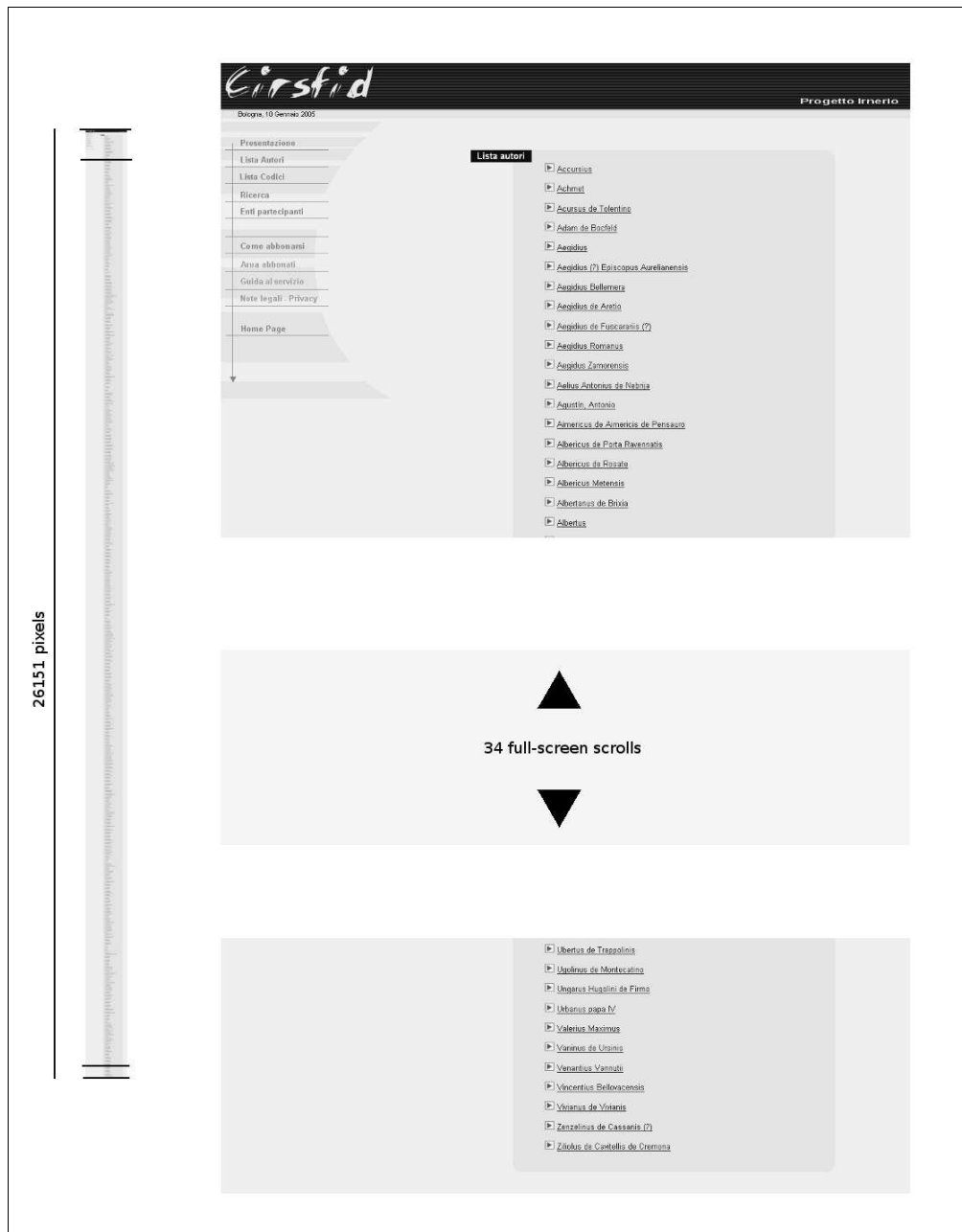


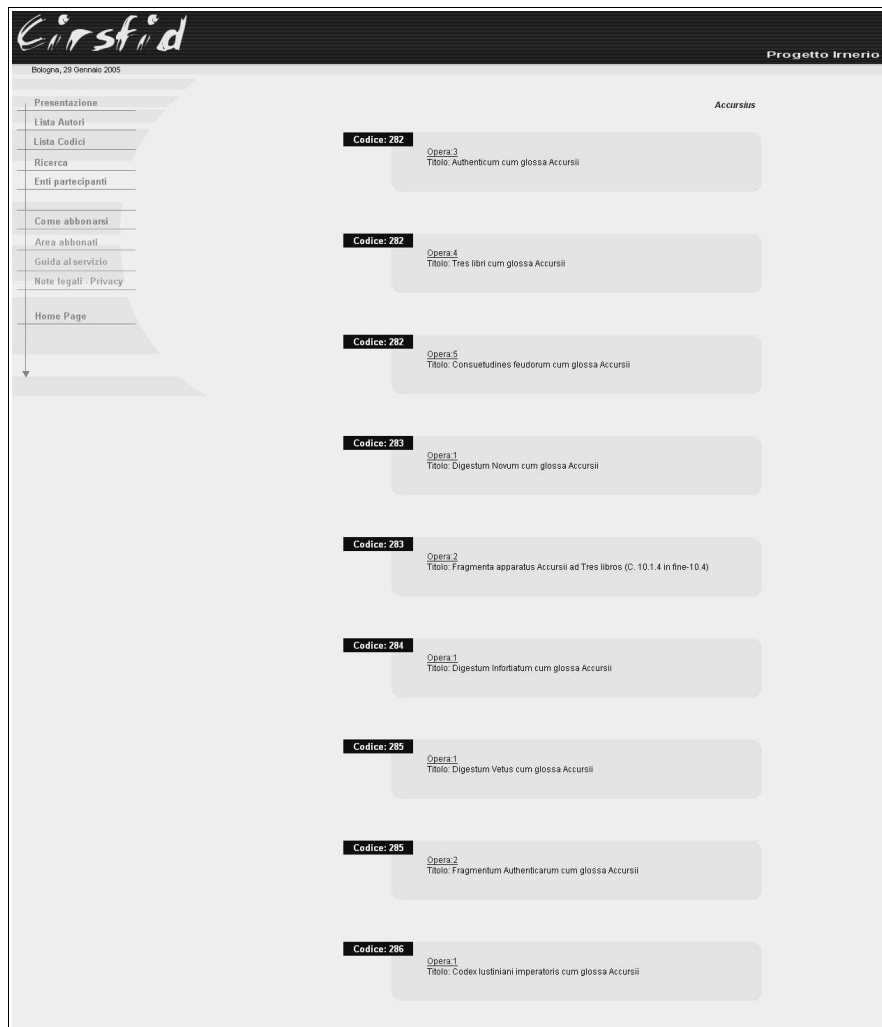
Figure 5: RI, an exemplificative composite image showing the actual length of the authors list page.

available in the database. This should also be used when searching, either by providing an alternate list or by building relationships between the terms.

Author (Autore)

This page provides a detailed view of an author's works (*opere*). It does not provide however any database-stored information about the author himself, such as date of birth, date of death, second name, etc.

Besides, single works are listed by codex but not grouped, so Accursius' works 3, 4, 5, all belonging to codex 282, are repeatedly printing that specific piece of information on every row, cluttering the list and building up a lot of visual noise.



The screenshot shows the Cirsfid website interface. The header includes the logo 'Cirsfid' and 'Progetto Innerio'. The date 'Bologna, 29 Gennaio 2005' is displayed. A navigation menu on the left contains links: Presentazione, Lista Autori, Lista Codici, Ricerca, Enti partecipanti, Come abbonarsi, Area abbonati, Guida al servizio, Note legali - Privacy, and Home Page. The main content area is titled 'Accursius' and displays a list of works. Each work entry consists of a codex number in a black box, the work number, and the title. The works are grouped by codex: Codex 282 (Opera 3, 4, 5), Codex 283 (Opera 1, 2), Codex 284 (Opera 1), Codex 285 (Opera 1, 2), and Codex 286 (Opera 1).

Codice	Opera	Titolo
282	Opera 3	Authenticum cum glossa Accursii
282	Opera 4	Tres libri cum glossa Accursii
282	Opera 5	Consuetudines feudorum cum glossa Accursii
283	Opera 1	Digestum Norum cum glossa Accursii
283	Opera 2	Fragmenta apparatus Accursii ad Tres libros (C. 10.1.4 in fine-10.4)
284	Opera 1	Digestum Infortium cum glossa Accursii
285	Opera 1	Digestum Vetus cum glossa Accursii
285	Opera 2	Fragmentum Authenticarum cum glossa Accursii
286	Opera 1	Codex Iustiniani imperatoris cum glossa Accursii

Figure 6: R1, codices by author. The list presented is that of Accursius.

Codices (Lista codici)

This page presents the complete expanded list of the 289 codices and their works the from the archive, in one single unfiltered column which, much like the author list, difficult to navigate and understand. It does not allow for reordering by different criteria, so users are either compelled to simply scroll or text-search within the page using browser functions.

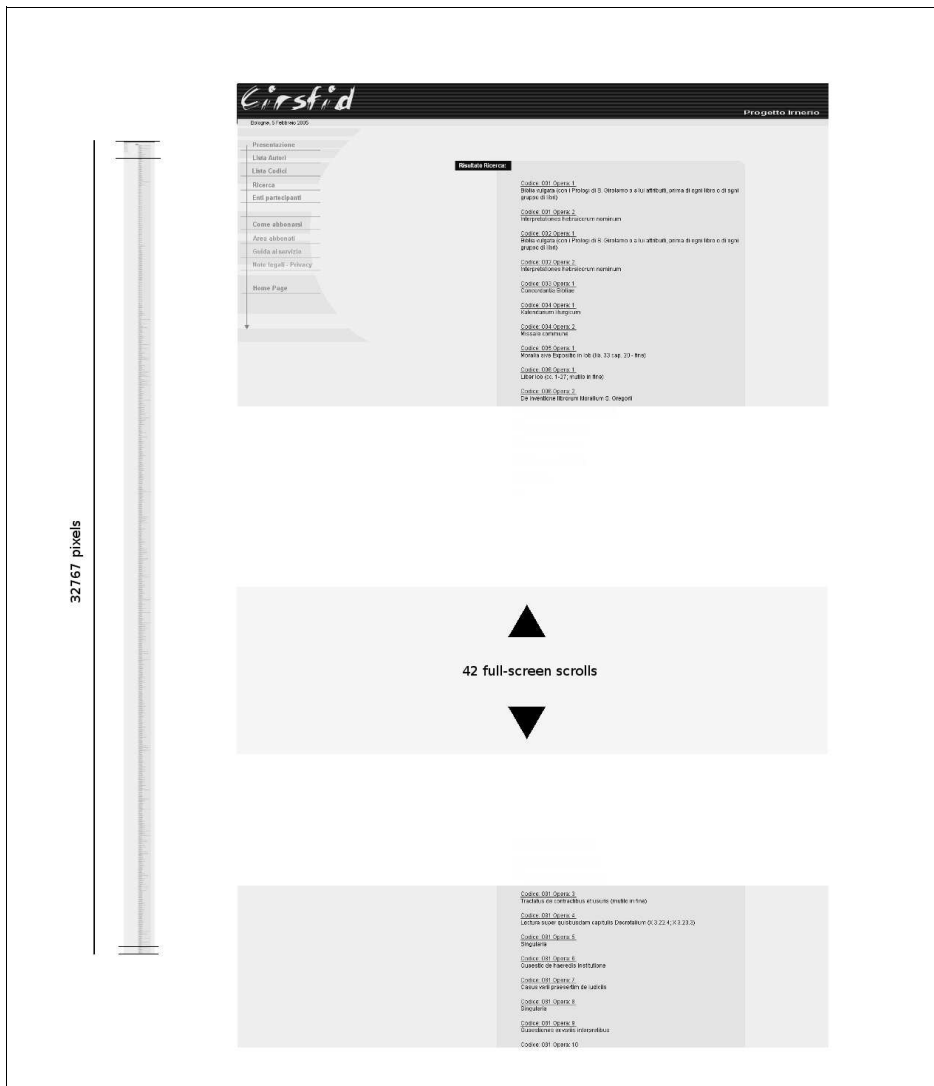


Figure 7: RL, an exemplificative composite image showing the actual length of the works list page.

Again, this is not semantically an XHTML or HTML list, but a series of DIV elements. Apart from splitting up the list into chunks, columns, or sections, the overall user

experience for this page could be improved by using categories already included in the database data and coming from the Maffei catalog such as “by size”, “by *foliation*”, “by century”, in order to provide restricted, more specific lists. Adding *Codex (Codice)* before any physical number or string which identifies every codex does not really improve use or comprehension, so that could possibly be deleted.

Codex (Codice)

The codex page opened up full screen, with the first page of the codex on the left, and the main meta data corpus on the right, including author, work, foliation, incipit, and explicit¹⁷. Navigation was possible in linear fashion by means of arrows just below the

The screenshot displays a digital interface for a manuscript viewer. On the left, a page from a codex is shown, featuring a large illuminated initial 'N' and two circular diagrams at the bottom. On the right, a metadata panel provides the following information:

- Chiudi** (Close) button and search, print, and download icons.
- Autore:** Hieronymus (S.)
- Opera 1:** *Biblia vulgata (con i Prologi di S. Girolamo o a lui attribuiti, prima di ogni libro o di ogni gruppo di libri)*
- Da Folio:** 001r a **A Folio:** 369r a
- Edizione:** con varianti rispetto a Patrologia Latina 28, p. 177-184, 1389-1394; 29, p. 63-66; 26, p. 1503-1506; 23, p. 1063; 29, p. 403-426, 443s., etc.
- Incipit:** (del Prologo al Pentateuco; sotto Incipit epistola sancti Ieronimi ad Paulinum presbiterum de omnibus divine historie libris) Frater Ieronimus (marg. Ambrosius) tua munuscula perferens (ed. Patrologia Latina 22, p. 540-549); (ex abr. della Bibbia f. 4ra) e batur adiutor similis eius
- Desit:** veni, Domine Iesu. Gratia Domini nostri Iesu Christi cum omnibus vobis. Amen.
- Segue:** Explicit liber Apocalypsis et tota Biblia omnium divinarum sacramentorum et sacrarum scripturarum contentiva. Rubrica.
- Note:** alcune glosse marginali e correzioni posteriori

At the bottom of the viewer, there are navigation controls: a search bar with the text "Vai a", and four arrows labeled "Primo", "Precedente", "Successivo", and "Ultimo". A copyright notice "© 2003 CIRSFID (Tutti i diritti riservati - All rights reserved)" is also visible.

Figure 8: R1, codex view

¹⁷ The incipit of a text, from the Latin for *it begins*, is its first few words or opening line. Before titles became common, texts were often referred to by their incipit. The explicit is conversely the the final words of a text. These are commonly used to identify manuscript works.

image, and regardless of the boundaries of works: a codex was browsable from start to end as a single artifact in one pass¹⁸. A number of issues were related to the way unnumbered pages were handled, and often *ad hoc* solutions for a specific codex had been introduced. This page provides access to user functions like high-resolution view and Cart as well.

Search (Ricerca)

The page allows free-text searching by author (*Autore*) and by work title (*Titolo Opera*). Given the limited and finite number of items in both lists, it could be argued that search is really not necessary. Nonetheless, users may expect to find it and may be more comfortable with a text-entry interface (Rosenfeld & Morville, 2002). As such, a hinting mechanism through a word wheel and auto-completion should be used to improve searching and to avoid unnecessary toil when the user is not sure of the spelling. This is a key point as the language in which the content is written is medieval Latin.

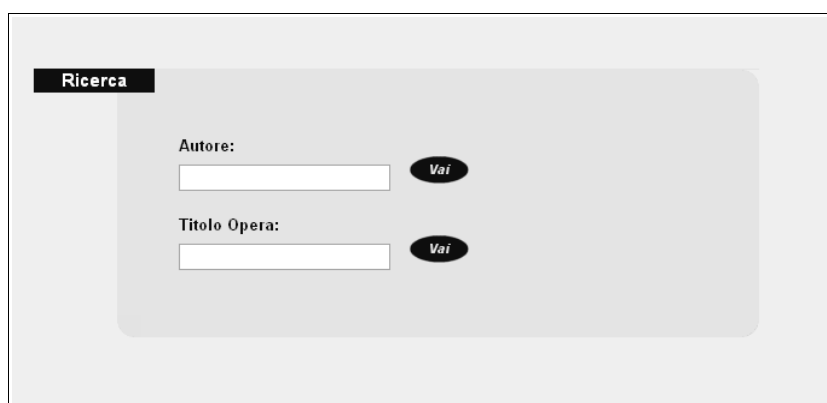
The image shows a search interface titled "Ricerca". It features two search fields. The first field is labeled "Autore:" and the second is labeled "Titolo Opera:". Each field has a white rectangular input box and a black oval button with the word "Vai" written in white. The entire search area is enclosed in a light gray rounded rectangle.

Figure 9: RI, search box in the Search page

Furthermore, when receiving result sets consisting of a single item (i.e. searching for *Accursius* produces exactly one entry), there is no need of a list / disambiguation page: the user should be taken directly to the details page for that author. Subworks (*sottoopere*) do not seem to be searchable directly and they should be added.

18 This following the physical form as opposed to the logic relationships of the elements is one of the most resilient legacies of library science and bibliographical records, and will be discussed in detail in Chapter IV when dealing with modeling languages and their logic.

Register (Come abbonarsi)

The page is quite confusing. The actual registration link should be moved to a more prominent position and renamed: the current terminology is overtly technical (“Persone fisiche” meaning natural persons as opposed to legal entities) and unnecessary, as there was no taxation distinction between subjects for the scopes of entering into a contract with CIRSFID. It should be simply named *Register*. The link to the printable (PDF) version of the contract has to be moved and renamed as well.

Cirsfid Progetto Innerio

Bologna, 10 Gennaio 2005

Presentazione

- Lista Autori
- Lista Codici
- Ricerca
- Enti partecipanti

Come abbonarsi

- Area abbonati
- Guida al servizio
- Note legali - Privacy

Home Page

Come abbonarsi [English version](#)

Descrizione del servizio

Il servizio telematico fornito dal CIRSFID, denominato «Progetto Innerio», ha ad oggetto la visualizzazione di un numero determinato di immagini digitalizzate ad alta risoluzione, con possibilità di ingrandimento, relative ciascuna a singole pagine delle opere presenti nel patrimonio culturale del Real Collegio di Spagna, meglio indicate nel catalogo liberamente consultabile nel presente sito (secondo diversi criteri di ricerca: «Lista Codici»; «Lista Autori»; «Ricerca»). L'Utente può liberamente scegliere un numero determinato di immagini da visualizzare ad alta risoluzione, nell'ambito del predetto catalogo completo, al cui interno le immagini sono consultabili in modalità «anteprima» a bassa risoluzione.

Il servizio di visualizzazione delle immagini ad alta risoluzione viene erogato, per un anno dalla sua attivazione, in favore del contraente che stipuli con il CIRSFID formale contratto in forma scritta, previa registrazione in apposita area del sito («Come Abbonarsi»; «Registrazione»).

Il servizio «Progetto Innerio» viene erogato a fronte di un contributo per la copertura parziale delle spese sostenute dal CIRSFID per la prestazione e la gestione del servizio medesimo, nonché per la manutenzione del sistema, determinato come da contratto.

Requisiti hardware e software

Hardware
 Processore: Pentium III 1000 Mhz (minimo), consigliato Pentium IV
 RAM: 256 MB (minimo), 512 MB consigliata
 Monitor: risoluzione 1024*768 pixel (minima)

Software
 Sistema operativo: Microsoft Windows 2000 o XP
 Web browser: Internet Explorer 6 o superiore
 Virtual Machine Java di Sun

Requisiti di sicurezza
 Esecuzione di Javascript abilitata
 Cookies di sessione abilitati

Registrazione

Il primo passo da compiere per poter usufruire del servizio è quello di registrarsi utilizzando il link che appare in alto a destra in questa pagina, scegliendo la categoria di cui si fa parte.
 È importante compilare attentamente i campi presentati, soprattutto quello dell'email, che verrà utilizzato dal sistema per l'invio del contratto da compilare a cura dell'utente.

Registrazione

- [Persone Fisiche](#)
- Facsimile contratto
- [Persone Fisiche](#)

Figure 10: R1, the Register page

The text of the page needs some corrections in the technical requirements section: PC specifications should be revised; there is no need to require a specific operating system on the client side; there is no need of a Java Virtual Machine on the client side; users should

be able to use any standard-compliant browsers. Furthermore, having Javascript enabled is not by any means a security requirement (as phrased in “Requisiti di sicurezza: Esecuzione di Javascript abilitata”) and “Cookies di sessione abilitati” (Session cookies have to be enabled) should be rephrased in something along the lines of “Please enable cookies to access all features of the Archives”. The text for the registration should offer no page-related physical hint as to where the link / button is relative to the page layout, as this is dependant on a number of factors (now it says “top-right”).

Login (Area abbonati)

The page just holds the login form for registered users. It should be redesigned to offer ways to recover a lost password or lost username (either via e-mail or through some other procedure). Both the menu and the page should also check on the user’s status and modify their behaviour to offer appropriate *Logout* mechanisms when a user is already logged in.

My Page (Pagina utente)

This page is only available to registered (i.e. paying) users and lists the codices, parts of a codex, or single pages a user has bought (meaning the user has access to high-resolution images), their last access date, their end of contract date, their residual credit (to acquire new high-resolution images). It directly links to Cart (*Carrello*), which is a pop-up window. This behaviour should be changed and brought back into normal flow, as this could possibly be confusing (Zeldman, 2003).

The list of codices is enriched by thumbnails, and by clicking the side link (change into clicking the image itself) the user is presented with the high-resolution Java Viewer Applet. A redesign might be necessary as this functionality would be better moved server-side for better maintenance and in order not to require third-party software client-side apart from a compliant browser (Reiss, 2006).

The user page is directly linked against a menu item child of Login (*Area abbonati*), but there is no global logout mechanism: logout is tied to an internal page. To log out, you

have to go to the main user page, which also holds the complete list of items in the user's cart. This is wrong functionally, as viewing / editing the user data and viewing / editing images from the catalog are two different tasks (Nielsen, 1993), and from a user experience point of view, since the users who want to access their data or cart have to wait for their current items to load. A possible redesign is outlined in Fig. 12.

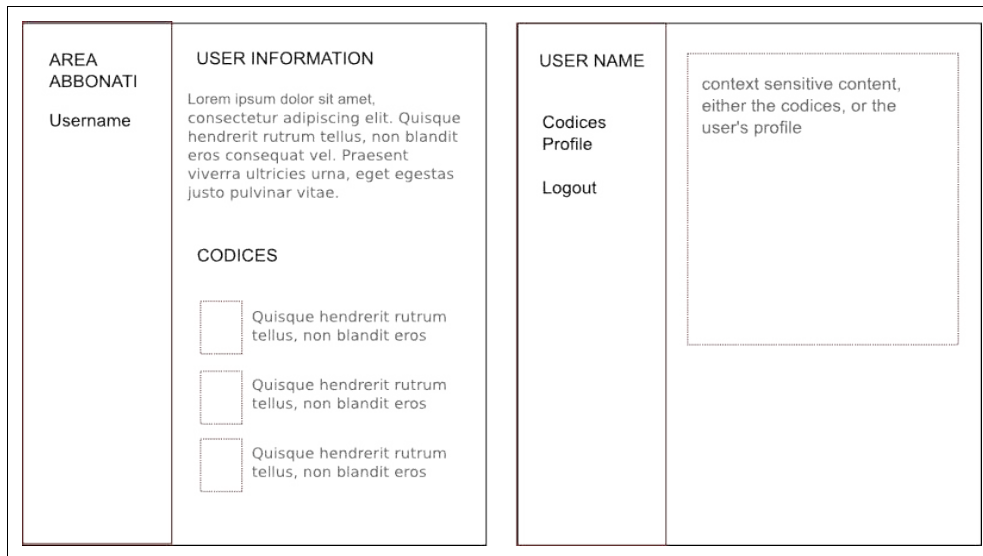


Figure 11: The current layout for the user's page (left), and an alternative design for the same (right)

Help (Guida al servizio)

The page suffers from some of the issues we outlined elsewhere as the text is too long and possibly too technical, with a couple of paragraphs rendering incorrectly due to bad HTML encoding, but it generally works and is better structured than most. It details how to access and use the Archives, and provides screenshots, a table of content, and HTML anchors for quick access to the relevant sections. It might be necessary to split it in different parts, but it will necessarily have to be rewritten to take into account the changes introduced in the UI. A ready-to-print version could be provided as well.¹⁹

¹⁹ For a screenshot, see end of chapter

Terms of Use (Note legali – Informativa privacy)

The page is not semantically correct and presents a number of encoded Windows-only-characters. It completely lacks a table of contents and anchors, which could greatly help readability. Structure-wise, it should be moved out of the main menu and into a utility menu (bottom of page, possibly), split in two (*Terms of use / Privacy*), and listed along the *Copyright* notices and the *Impressum* containing all web site related info (contacts, authors, compliance, etc).²⁰

Partners (Enti partecipanti)

The page has no real information for users: it is just a number of logos and links which point away from the web site. As such it could easily be turned into a sidebar if there is no contractual reason to maintain it as it is. If it has to be maintained as an individual page, a few notes and contact information for the partners would make it more useful.



Figure 12: R1, partners page

²⁰ For a screenshot, see end of chapter

The logos are not graphically well-composed: they are not evenly spaced along the axis and they are of different sizes. The background colour should not be hard-coded around the image (currently set to #FEF79C). The CIRSfid logo should be replaced with the one currently showing on www.cirsfid.unibo.it. There are miscellaneous issues with the links provided for the Cassa di Risparmio Foundation and MURST / MIUR.

Comparing Innerio with the Case Studies

“Non-goal oriented immersion in a digital library resembles similar experiences in other milieus, namely museums and art galleries, and shopping” – Elaine Toms, 2004

A very quick comparison can simply point out a number of common patterns in the handling of navigation and lists as the most visible similarities.

Access	Partially open: medium-resolution available to everyone, high-resolution needs paid-for subscription
Navigation	Via a main menu, and via linear horizontal navigation inside codices
Labeling	Minimal, sometimes technical
Grouping	By codex; by author
Search	Free-text, on title and author
Images	Available in medium-resolution (1000px) and high-resolution (3400px) (paid-for subscription)
Information	Not available
User services	User profile, with cart and administrative pages (paid-for)
URL structure	-
Harter compound	A1; B2; B3; A4; A5; A7; A8; C10; C11

Table 2: Heuristic-based evaluation of the Innerio Archives R1

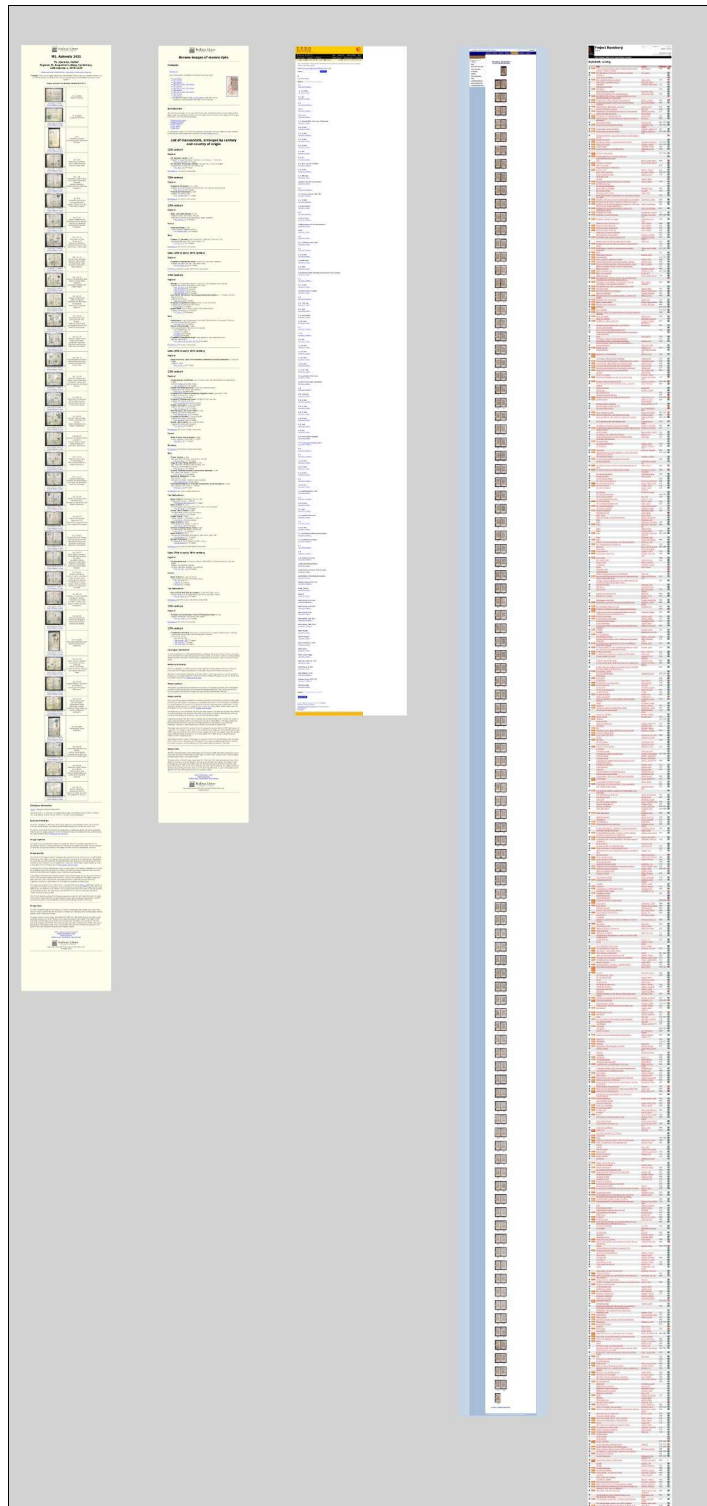


Figure 13: Comparative image showing list pages in a number of different digital archives

Fig. 13 is a a very simple and down-to-earth comparative table showing the length of some of the list pages in the case studies from Chapter II and other digital libraries: from left to right, the Bodleian Library (part of the Early Manuscripts at Oxford), the Early Manuscripts at Oxford Digital Library, the Early English Books Online (EEBO)²¹, the Laurentius at Lund, and Project Runeberg.

These lists are unfiltered and present no paging or reordering capabilities, and they produce extremely long pages. It is interesting to note that Project Runeberg tops at more than 23000 pixels, followed by Laurentius at more than 13000 pixels: that means between 30 and 18 full screen to scroll these lists from top to bottom at a normal resolution. Innerio R1 scores even higher, with list pages such as that for works topping at more than 32000 pixels²². Even though this research makes a point that “experiences with libraries (whether physical or virtual) need not be for the explicit purpose of finding, acquiring and using information” (Toms et al, 2004), as “(t)he experience and its playfulness and pleasure have equal value”, it is difficult to argue that this is a suboptimal way of presenting information.

Making the Transition

The reboot of the Innerio Archives was started in 2007, under the supervision of Monica Palmirani. The author was to act as technical and architecture lead on the project, and Luca Cervone and Regis Riveret, both of CIRSIFID, as *prima facie* developers. A note from an early project meeting reports that the team was looking especially into following:

“Language issues (Latin), incomplete information concerning the authors, duplication, synonymy, necessity of central control, authoritativeness, and openness. Labeling, menu structures, linking, and general application vision and deployment are to be considered.

21 Early English Books Online (EEBO) is a digital library of works from Pollard and Redgrave's 'Short-Title Catalogue', Wing's 'Short-Title Catalogue' and the 'Thomason Tracts'. In total, this database provides full-text images of over 125 000 individual titles. EEBO is maintained by the University of Michigan.

22 See early on in the chapter and the listings at the end of chapter.

In the end, R2 had the overall goal of building a more modular, robust system while addressing a number of issues with the data layer and the representation layer that had become evident since R1 was released. It also had to introduce a couple of features which had been originally contracted but not implemented for technical reasons yet, such as the use of a thesaurus to improve the internal search function, a complete English version of the application, or a way to add comments to codices.

The application was to be set up as a single correlation engine: the final goal was to allow the maximum freedom of informed movement inside the codices to the users.

Technical challenges were also briefly discussed, and the possibility to abandon a PHP framework in favor of Java was considered and finally discarded because of the available skill sets and because of development and architectural considerations.

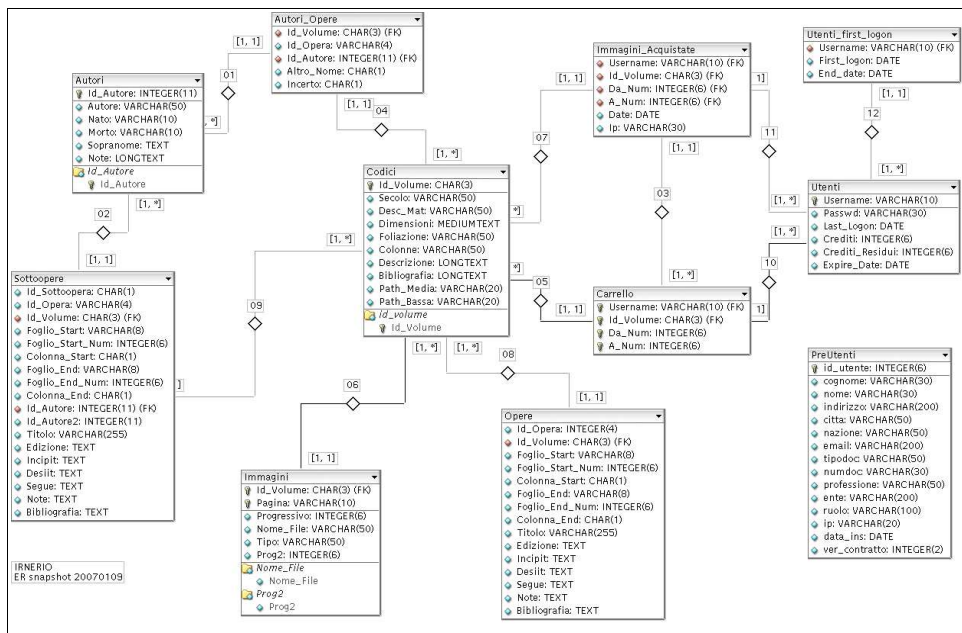


Figure 14: The database schema for R1

It was also decided to improve the underlying database schema, taking it to normal form, and to correct and integrate the data wherever possible, but to invest little time in architectural changes which were hardly justified by the goals at the time.

Schematically, the new release (R2) was to:

- ◆ improve the quality of the code using logic / presentation separation;
- ◆ improve the engineering approach via an hybrid MVC approach;
- ◆ improve the search function;
- ◆ add missing functionalities;
- ◆ improve the user experience.

This latter item was finally reduced to a number of qualitative requirements tied to the information architecture evaluation heuristics, navigation, labeling, grouping, and place-making. R2 was in this respect to:

- ◆ generally improve the navigation systems within the web-site: the global site navigation, and the local codex navigation;
- ◆ streamline labels;
- ◆ improve the user experience by making searching, finding, and re-finding better, by means of logical grouping, paging, and general attention to the way information is presented to the user;
- ◆ improve the correlation, the horizontal linking, between elements;
- ◆ improve place-making by creating a clear hierarchy of elements, and by making their ties explicit and recognizable;
- ◆ resolve local issues with single pages in terms of language, layout, and flow.

Finally, it was decided to look for open source solutions to modify and tie in for the managing and use of the thesaurus²³, and to make the language layer completely

²³ This was finally identified in the GPL web application OpenTes. The thesaurus will not be discussed in detail because of its limited importance in respect to both the development of the application and the goals of this research. For more on OpenTes, see <http://www.openthesaurus.de/>.

independent of code and presentation, so to make it easier to add further locales in the future. As for the requirement concerning the possibility to add comments, in the light of the move away from a Java solution for the high-resolution image viewer, it was decided to pursue a full Javascript solution, with the final goal of being able to add notes directly over the images and saving these as conversational threads in the database. This was the first conception of what was to become the annotation engine, the Irnerio in-line editor for adding comments directly on the folia themselves. The full rationale for the move from Java to Javascript was articulated as such:

“The original idea (confirmed at the kick-off meeting) was to implement the viewer as a Java applet, client-side. Pros in favor of this approach are: Environment control, so that users were prevented from printing images or saving them for copyright reasons, as per the project’s requirements (4. Irnerio Web – sottofase A, Lato client, par. 1, although it must be noted that the specs mention registration and authentication as a means to prevent security breaches); Quick responses after the initial loading (no continuous requests to / from the server); Possibility to zoom into the image. Cons are mostly related to: Java on the client may be unavailable, or in a non-tested configuration (release number in the past or in the future), or working unreliably due to client-side issues; Adding notes, thumbnails, comments and zooming requires a full-fledged small application which is to be written from scratch; Thumbnails should be somewhat saved to the server, with two direct consequences: a client application has to have write capabilities in server space, and an exchange protocol has to be figured out. However, since working server-side could possibly cut down on time and new code due to the large number of open source solutions for server-side scripting languages, we decided to investigate possible alternatives, which we found out mostly consist of cross-over PHP / AJAX solutions. Pros: No need to concede control to client applications; Integrated and centralized approach; No client-side environment issues. Javascript has to be enabled, but this is a requirement for most web applications as of today; No client-side update issues; Greater degree of UI control and integration. Foreseeable cons we thought could impact this solution were: Slow connections and loading times; Possible UI issues when dealing with large images (zoom, move, etc); Necessity to block and prevent hi-res image downloads. The use of AJAX actually solves the first, since the way the application works is and second issue, since Of these three issues, the AJAX implementation actually leaves out.”

R2: Second Iteration



Figure 15: Homepage, Irnerio Archives R2

The refactoring of the code and the redesign of the information architecture of the application proceeded in parallel. I already outlined the general goals of R2: I should also add that the most poignant goal of the code and engineering side was to develop something which could be used modularly and maybe even reused for other projects concerning digital libraries.

As a consequence, a number of important principles made their way into the software engine but not into the actual on-line digital archive, meaning that they produced no real changes in the information architecture of R2, or in its presentation layer. Among these, the most significant were that the new software engine was written from scratch in PHP5, with the front-end being completely XHTML / CSS / JS. No add-ons, third-party software or modules were necessary to access any part of the DL. The presentation layer was completely abstracted into a template system written in pure XHTML / CSS, so that it is

possible to change the Archives look and feel without any technical coding competence and without the need to put the application in maintenance mode²⁴.

Similarly, the language layer was completely redesigned and moved to a plug-in system. It is now currently possible to add new languages by simply modifying a configuration setting and by providing a plain-text list of the application messages in that language. The engine also identifies the browser language during client-server content-negotiation and sets it as the default on first visits: users can change this transparently and at will at any time.

Finally, the new design presents an MVC-like approach, with one single point of entry through the *index.php* file which acts as a controller and performs all runtime and security checks before initializing the application and returning pages to the client²⁵: this makes spotting bugs and issues easier, and improves control of the output.

Most of the navigational, labeling, and place-making issues were solved by a complete redesign of the information architecture: where R1 only allowed hierarchical navigation from the homepage to list or search pages and from these to the codices, R2 allows easy horizontal navigation by means of correlation: users can move between works belonging to the same codices, between works written in the same period, and between works written by the same authors. They can even jump from one author to authors who appear in the same codices.

At any single time it is possible to either filter results by predictive typing or alphabetical order (authors), time period (codices), catalog number (codices), physical location (works), authorship (works), and string matching (works).

24 As a matter of fact, the current look and feel was designed to resemble the original one from R1 as closely as possible for contractual reasons, and was dubbed “Old school”.

25 Pages are buffered and only rendered to the browser viewport upon successful completion.

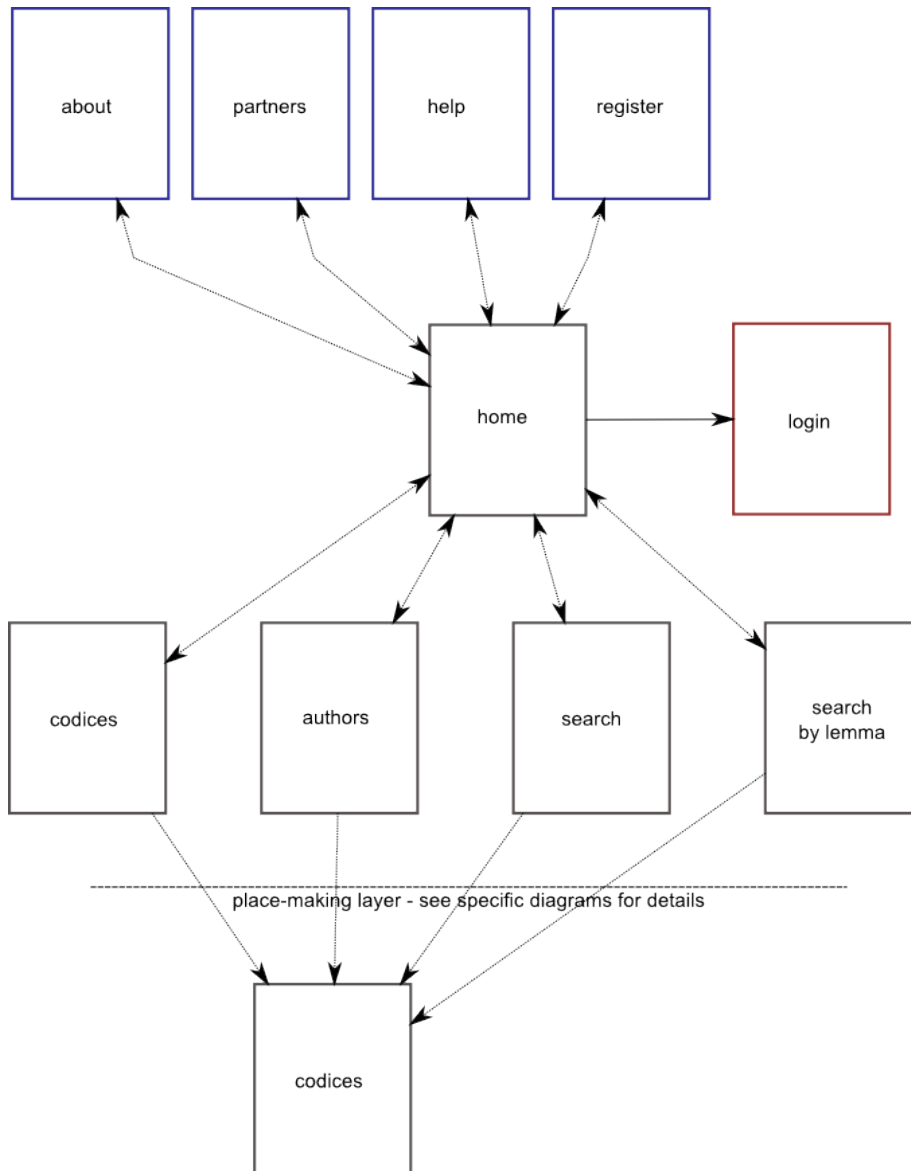


Figure 16: R2, general information architecture

R2 works effectively towards bringing the users to the codices by using whatever approach suits their seeking behavior best, providing something similar to polyhierarchical access (Kalbach, 2007) to the information available in the DL and improving the general information scent all through the initial stages of seeking (Rosati, 2007).

The main menu has been reshaped to offer distinct, task-oriented sections, and to group all codex-related functions under one label, *Codices*. The menu is now also completely

dynamic, reflecting the user's status. For example, *Login* and *Register* change to *Logout* and *disappear respectively when a user is logged in*. Some items, namely Terms of use, Privacy, and Language have been moved to the bottom of the page, in the footer. A short message also recaps the user's status here as well.

The URLs behavior has been completely redesigned to offer recognizable paths and addresses: every single action or page has been made human-readable, and it is now possible to predict the logic with which pages are laid out. For example: if `<site>/author/` is the authors general page, `<site>/author/A/` will be the restricted list of all authors whose name begins with the letter "A", and `<site>/author/A/12` will be the landing page for Accursius. Similarly, `<site>/codex/` will be the codices list page, `<site>/codex/002` the landing page for Codex 002, and `<site>/codex/002/001r` the codex first *recto folio*.

Even when accessing as a paid-for, logged in user, these URLs remain stable, as all security and privileges checks happen before. There is no URL redirection, but simple escalation of a user's status. This greatly enhances the navigational and place-making capabilities of the application.

I will now briefly detail some of the major changes in the various pages before describing a sample search session and then moving on to describe the annotation engine and its significance for later developments.

Authors

The authors list and the author page were heavily redesigned to allow improved and better interaction.

The authors page breaks down authors in alphabetical groups with query previews, meaning that mouse-overs on a given letter offer the number of authors which can be reached by clicking. It also offers a predictive entryfield which allows user to start typing and be offered a dynamic list of authors which checks both names and second names.

AUTHORS

Authors have been split according to their name's first letter to increase findability.
It is also possible to enter names or nicknames or parts of them in the search box below for faster access.

A B C D E F **G** H I L M N O P Q R S T U V Z

(48 authors)

Search for authors

Start typing either a name, part of a name or a nickname.

Figure 17: R2, Authors page with the entryfield for searching and the alphabetical groups. Query preview is visible.

Conversely, the author landing page, the one a user is taken upon choosing a single author, has been similarly streamlined and made more rational. Works are grouped by codex, with no redundant text attached, and the list is offered after the author's name and all available author's data²⁶.

Horizontal navigation and correlation is provided via a complete list of all other authors who have works in any of the currently displayed codices. This implicitly brings time and space correlations²⁷ into the conversation, and augments the serendipity while searching (Quintarelli, 2005).

26 Most of the time this is either simply the date of birth or the date of death.

27 As having work in the same codex is not a sufficient condition for being either contemporary in time or space. Works could have been added later, copied later, or simply bound together for reasons that have nothing to do with the authors.

<p>AUTHOR: ACCURSIUS</p> <p>Born: 1182</p> <p>Dead: 1260</p> <hr/> <p>Codex 282</p> <hr/> <p>3. Authenticum cum glossa Accursii 4. Tres libri cum glossa Accursii 5. Consuetudines feudorum cum glossa Accursii</p> <hr/> <p>Codex 283</p> <hr/> <p>1. Digestum Novum cum glossa Accursii 2. Fragmenta apparatus Accursii ad Tres libros (C. 10.1.4 in fine-10.4)</p> <hr/> <p>Codex 284</p> <hr/> <p>1. Digestum Infortiatum cum glossa Accursii</p> <hr/> <p>Codex 285</p> <hr/> <p>1. Digestum Vetus cum glossa Accursii 2. Fragmentum Authenticarum cum glossa Accursii</p> <hr/> <p>Codex 286</p> <hr/> <p>1. Codex Iustiniani imperatoris cum glossa Accursii</p>	<p>Other authors</p> <p><i>Other authors whose works belong to the same codices</i></p> <ul style="list-style-type: none"> ▪ Fredericus Imperator II ▪ Ioannes Bassianus ▪ Ioannes Egitanensis ▪ Iustinianus imperator
---	---

Figure 18: R2, single author landing page. Accursius is shown here. Works are grouped by codex.

Codices

The *Codices* list page produced a lot of discussion. As the only proper way to organize the volumes is their catalog number, since we have no title or single author, or publisher as in modern books, the list will always remain up to a certain degree obscure to anyone not privy to the specific materials contained within the Innerio Archives. It was decided that this was to be addressed differently, providing better and quicker search, and that everything that could be done amounted to better visual organization, and filtering, which was already happening in different parts of the application.²⁸

²⁸ Codices can be retrieved by author from the Author page, for example.

Secondarily, a codex preview mouse-over, appearing on the page when placing the mouse on a certain codex number, was added to increase the information scent. The codex preview gathers and prints codex information such as material, century, foliation, size, and columns (Fig. 19).

The codices are listed in a multi-column display divided into chunks of 50 items each, for better scanning, visually separated and clearly labeled. On the right, a simple drop-down select box allows to filter the codices by century (Fig. 20) thus providing shorter and more focused listings.

Once users select a codex, they are taken to the codex landing page (Fig. 22).

The codex page provides a complete synopsis of all available information and metadata from the Maffei catalog, a clickable list of works, a bibliography, and the list of unnumbered pages belonging to the codex. These are accessible on a one-on-one basis, and lie completely outside the normal navigation flow²⁹. By clicking on the link for a certain work, users are then taken to the work's landing page where actual navigation through the pages is possible³⁰.

This behavior marks a major departure from the information architecture model of R1: R1 allowed for unrestricted navigation across a codex, from first page to last page, exactly mimicking the way the physical volume could be browsed. R2 avoids that, and imposes limits per work, as this is the logical base unit. If a user browses for example the *Interpretationes hebraicorum nominum*, work 2 from Codex 002, the only way to browse work 1 on the same codex is by means of the correlation links on the right side of the page.

29 In this author's opinion, this still remains one of the unsolved issues R2 did not address completely.

30 Either at medium-resolution (open access), or high-resolution (paid-for subscriptions).

CIRSFID
Irnerio

Project

- Homepage
- About the project
- Partners
- Help
- Registration
- Login

Codices

- Codices
- Authors
- Search
- Search lemma

Contract

- Contratto
- Contract

CODICES

Complete list of codices

Da 001 a 050

001	002	003	004	005	006	007	008	009	010	011
012	013	014	015	016	017	018	019	020	021	022
023	024	025	026	027	028	029	030	031	032	033
034	035	036	037	038	039	040	041	042	043	044
045	046	047	048	049	050					

Da 051 a 102

051	052	053	054	055	056	057	058	059	060	061
062	063	066	067	068	069	070	071	072	073	074
075	076	077	078	079	080	081	082	083	084	085
086	087	088	089	090	091	092	093	094	095	096
097	098	099	100	101	102					

Da 103 a 153

103	104	105	106	107	108	109	110	111	112	113
114	115	116	117	118	119	120	121	122	123	124
125	126	127	128	129	130	131	132	133	134	135
136	137	138	139	140	141	142	143	144	145	146
147	148	149	150	152	153					

Da 154 a 207

154	155	156	157	158	159	160	161	162	163	164
165	170	171	172	173	174	175	176	177	178	179
180	181	182	183	184	185	186	187	188	189	190
191	192	193	194	195	196	197	198	199	200	201
202	203	204	205	206	207					

Da 208 a 258

208	209	210	211	212	213	214	215	216	217	218
219	220	221	222	224	225	226	227	228	229	230
231	232	233	234	235	236	237	238	239	240	241
242	243	244	245	246	247	248	249	250	251	252
253	254	255	256	257	258					

Da 259 a P

259	260	261	262	263	264	265	266	267	268	269
270	271	272	273	274	275	276	278	279	280	281
282	283	284	285	286	A	B	C	F	G	H
I	J	K	L	P						

Filter by century

X-XI

Select any period to filter the codices accordingly

Codex preview

Century:
XIV

Material:
membranaceo (1 cartaceo)

Size (mm):
359x265 (285x185)

Foliation:
1493

Columns:
2

2003-2007 © CIRSFID - Tutti i diritti riservati - All rights reserved - Col sostegno di Fondazione CARISBO e MIUR e per gentile concessione del Real Collegio di Spagna

User: Not logged in Terms of use Privacy statement Language: Italiano

Figure 19: R2, complete list of codices. Filtering by period is visible on the right. Below filtering the codex preview has been triggered by mouse-oving on a specific codex.

This apparent restriction tries to emphasize via navigational constraints the atomic nature of works, its place-making connotation, as opposed to the mere physical bounds of

codices, and to encourage users to better build a complete mental model of the DL (Ding et al, 2009).

Once a user navigates to a work's landing page (Fig. 22), she is presented with an overview page that shows the Codex number top-most, the work's title, edition, incipit, explicit, segue, and eventual annotations. On the left, an image of the first folio is displayed, with horizontal navigational aids just below. On the right, links to other works from the same codex are provided for correlation.³¹

While schematically this page does not differ that much from the view provided by R1 in its codex view, there are a few important structural changes, as the page has been made atomic (self-supporting) and clearly placed and connected in both the global structure and the local codex hierarchy.

The screenshot shows the 'CIRSFID Innerio' website interface. The main content area is titled 'CODICES: XV CENTURY' and displays a grid of 182 codex numbers, arranged in 18 rows and 10 columns. The numbers range from 009 to 270, with the last row containing 271, F, and I. A 'Filter by century' dropdown menu is active, showing 'X-XI'. Below the grid, there is a 'Back to complete list -' link. The left sidebar contains navigation links for Project, Codices, and Contract. The right sidebar contains a 'Codex preview' section. The footer includes copyright information and user status.

Figure 20: R2, codices list page with filtering by period active

31 The original design called for more correlation boxes linking to other authors, or works from the same author in other codices, or works from the same period, which proved too unreliable to be added successfully at the time.

Consistent and consistently enforced labeling allows for optimal place-making: not only are page numbers with recto (*r*) and verso (*v*) printed as the central element of the local navigation widget under the image, but the URLs themselves follow this notation, and the same do medium-resolution images filenames for users who download them.

CIRSFID Irnerio		
<p>Project</p> <ul style="list-style-type: none"> ■ Homepage ■ About the project ■ Partners ■ Help ■ Registration ■ Login <p>Codices</p> <ul style="list-style-type: none"> ■ Codices ■ Authors ■ Search ■ Search lemma <p>Contract</p> <ul style="list-style-type: none"> ■ Contratto ■ Contract 	<p>CODEX: 002</p> <hr/> <p>Century: XIV/1 Material: membranaceo Size (mm): 425x265 (300x180) Foliation: 1-II+399+III-IV Columns: 2</p> <hr/> <p>Description</p> <p>Altre segnature: IV-4; 8-II-4. Due mani, corrispondenti alle due parti del codice. La prima parte (ff.1r-369r), in littera bononiensis del sec. XIV/1, rivela la sua provenienza scolastica grazie alla scrittura, alla costante indicazione delle peciae e al regolare intervento del correttore, in conformità alla normativa bolognese; la seconda parte (ff. 369v-399v) è in gotica libraria di mano transalpina. Fascicoli di ff. 5, con richiami. Colonne di linee 63 nella prima parte, di 65 nella seconda. Due foliazioni di epoca diversa: una in numeri arabi, che abbiamo seguito, l'altra in numeri romani. Nel primo pezzo sono indicate le peciae in quattro parti: la prima dall'inizio alla fine del libro di Giobbe (ff.1ra-161vb) in 36 peciae, la seconda comprende i Salmi (ff. 166rb-180vb) in 4 peciae, la terza dai Proverbi al II dei Maccabei (ff.186ra-293vb) in 16 peciae, la quarta comprende il Nuovo Testamento. Alla fine di ogni fascicolo della Bibbia è indicato il segno del correttore. Per la perdita di un foglio, la Bibbia inizia a f. 4ra con la Genesi c. 2 mutilo. Nel foglio di guardia membranaceo in fine (gp_c) un frammento del Breviarium decretorum di Ioannes de Deo (dalla C. 10 alla 31), ripetutamente edito, ma con adattamenti, omissioni e varianti (Inc. Decima causa habet tres questiones). Legatura in assi con pergamena nel dorso. Scheda di Martínez de Pons al f. ga_br.</p> <hr/> <p>Works</p> <ol style="list-style-type: none"> 1. Biblia vulgata (con i Prologi di S. Girolamo o a lui attribuiti, prima di ogni libro o di ogni gruppo di libri) 2. Interpretationes hebraicorum nominum <hr/> <p>Bibliography</p> <p>A. Martínez de Pons, Bibliotheca Albornotiana. Continet manuscriptos codices qui adservantur in Regali ac Majori Collegio Hispanorum gratia ab eximio Card. Albornotio erecto..., ad cod. [MS nella Biblioteca del Collegio. Le descrizioni relative ai codici civilistici sono state parzialmente edite a stampa da M.J. Peláez, 'Catálogo de la documentación' p. 485-548, specialm. 504-528]; I. Gil Fernández, De codicibus albornotianis ad graecas latinasque litteras pertinentibus commentarius, p. 6, 11; A. García y García e C. Piana, 'Los manuscritos filosófico-teológicos, históricos y científicos del Real Colegio de España de Bolonia', Salmanticensis 14 (1967), p. 82, 84 nu.23 [Inventario del 1453] 87, 92, 94s., A. García y García, 'El legado de libros del cardenal Gil de Albornoz al Colegio de España de Bolonia', Studi Senesi 84 (1972), p. 21, 38 nt. 57.</p>	<p>Unnumbered pages</p> <ul style="list-style-type: none"> ■ Copertina anteriore ■ Copertina posteriore ■ Foglio sparso ■ Foglio sparso ■ Foglio sparso ■ Foglio sparso ■ Foglio sparso ■ Foglio di guardia anteriore - recto ■ Foglio di guardia anteriore - verso ■ Foglio di guardia anteriore - recto ■ Foglio di guardia anteriore - verso ■ Foglio di guardia anteriore - recto ■ Foglio di guardia anteriore - verso ■ Foglio di guardia posteriore - recto ■ Foglio di guardia posteriore - verso ■ Foglio di guardia posteriore - recto ■ Foglio di guardia posteriore - verso ■ Piatto anteriore ■ Piatto posteriore
<p>2003-2007 © CIRSFID - Tutti i diritti riservati - All rights reserved - Col sostegno di Fondazione CARISBO e MIUR e per gentile concessione del Real Colegio di Spagna</p> <p>User: Not logged in Terms of use Privacy statement Language: Italiano</p>		

Figure 21: R2, codex page. On the right the list of unnumbered pages related to the codex is visible

	<p>Codex: 002</p> <hr/> <p>1 - Biblia vulgata (con i Prologi di S. Girolamo o a lui attribuiti, prima di ogni libro o di ogni gruppo di libri)</p> <hr/> <p><i>From folio 001r to folio 369r</i></p> <p>AUTHOR <i>Hieronymus (S.)</i></p> <p>EDITION <i>con varianti rispetto a Patrologia Latina 28, p. 177-184, 1389-1394; 29, p. 63-66; 28, p. 1503-1506; 23, p. 1063; 29, p. 403-428, 443s., etc.</i></p> <p>INCIPIIT <i>(del Prologo al Pentateuco; sotto Incipit epistola sancti Ieronimi ad Paulinum presbiterum de omnibus divine historie libris) Frater Ieronimus (marg. Ambrosius) tua munuscula perferens (ed. Patrologia Latina 22, p. 540-549); (ex abr. della Bibbia f. 4ra) e batur adiutor similis eius</i></p> <p>EXPLICIT <i>veni, Domine Iesu. Gratia Domini nostri Iesu Christi cum omnibus vobis. Amen.</i></p> <p>SEGUE <i>Explicit liber Appocalippsis et tota Biblia omnium divinorum sacramentorum et sacrarum scripturarum contentiva. Rubrica.</i></p> <p>NOTES <i>alcune glosse marginali e correzioni posteriori</i></p>	<p>Other works</p> <p>Other works belonging to the same codex</p> <ul style="list-style-type: none"> ▪ 2. Interpretationes hebr...
<p>« Previous 001r Next »</p>		

Figure 22: R2, the landing page for work 2 in Codex 002. Local work navigation and place-making is visible under the image and correlations are visible on the right.

The Annotation Engine

R1 allowed easy zoom in and zoom out over the high-resolution digital rendition of the pages via an external Java viewer: R2 was required to deliver a proper way to comment the codices and to make the requirements simpler to fulfill and more standard-compliant.

For this reason, a totally new Javascript engine for handling zooming and commenting was put in place, based on publicly available open source libraries.³²

The new engine, dubbed the annotation engine, allows not only seamless zoom in, zoom out, and scrolling of the high-resolution images (Fig. 23), but it also allows users to graphically select one or more areas on the image with the mouse, move them around, resize them at pleasure, and connect them to text notes which are then displayed on the side of the page (Fig. 24 and 25).



Figure 23: R2, the zoom area in high-resolution view with the action commands visible above

These notes are recorded in the database and connected on a 1:1 relationship with the page. They are also attributed with full authorship to the user who added them, and time-stamped, and can therefore be structured in a classic threaded conversation, with other users adding comments to a note. Notes are versioned as well, so later changes do not alter

³² The Javascript application was entirely coded by R. Riveret and L. Cervone as an extension of the Fotonotes library (www.fotonotes.net).

the original wording and it is always possible to accurately follow the course of a discussion as it develops (and developed).

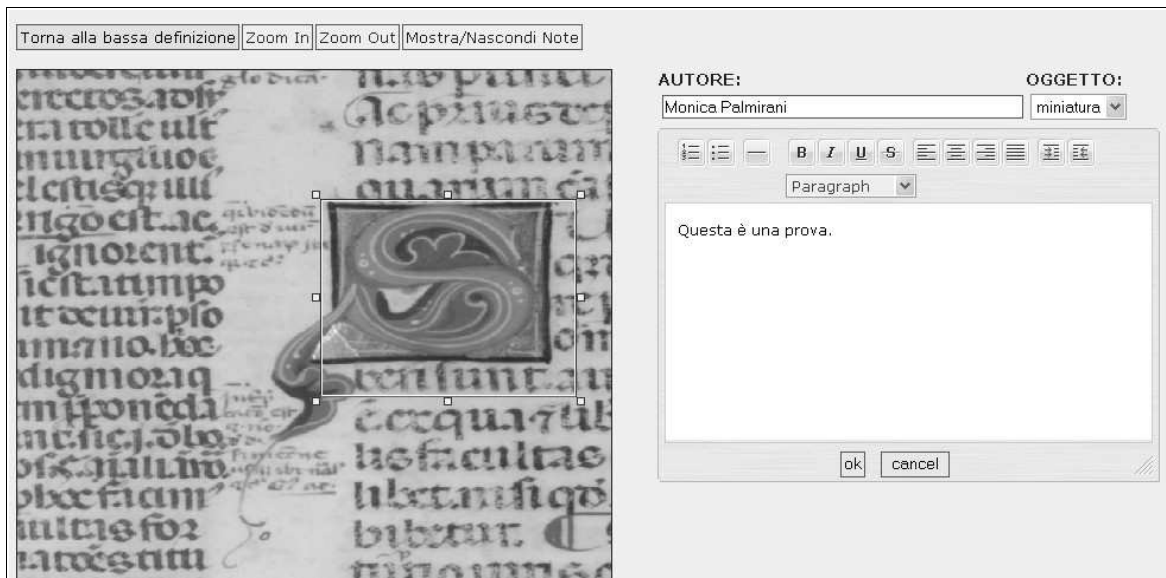


Figure 24: R2, the annotation engine with the note-taking tool ready for writing. Note the selection area surrounding the glyph in the image

The notes and their hot-spots can be hidden at any time for unhindered view over the image by means of the action menu available directly above the image.

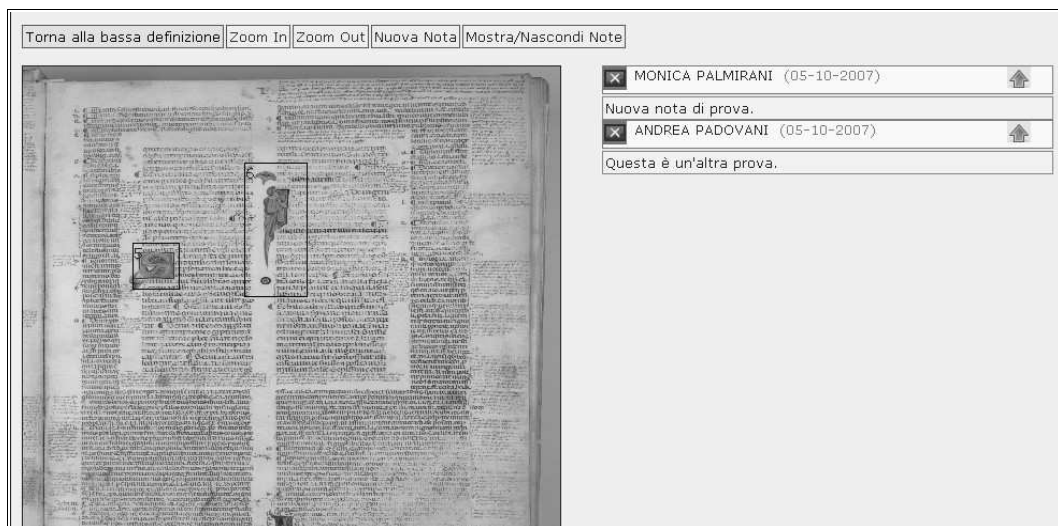


Figure 25: R2, the annotation engine with image notes displayed on the right

The annotation engine is a most remarkable tool, which somewhat makes the very idea of glossing once more a reality for manuscripts. Its distinctive and conversation-structuring

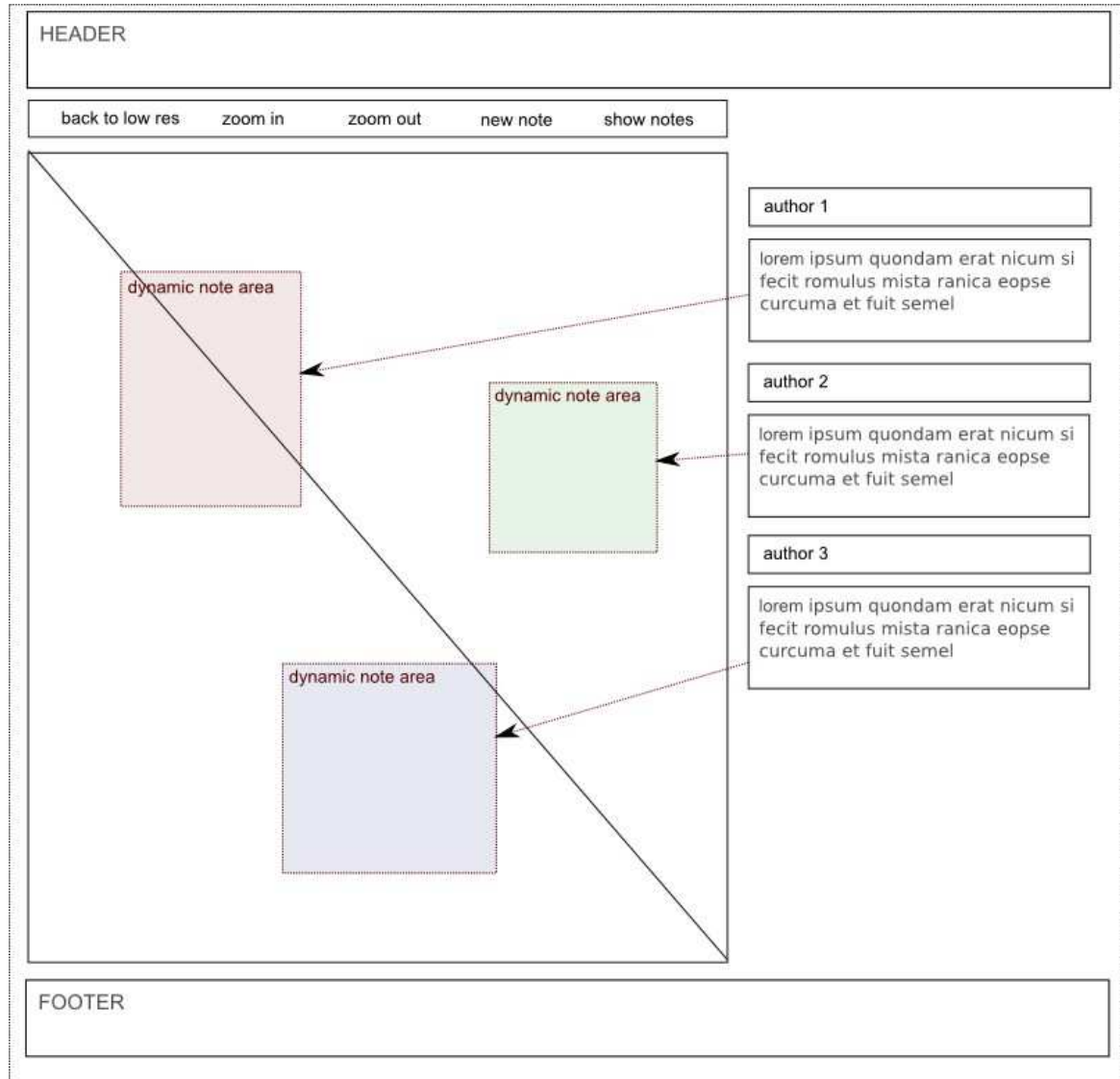


Figure 26: R2, wireframe of the annotation engine

features, such as the facts that notes are specific to areas of the image; that there can be more than one note per page; that notes are attributed with full authorship and timestamps; that users can comment notes, effectively starting conversations; that these conversations are saved and versioned.

The annotation engine also clearly demonstrated that the in many different ways, technical and architectural and conceptual, the boundaries for a proper handling of historical juridical manuscripts and their conversations had been reached within the current framework and that a new, integrated approach was required.

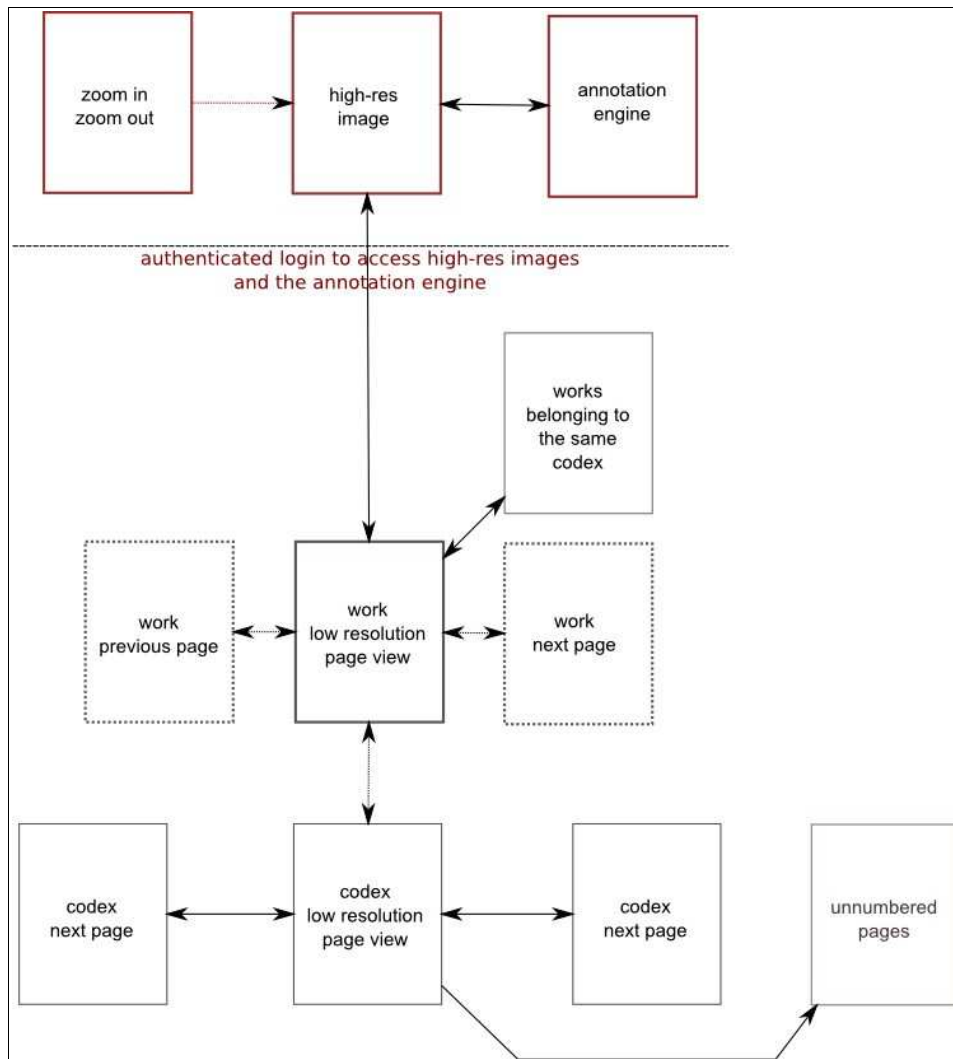


Figure 27: R2, general information architecture at the codex and work level

Searching the Innerio Archives

CIRSFID Innerio

Project

- Homepage
- About the project
- Partners
- Help
- Registration
- Login

Codices

- Codices
- Authors
- Search
- Search lemma

Contract

- Contratto
- Contract

SEARCH

It is possible to search for words and sentences in any work's incipit, explicit or title.

2003-2007 © CIRSFID - Tutti i diritti riservati - All rights reserved - Col sostegno di Fondazione CARISBO e MIUR e per gentile concessione del Real Collegio di Spagna

User: Not logged in Terms of use Privacy statement Language: Italiano

Figure 28: R2, the search page

Searching for content in R2 can be done either by means of the internal thesaurus based on code from the OpenThesaurus project, or via a Google-style autocomplete entryfield. Based on feedback that was received from project management, it was decided not to allow one single , but to preserve a distinction between searching for authors, which can only be done from the Authors page, and searching for all other types of content in the codices, which can be done via the Search page. To alleviate the problem , as soon as a user starts typing, the predictive search box starts to provide results from the database. The more a user types, the more precise the results and the shorter the list of matches.

The list, which can be seen in Fig. 22, is dynamic, so if the user keeps typing the result set is zoomed in (i.e. restricted), if the user deletes characters the list broadens accordingly or zooms out.³³

³³ Of course, on matches. To effectively produce either a broader or narrower selection, the string in the entryfield has to match a string in the database. For example, while typing *testamentum*, there might be no difference between “*testamentum*” and “*testamentu*”, but there could be between “*testamentum*” and

CIRSFID Innerio	
<p>Project</p> <ul style="list-style-type: none"> ■ Homepage ■ About the project ■ Partners ■ Help ■ Registration ■ Login <p>Codices</p> <ul style="list-style-type: none"> ■ Codices ■ Authors ■ Search ■ Search lemma <p>Contract</p> <ul style="list-style-type: none"> ■ Contratto ■ Contract 	<p>SEARCH</p> <p>It is possible to search for words and sentences in any work's incipit, explicit or title.</p> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">testamentum</div> <p>[Codice 123] Allegationes super materia iudaeorum <i>(Inc.) (sotto Allegationes Jacobi de Monello consiliarii et vicecancell ...</i></p> <p>[Codice 070] Allegationes super Trebellianica <i>(Inc.) (sotto, al f. 59r e 60v, Allegationes super trebellianica et an ...</i></p> <p>[Codice 070] Consilium de donatione et testamento <i>(Inc.) (marg. In donatione causa mortis iurata cum clausula rati et non ...</i></p> <p>[Codice 083] Consilium de executione legati testamentarii <i>(Inc.) Quidam Antonius Laudubii (?) in m^o.ccc.lxxxii. suum ultimum condi ...</i></p> <p>[Codice 126] Consilium de executione testamenti et de portione canonica <i>(Inc.) (sotto In Dey nomine amen. Hec est copia cuiusde (!) consilii pu ...</i></p> <p>[Codice 083] Consilium de executione ultimae voluntatis mutato valore rerum <i>(Inc.) Punctus talis est: Quidam nomine Antonius condidit testamentum e ...</i></p> <p>[Codice 070] Consilium de falso <i>(Des.) et dolose accessit ad eius testamentum condendum asserendo ipsam ...</i></p> <p>[Codice H] Consilium de fideicommissaria substitutione <i>(Inc.) (sotto De fideicommissaria substitutione et de intellectu I. Her ...</i></p> <p>[Codice 083] Consilium de fideicommissariis <i>(Inc.) Nutus fecit testamentum et codillos ut supra patet ...</i></p> <p>[Codice 070] Consilium de heredis institutione <i>(Inc.) (marg. Relictum omnium bonorum importat heredis institutionem) A ...</i></p> <p>[Codice 083] Consilium de institutione heredis <i>(Inc.) Nicolaus fecit testamentum in quo plura legata reliquit ...</i></p> <p>[Codice 083] Consilium de legatis <i>(Inc.) In nomine Domini nostri Yesu Christy amen. Factum super quo cons ...</i></p> <p>[Codice 122] Consilium de legato conditionali <i>(Inc.) Quidam nomine T. suum condidit testamentum in quo inter cetera s ...</i></p> <p>[Codice 122] Consilium de legato facto Fratribus Minoribus <i>(Inc.) In nomine Domini amen. Casus talis est. Magister Johannes de Mer ...</i></p> <p>[Codice 070] Consilium de legitima <i>(Inc.) (marg. De filia instituta herede in quota ita quod plus petere n ...</i></p> <p>[Codice 070] Consilium de poena falsi notarii <i>(Inc.) (marg. Notarius qui scribit testamentum eius qui non potest arti ...</i></p> <p>[Codice 083] Consilium de portione episcopi et parochi <i>(Inc.) In nomine Domini amen. Quidam faciens testamentum ... aliqua por ...</i></p> <p>[Codice 083] Consilium de probatione solutionis <i>(Inc.) Testator fecit testamentum ...</i></p> <p>[Codice 070] Consilium de pupillari substitutione <i>(Inc.) (marg. Nota de substitutione pupillari quando est mater in medio ...</i></p> <p>[Codice 070] Consilium de pupillari substitutione <i>(Inc.) (marg. In materia substitutionis pupillaris) In Christi nomine a ...</i></p>
<p>2003-2007 © CIRSFID - Tutti i diritti riservati - All rights reserved - Col sostegno di Fondazione CARISBO e MIUR e per gentile concessione del Real Collegio di Spagna</p> <p>User: Not logged in Terms of use Privacy statement Language: Italiano</p>	

Figure 29: R2, the search entryfield and the dynamic list of results

Fig. 23 shows an exact match on one single record for the string being typed in the entryfield. By clicking on the item in the list, the user is taken to the relevant work.

“testa”.

CIRSFID Innerio

Project

- Homepage
- About the project
- Partners
- Help
- Registration
- Login

Codices

- Codices
- Authors
- Search
- Search lemma

Contract

- Contratto
- Contract

SEARCH

It is possible to search for words and sentences in any work's incipit, explicit or title.

[Codice 083] Consilium de fideicommissariis
(Inc.) Nutus fecit testamentum et codicillos ut supra patet ...

2003-2007 © CIRSFID - Tutti i diritti riservati - All rights reserved - Col sostegno di Fondazione CARISBO e MIUR e per gentile concessione del Real Collegio di Spagna

User: Not logged in Terms of use Privacy statement Language: Italiano

Figure 30: R2, an exact match on one single item, Codex 83, Consilium de fideicommissariis

Synopsis of List Pages from R1

The image displays a series of screenshots from the 'Cirsfid' website, illustrating the structure and content of its list pages. The screenshots are arranged in a grid-like fashion, showing the layout of the site and the text of the pages.

The first screenshot shows the main navigation menu, including options like 'Home', 'About', 'Contact', and 'List of pages'. Below the menu, there is a section titled 'List of pages' which contains a list of page titles and their corresponding dates.

The subsequent screenshots show individual page content, including text, images, and links. The pages are organized into sections, with each section containing a list of items. The text is presented in a clear, readable format, with headings and sub-headings used to organize the information.

The final screenshot shows a page with a large heading and a list of items, similar to the other pages. The layout is consistent across the pages, providing a cohesive user experience.

Chapter IV

Modeling Complex Historical and Juridical Conversations

Current Challenges

“What I hope to do is change your thinking from ‘build it and they will come’ to ‘build it right and they will come back’.”

– Kim Guenther¹

The traditional library science-based model of archival for on-line consumption which is a large part of the on-line search experience in large catalogs is also the *de facto* model for the traditional digital library or digital archive.

These might be perfectly fit for a world of simple descriptors, but they are hardly satisfactorily for complex, user-manipulable or multi-layered architectures like those we supposedly should use for the digital rendition of collections.

Information architecture has had a large impact in making library cataloging and categorizing methodologies a standard part of the design of on-line repositories and web sites for improved user access to documents and information. and has managed through the years to reconcile the push and pull of third-order² user-generated-content and folksonomies with traditional practice (Hearst et al, 2003; Quintarelli et al, 2008) in almost every conceivable use or industry. Bussolon (2010)³ argues that

1 Guenther, K. (2000). The Evolving Digital Library. Computers in Libraries. Vol. 20. No. 2. Pp. 48-50

2 For a complete discussion of the challenges of designing categorization in the third order, see Weinberger, 2007.

3 In a conversation with the author and Luca Rosati while discussing Bussolon’s article for the Journal of Information Architecture. See references.

“(...) classical information architecture implicitly assumes an hypothesis dear to classical classification theory: in every domain there is one single rule to classify information. The information architect’s duty is to identify and apply this rule. But (...) classical classification theory has been radically contested in the cognitive sciences, and models such as the one proposed by Barsalou (2003) explicitly maintain that there is no single way to classify a domain, and that classification is dependent on the context, on circumstances, on people’s goals. The debate that has emerged within the cognitive sciences should probably suggest to information architecture a less prescriptive and certainly more user-centered approach.”

Nonetheless, active research being done from the perspective of contextual experience grounding and information architecture-informed design in connection with digital libraries is scarce. It is maybe important to remember that none of the case-studies examined during development of R2 presented anything more than a simple hierarchical system for navigation and, apart from a few offering download of materials or a user login because of limited access policies, very little user interaction.

As a result, the current on-line model for publishing digitized collections does not yet differ that much from the traditional archival form, and that amounts to the use of what Bussolon calls “classical information architecture”, often performed either unconsciously or naively: static descriptors and toned-down bibliographical records for single, simple artifacts, rendered in digital format.

Even the more interesting, visionary and wide-reaching endeavors, such as the Europeana library⁴ explicitly conceived to connect on-line the different cultural collections of Europe, do not provide much more than those (important) social navigation mechanisms already made popular by commercial web sites like del.icio.us⁵ or more recently Facebook⁶. Many efforts, among these for example existing digital libraries such as the Illuminated Manuscripts at Oxford or the Manuscriptorium I described in Chapter III, seem to

4 Europeana, <http://www.europeana.eu>.

5 One of first and most famous social bookmarking web applications for saving URL references in shared, taggable libraries. See <http://del.icio.us/>.

6 Currently the most popular social web site. See <http://facebook.com/>.

underestimate the need for good user experience; many hard-working standard-producing bodies seems to support the idea that going from average to good simply involves a change in the reference metadata model. Of course most on-line digital libraries and archives allow for complex research to be carried out, and some like the Innerio Archives allow greater interaction. Some but not all allow some kind of free-form or Amazon-like⁷ navigation of the content, not limited by the physical boundaries of the original support: the Innerio Archives R2 do precisely this by presenting either works by the same author, works within the same codex, or works from the same period.

Some like Europeana again offer support for user-generated content, advanced visualization, social navigation, or personalized paths. The Innerio Archives R2 takes this

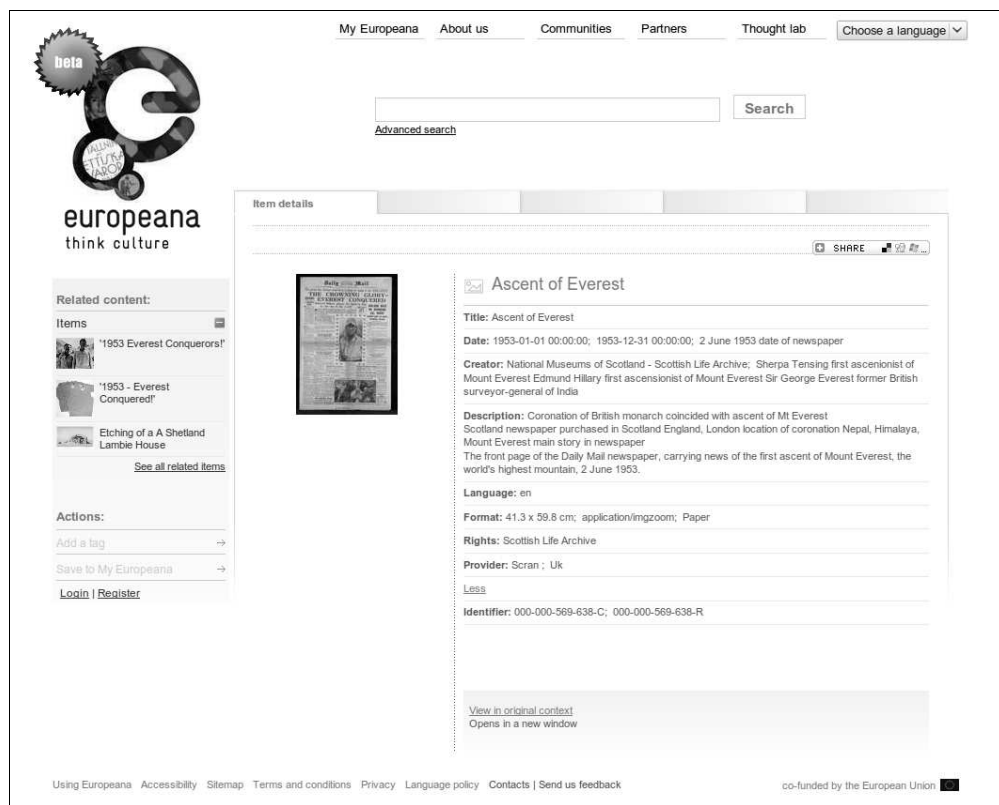


Figure 1: Europeana: related content navigation and social remediation tools are prominent elements

“if you liked x you will probably like y” suggestion engine which correlates items with past choices made by the same user, effectively suggesting books or other materials which are connected (in different ways) to the items currently visualized. See <http://www.amazon.com/>.

one step further by allowing direct manipulation of content via its annotation engine.

Very few though are concerned with the digital archive as a living, historical organism with its internal and external chains of causality and relationships. Such an organism should provide a satisfying user experience, require user involvement, and be capable to preserve an authoritative, or legally valid, chain of causalities for its artifacts: this is a strategic, core necessity for scholarly digital archives in the historical and juridical domain.

I see a striking resemblance between this, the necessity of moving from static records to managing complex information in digital libraries, and the gap that separates usability from user experience. Emotions, engagement, evolution through time, these were and are not accounted for in usability as much as complexity and the flow of historical and juridical discourse is not accounted for in the current batch of existing DLs.

Figure 1 shows a sheet, a recto from one of the codices in the Irnerio Archives. Even though it sure is not immediately apparent to the non-initiated to medieval manuscripts or paleography, the text proper, the original work so to speak, is just the two small areas at the center.

The remaining writings all around are glosses, comments, penned in later times⁸ by scholars and commentators. It might come as a surprise to anyone but historians that the real value of this page lies within the glosses, and in the small glyphs, signs and drawings that are not part of the original text. Sheets like this tell a story, weaved through time, sometimes long stretches of time, by different hands: that is their treasure.

8 Or immediately after, but in no way part of the original body of text and normally by different authors.

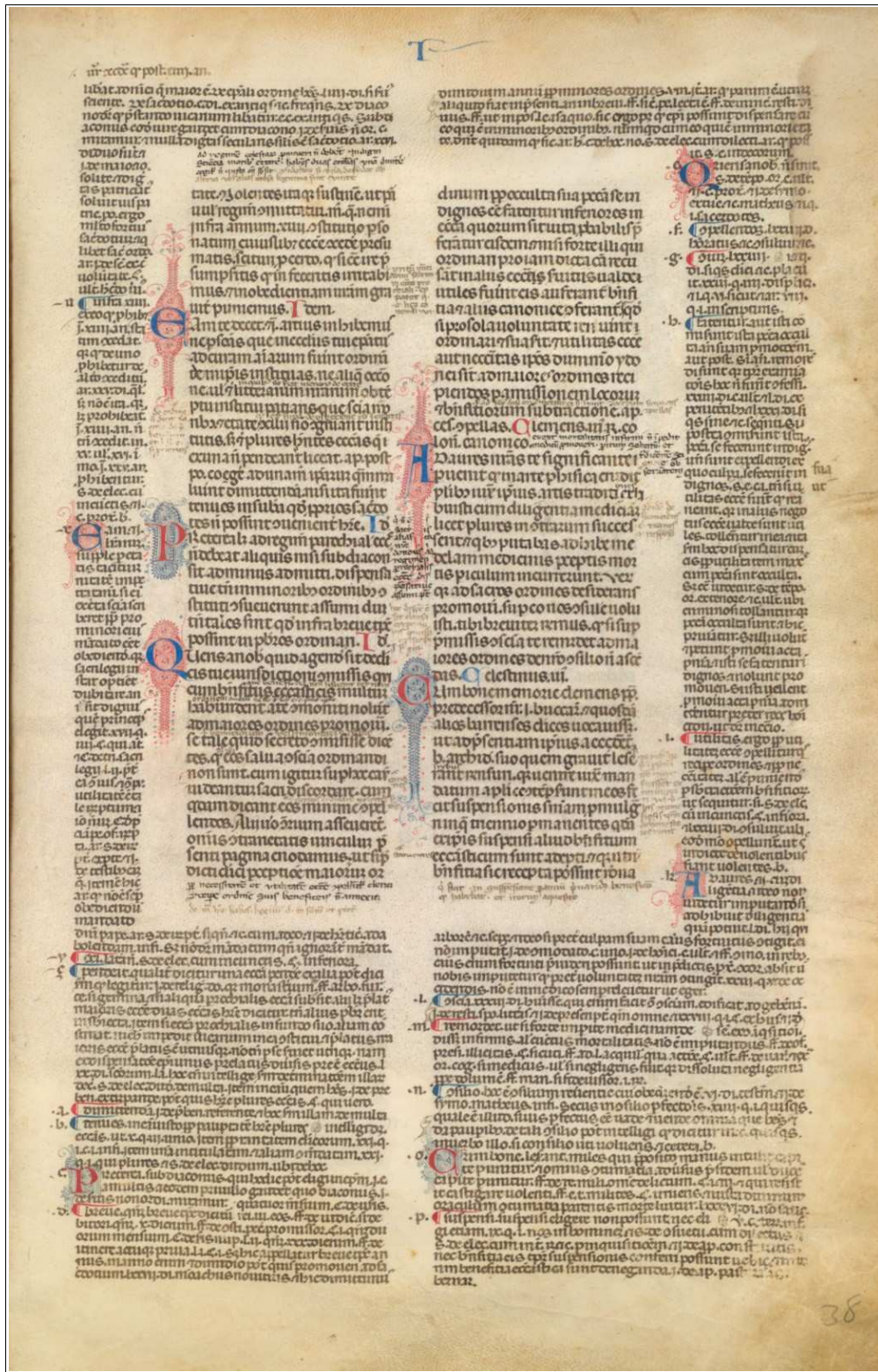


Figure 2: Recto sheet with glosses, Irnerio Archives.

The Authenticum Manuscript

An example of this attribution, linking, causality and complexity problem is perfectly described by Cristina Vano⁹ in a working paper commenting a compound work constituted by a medieval manuscript and its XIX century commentary. Vano describes two main elements:

1. a pre-accursian manuscript with Authenticum. This is actually a compound itself, being a multimedia artifact consisting of one photograph of the page, reference 12(24) Sankt Petersburg (item A.1 in his description), and a description of the manuscript by Loschiavo, preserved at Gosudarstvennyi Universitet Naučnaja Biblioteka, Lat. 5 SP (item A.2)
2. a hand-written note by F. A. Biener, dated XIX century, in SP 5. This again is a compound document, comprising a photograph (B.1), a description and attribution by Vano (B.2), and his philological transcription (B.3).

When this gets connected through the World Wide Web, an additional layer, the online layer, enters the picture. Vano describes it simply as link to content. But of course, even this simple link to content is a compound resource, and potentially some order of magnitude larger. Vano again identifies these elements in the Authenticum case:

1. Geschichte der Novellen (Biener 1824:bib. record) (C.1)
2. The text online and / or its rendition in PDF (or similar formats)¹⁰ (C.2)
3. additional bibliographical records on authors and / or manuscripts owners (C.3)

It is immediately evident how the historical and juridical value of this resource is only partially fulfilled by describing say item A.1, B.2, and C.1 using any current standard for

9 Vano, C. (2009). Elementi per una catena di più livelli d'informazione relative a un ms medievale e a un ms ottocentesco collegato. University of Napoli. FIRB.

10 For example and for this case: <http://books.google.it/books?hl=it&q=geschichte+der+novellen+biener>

metadata.¹¹ The real challenge and the real value reside in capturing the whole process in its development through time and space, and preserving its chain of causality. This is something that can only be rendered and modeled through a proper information architecture approach.

The Montecassino 266 Manuscript

Another very interesting example of the challenges involved in the rendition of complex, layered documents, is Bertram's hypertext rendition of his inquiries on the Montecassino 266 Manuscript (Bertram, 2009), again part of FIRB.

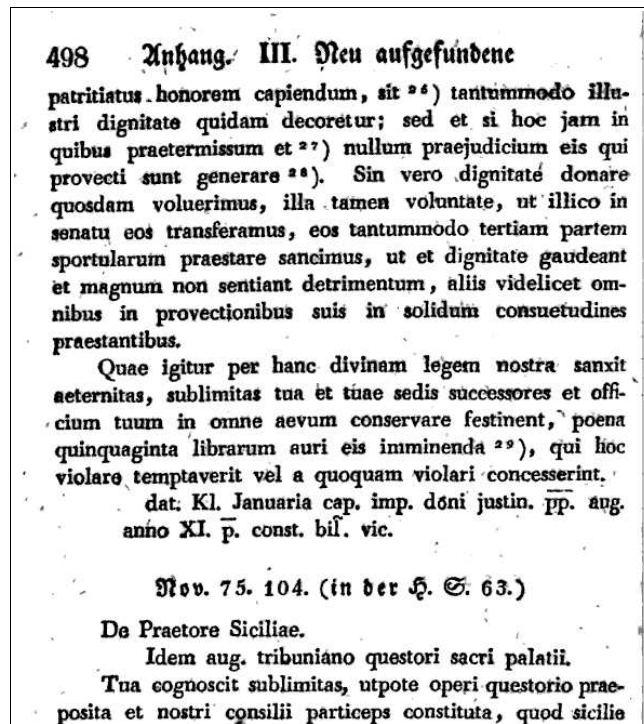


Figure 3: Page 498 of the *Geschichte der novellen Justinian's*, from Google Books

The manuscript, part of the *Corpus Iuris Canonici*, contains Goffredo of Trani's *Apparatus*

¹¹ A passing note might also be dedicated to pointing out that in this case, and this is not at all uncommon, there is no primary source either. The original document is not part of the chain, which relies entirely on the authority of the scholars involved.

decretalium, a gloss of pope Gregorio IX's decretales, and is as of today still unpublished. Historians maintain that a modern critical edition and a complete transcription are unrealistic goals because of the sheer complexity, and length, of the text itself. For this reason, releasing one of the manuscripts in a fac-simile format seemed the best possible solution. The Codex Cassiniensis, on sheepskin parchment, 300 pages dating back to the XIII century, reports Goffredo's works without the decretales, in a narrative form.

Bertram's commentary of the Montecassino 266 Manuscript commentary is itself a compound document, a hypertextual essay structured with links. Some of them internal, some external, to resources available somewhere on the World Wide Web.

The physical structure of this piece of critique itself does not lend to a simple rendition: how to exemplify links, back-links, and correlations using one of the current standards for the description of manuscripts? Basically, all of Bertram's work would be constrained in a single, all-containing text field. It would be either that way, or, treated as a separate work, it would lose all connection to the manuscript it glosses and definitely owes its existence to in the first place. Either lose complexity, and richness, and its implicit chain of reasoning, or lose relevance and correlations.

Early Findings and Four Basic Goals

In 2007, presenting some early findings connected to the release of the Innerio Archives R2 at the 2nd DELOS Conference on Digital Libraries in Pisa, Palmirani (2007) identified a number of central nodes that needed to be acted upon, and a number of related issues. Among the former were:

- ◆ the necessity to properly understand the specific requirements for the modeling of information systems capable of handling digital collections for historical and juridical documents;
- ◆ the preservation of the legal axiology of the documents;
- ◆ the modeling of juridical knowledge on said documents;

- ◆ the design of new capabilities for data and metadata correlation;
- ◆ the design of new interfaces for improved user access to these documents.

Some of the issues which could be foreseen included:

- ◆ legal challenges connected to digitization;
- ◆ security;
- ◆ integrity;
- ◆ authorship and IP rights management of the original work;
- ◆ author rights on user-contributed materials;
- ◆ on-line service modeling.

Among these latter, the ones concerning security, integrity and authorship are especially important, as they correctly imply that from a history of Law perspective descriptions, comments, and metadata are to be treated as being fully legitimate sources themselves and, to a degree, as fully autonomous works. Palmirani maintained that some of these challenges could be overcome by looking into new emerging standards which took into account collections and documents dynamics, such as FRBR for the modeling of complex relationships and digital signature techniques for the physical integrity of works, while others required a shift of perspective and a new approach which holds the preservation of a document's chain of causality as paramount (de Oliveira Lima, Palmirani & Vitali, 2008).

From the above, three basic goals or targets can be preliminarily outlined for information architecture modeling of historical and juridical manuscript collections:

1. dynamism: to move past the static bibliographical record to preserve the historical and juridical conversation commenters, historians and the flow of time have weaved around the artifact;
2. complexity: to render the multi-layered structure of complex and compound documents and artifacts, where the metadata is a work *per se*;

3. correlation: to link across the vast heterogeneity of descriptions, sources, formats, and connect across separated digital collections.

Point 1 is concerned with continuity through time: items in a collection are not simply photographed and crystallized in isolated points in time, they tell a story and that story is an important part of their value.

Point 2 is concerned with continuity through space. Different documents are often related by chains of causality, such as the different pieces that make up the complete picture of the Authenticum described by Vano. Similarly comments, like glosses or catalogs, or digital metadata, are necessary pieces of a historical and juridical view of the document without clearly being the document itself.

Point 3 is concerned with continuity through hyperspace, or Web space, and it is a core issue in the broader view of digital archives as a scientific part of the ongoing conversations happening on the Internet. Linking is what makes the Internet a different medium (Hinton, 2009). As we have seen from the brief examples of the Authenticum and Cassino, there is a rather large array of vastly different materials that require, or are thought to require, a number of different standards, conventions, and syntaxes to be described. Palmirani (2009) describes, for these examples only and hence with no pretense of exhaustivity, at least 15 autonomous artifacts:

a description of the digital manuscript	an article in the doctrine	books or parts of books
a description of the catalog	first-level metadata	e-books (critical editions)
the digital manuscript	a codex-related part	hypertexts
a PDF rendition of the XIX century edition	a description of the manuscript / work	author comments
a PDF rendition of the original handwritten card	a thesaurus	user comments and notes

These often are the abstract, descriptive layer over very different primary source

situations: manuscripts with accompanying records¹², records without manuscripts (catalogs), manuscripts without records. These all require different treatments. Issues pertaining to correlation will be dealt when discussing *correlation*.

Finally, a fourth issue is connected to the persistence of scientific documents on the World Wide Web, and will be examined in the context of the larger discussion on identifiers (Gomes & Silva, 2006) in connection with the different requirements imposed by a solid user-aware information architecture and by a legal informatics field of operations. This will be dealt with when discussing *persistence*.

Dynamism

Dynamism is continuity through time. Artifacts in a historical and juridical conversation have to be modeled and rendered as the complex objects they are, with their stratifications of changes. To properly address the issue, a brief introduction to some of the current, most widely used standards for manuscripts is a necessary precondition before moving on to the dynamic view brought on by FRBR.

A Brief Introduction to Standards

The data necessary for proper interpretation of digital formats is known as “representation information” as defined by the OAIS Reference Model (ISO, 2003). Two categories of representation information have to be present for effective digital preservation: representation information about the formats used to encode digital objects; and representation information about individual encoded digital objects. (Hedstrom & Lee, 2003). This is what is usually called metadata.

The former supports repository activities on digital formats at the general, higher level; the latter is concerned with the capture of “significant properties of actual instances of those classes” (Abrams, 2003). A typical example is that the first type of representation dictates

¹² This is the case with the Innerio Archives, for example, where the codices have been categorized and classified by Maffei. Maffei’s catalog provides most of the metadata in the Archives.

the paper formats and font schemes that can be used in PDF files, while the second details the actual paper format and fonts used in a particular PDF file.

Their use is different as well: representation information about formats is useful in developing systems that will process compliant digital objects, while representation information about these individual compliant objects helps application work-flows: for example, substitute font *X* with font *Y* on all PDFs of page size *Z*.

Abrams makes a point that digital artifacts are opaque objects, strings of bits, if there is no intervening mediation of a technical system that

“(...) is capable of transforming those bits into an appropriate representation interpretable by the human perceptual and cognitive facilities. This intermediation requires a detailed understanding of the particular means used to encode an object’s intellectual content into a digital form; in other words, its format.”

Understanding how these particular means work and how they are used to encode digital objects is then vital, as

“(w)ithout a thorough understanding (...) the long-term preservation of those objects is simply not feasible. Formats need to be understood in the general sense so that generic tools and services can be created. Specific instances of formatted objects must also be interpretable so that the significant properties of those objects can be retrieved.” (Abrams, 2004)

Finally, Abrams (2004) maintains that the idea of a standard format for representation of information in digital libraries

“permeates all technical aspects of digital repository architecture and is, therefore, the foundation of many, if not all, digital preservation activities. While analog content is generally directly usable by a human agent despite the passage of time (consider, for example, the ongoing accessibility and utility of 500 year-old incunabula), digital content is always dependent upon technical mediation for effective access.”

TEI

Historically, three standards have been developed and then widely adopted for describing manuscript resources through a machine-readable language: TEI, TEI/MASTER, and TEI P5 ENRICH¹³.

The Text Encoding Initiative (TEI) is long-standing organization pushing for open standards in the representation of texts in digital form (Bournaud, 2008). It was started in 1987 as a research initiative with the aim to “develop, maintain, and promulgate hardware- and software-independent methods for encoding humanities data in electronic form” (TEI, 2009), and later incorporated as an international membership consortium. TEI’s Recommendations and guidelines, currently over 1300 pages long in their printed form, span more than 23 chapters and define over 500 XML elements to distinguish or discuss textual phenomena and metadata. It must be remembered that TEI always sought after generality of application:

“The solution of the TEI, from the beginning, has been to create standards which are ‘broad churches’: which would accommodate within themselves many different kinds of records, from the most simple to the most complex. The case for this is particularly cogent for manuscript descriptions, where each country, each intellectual domain, even each cataloguer for each different cataloguing project, has developed a distinct cataloguing style. We have therefore sought, above all, for flexibility and for perfectibility in these proposals.”

TEI has also clearly stated that they are not seeking perfection, but flexibility and perfectibility. Flexibility

“permit(s), on the one hand, the making of simple manuscript inventories which contain no more than a list of manuscript identifiers, to which may be

¹³ TEI, TEI/MASTER and TEI P5 ENRICH are all XML-based standards. These are standards built for generic cataloging and conceived to be suitable for many different kinds of manuscripts. Early database-systems were conceived to exploit computing power to achieve precision inside a particular domain, and lie largely outside the scope of this research. For a comprehensive summary, see Stevens, W. (1991). *Bibliographic Access to Medieval and Renaissance Manuscripts: A Survey of Computerised Data Bases and Information Services. Primary Sources and Original Works. Vol. 1, No. 3/4.*

added just a few words of description of each manuscript, or an image of the manuscript. On the other hand, (...) to enable the most highly-formalised descriptions, with elaborate structural mark-up distinguishing the various elements within the description and containing complete manuscript transcriptions and complete digital facsimiles of the manuscript.”

while perfectibility accounts for the fact that

“(…) one might begin by making a simple inventory, and elaborate this by progressive addition of information. Or, one might begin by importing information from an existing manuscript database into unstructured paragraphs, which later workers might reformulate to distinguish statements of date, place, provenance, or description.”

TEI’s current incarnation, TEI P5, was released in November 2007¹⁴. TEI is definitely a complete and thoroughly defined standard, with a clear vision in respect to the narrow line any such effort is bound to walk¹⁵.

TEI / MASTER

In the mid 1990s, the booming World Wide Web and its wealth of text and images was a clear signal that an on-line catalog of manuscripts was indeed technically feasible, and that digital imaging could offer new unimagined ways to access and experience manuscripts.

14 According to the TEI web site, P5 provides important new features, including ” new support for manuscript description, multimedia and graphics, standoff annotation, and representation of data pertaining to people and places ; improvements to elements like <sic> and <corr> to provide more powerful methods of encoding textual alternatives ”, and, worth of note within the scope of this research, “changes to the way that linking mechanisms are expressed, so that pointing to other documents will be easier”.

15 “A design principle which emerged in the often-intense discussions surrounding these proposals concerned their architecture. Should we aim for an architecture which would permit the cataloguer to say almost anything, in almost any way imaginable? Or should we aim for a more formal prescription, which declares: if you want to say something about the binding of the manuscript (for example) you can only say it in a <binding> element within a <physDesc> element. The advantage of the first approach is that it permits the cataloguer near complete freedom; its danger is that the heterogeneity of descriptions might be at the cost of efficient retrieval based on predictable use of agreed descriptors: the prime justification for making the computer-readable records in the first place. The disadvantage of the second approach is that over-rigid formalism would lead to frustration among cataloguers, and (rather quickly) refusal to adopt the standard.” TEI. <http://www.tei-c.org/index.xml>. Accessed Jan 12 2010.

The missing piece was a language which could be used to describe manuscripts, and this could only be achieved after agreeing on the descriptions themselves: what is relevant, how they are to be structured, how they are to be accessed. As no such common vision existed, projects were founded to address the issue. Two American projects, the Electronic Access to Medieval Manuscripts project (EAMMS) and the Digital Scriptorium, started in 1996, brought together American and European experts and were instrumental in the creation of the European effort that was to become MASTER.

At the end of 1996, near Oxford, England, a workshop had representatives of a number of manuscript holding institutions in Europe and experts in metadata standards sitting at the same table.

The meeting was so successful that it was grown into the EU-funded project MASTER, Manuscript Access through Standards for Electronic Records, with the goal to develop an XML-based standard for the description of manuscripts. Once completed, this standard, called TEI/MASTER, was further elaborated by a group from TEI and finally added to the TEI chapter on Manuscript Description.

The basic innovation that TEI/MASTER introduced was a new TEI element, *<msDescription>*, explicitly aimed at manuscripts. The *<msDescription>* element may contain one, and only one, description of a manuscript. A composite manuscript is regarded as a single manuscript, with the distinct manuscripts which it contains each being held in distinct *<msPart>* elements, contained within the *<msDescription>* element. One single *<msDescription>* element may contain any the following elements:

<i><msIdentifier></i>	contains the identifier of the manuscript: where it is, and its exact shelfmark or other identifier. This element is mandatory.
<i><msHeading></i>	can be used to give basic information about the manuscript, such as might appear in summary catalogue (for example: author, title of work, date, place, language)
<i><msContents></i>	contains a description of the intellectual content of the manuscript, either as a series of paragraphs or structured into defined sub-elements

<physDesc>	contains a description of the physical aspect of the manuscript, structured into defined sub-elements
<history>	contains a history of the manuscript, structured into defined sub-elements
<additional>	contains additional information related to the manuscript but not part of its formal description (e.g.: bibliography; availability of facsimiles or images; curatorial information)
<msPart>	used for composite manuscripts. It may contain all the elements listed above except for <summary> (<summary> is intended for description of the entire manuscript alone). This element is required when dealing with composite manuscripts

Table 1: The elements in *mDescription*

TEI P5 ENRICH

ENRICH, European Networking Resources and Information Concerning Heritage, is a project funded under the EU eContent+ program in 2007, with the objective to provide “seamless access to distributed digital representations of old documentary heritage” in Europe, to foster a “shared virtual research environment especially for study of manuscripts, but also incunabula, rare old printed books, and other historical documents” (ENRICH, 2007).

The ENRICH project’s own format specification is based primarily on TEI P5’s chapter for *Manuscript Description*, but benefits as well from recommendations relating to the description of digital images, of non-Unicode characters, and of paleographic or transcriptional data. This was a necessary step as TEI is a very flexible and general scheme that covers bibliographic, linguistic, historical, literary and textual materials, and not simply manuscripts¹⁶. Furthermore, the ENRICH specification was structured so that it maintains full compatibility with the published TEI P5 standard:

¹⁶ As any application of the standard requires a customization by means of an ad hoc specialized XML vocabulary, TEI provided rules for its development in order to maintain a general compliance. Specifications need to be written, with the specific goal of documenting “which aspects of the TEI are being used”, and if “any additional constraints (are) required for the project, for example to determine the legal values of certain attributes, or to remove unwanted elements”. Finally, the specifications are used to generate formal XML schemas for document validation.

“A document prepared in conformance to the ENRICH schema is therefore also a TEI-compliant document, and can be used by any TEI-aware software. Furthermore, because of its use of the TEI, the ENRICH specification provides a complete suite of encoding possibilities, covering not simply the cataloguing and description of manuscripts or early printed books, but also the encoding of a digital edition in which metadata, digital image, transcribed text, edited text, and editorial annotation are all integrated in a standard framework. “

These standards are widely used. They solved many problems which were connected with the earliest database-driven systems¹⁷, thanks to the general availability as XML at the time the earliest EAMMS efforts were being carried out. Since the end of the 1980s, the TEI had created a set of guidelines for the encoding of a very wide range of materials in the humanities: these were instrumental in the process, and “a consensus developed, among the scholars, archivists and computer experts involved in the EAMMS, Digital Scriptorium and MASTER projects” that XML was the way to go because of two very simple reasons¹⁸:

“(a)t one end of the scale: SGML/XML encoding of very simply structured records accompanied by manuscript images would lend itself to immediate web mounting. At the other end: the most complex manuscript descriptions could be housed in SGML/XML encoding, without distortion of the description and with no compromise of scholarly detail. Similarly, the development of a universal interchange format in SGML/XML might permit decanting of the many manuscript records from the many databases into a

17 For example, database systems had issues with the compatibility of data across systems and the binding of particular databases to specific and perishable combinations of software or hardware.

18 “Several further factors contributed to the choice of the TEI implementation of SGML/XML, as the base for the proposals here outlined. Firstly, there was the perceived success of the TEI guidelines in addressing many of the issues, which would have to be resolved in any scheme of computer readable manuscript descriptions. Such a scheme would need to establish encoding for transcriptions of text from the manuscripts: the TEI had already set forward guidelines for transcription of primary textual material, and these guidelines had been well tested. A vocabulary would be needed for encoding of various kinds of textual division, for encoding of metadata about the description, and more: the TEI had already done much of this work. Secondly, the TEI workgroup system offered a well-proven means of bringing together domain experts and encoding experts, and harnessing their various skills towards a consensus which could form the basis of a widely-accepted standard. Thirdly, several of the people active in framing the TEI guidelines had strong interests in medieval manuscripts, and so were involved in the projects named above from their beginning.”

single, and therefore more malleable, form.” (Centre for the Technology and the Arts, 1999)

While true, this still does not account for a better understanding of how to structure complex information in such a way that its dynamic nature is conveyed and preserved and its users are provided with good user experience. TEI P5 ENRICH still falls short in the way it can handle all of dynamism, complexity, and correlation¹⁹: the model still pushes forward a library science approach which physically describes the book more than what the historian of law looks for in the book (Palmirani, 2009):

“TEI P5 ENRICH pur nascendo specificatamente per modellare i metadati del manoscritto e del manoscritto riprodotto digitalmente, non rispetta invece il flusso descrittivo e scientifico dello storico del diritto il quale spesso argomenta in modo fluido per condurre la sua tesi.”²⁰

This “descriptive and scientific flow” that needs to be preserved, modeled and rendered, is the core issue this research tries to address. Sources of historical and juridical value have a need to go past the static descriptors provided by standards such as TEI. I maintain that architecturally-wise the heuristic information architecture evaluation and consequent adaptation to user experience of navigation, labeling, grouping, and place-making can help provide a better sense of context and flow (McCulloch, 2005), on the technical, modeling front there seems to be enough promise for handling dynamic sources in a relatively new standard, FRBR.

FRBR

The Functional Requirements for Bibliographical Records, usually shortened to FRBR, is

19 For a balanced discussion on what scenarios might be appropriate for the use of TEI, see Lavagnino, J. When Not to Use TEI? [http://www.tei-c.org/About/Archive_new/ETE/Preview/lavagnino.xml]. Accessed Dec 15 2009.

20 Palmirani, M. (2009). FIRB. “The TEI P5 ENRICH model, although specifically built to handle manuscript metadata and digitally reproduced manuscripts, does not always follow the descriptive and scientific flow typical of historians of law, who typically argument their thesis and discourses fluidly”. Translation by the author.

an entity-relationship model developed by the International Federation of Library Associations and Institutions (IFLA) that presents a holistic user's perspective on retrieval and access tasks in online catalogs. By providing relationships between the entities FRBR provides links to navigate not only across the records but through the hierarchy of relationships as well (IFLA, 1998). Its basic elements are the result of logical analysis of data in current bibliographic records. (Pisanski et al, 2009).

FRBR has been well received, and while FRBR's abstract terminology may not be common and may require a different approach, many authors recognize three major advancements the standard brings forth: it provides a framework for more user friendly displays of bibliographic data; it represents a necessary break from the traditional usage of same appellations for different concepts (most notably edition); it turns linear library databases into a true network ed environment by relying heavily on correlations (Pisanski et al, 2009). This accords perfectly with the general principle to consider context a primary element of the information architecture (Resmini & Rosati, 2009b) and with the place-making heuristics I introduced in Chapter II.

FRBR works by providing an abstract framework which categorizes generic bibliographical entities in three different groups:

- ◆ Group 1 entities are Work, Expression, Manifestation, and Item. These represent the products of intellectual or artistic endeavor.
- ◆ Group 2 entities are person and corporate body, responsible for the custodianship of Group 1's intellectual or artistic endeavor.
- ◆ Group 3 entities are subjects of Group 1 or Group 2's intellectual endeavour, and include concepts, objects, events, places.

FRBR also defines four user tasks in connection with the above entities: find, identify, select, and obtain. Users should be able to perform any of these tasks on any of the entities or relationships. A fifth, unofficial, task is to navigate or relate (IFLA, 1998).

Group 1 entities, work, expression, manifestation, item, often collected under the

shortened name WEMI from their initials, are explicitly aimed at the unambiguous identification of artistic or intellectual artifacts, and they are the ones commonly used for information modeling in connection with digital libraries and digital archives. According to the FRBR specifications, Group 1 entities are:

- ◆ Work: a distinct intellectual or artistic creation;
- ◆ Expression: the specific intellectual or artistic form that a work takes each time it is realized;
- ◆ Manifestation: the physical embodiment of an expression of a work. As an entity, manifestation represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form;
- ◆ Item: a single exemplar of a manifestation. The entity defined as item is a concrete entity.

Hence, in FRBR WEMI notation, a codex part of the Innerio Archives, for example the *Summa Glossa*, could be be FRBR-ised as such:

Work	Summa Glossa
Expression	Many different renditions, one of them belonging to the Royal College of Spain
Manifestation	The TIFF or JPG rendition digitized by CIRSIFID and now available online
Item	The actual files stored in server space

The *Summa Glossa* (work) is realized (among others) by its manuscript copy in custody at the Royal College of Spain (expression), embodied in the digitized version hosted by CIRSIFID (manifestation), and exemplified by the more than 130,000 files stored on the Innerio Archives’ servers (items).

In this light, the Innerio Archives as a whole, as many other digital archives are, is a manifestation. Let’s take an in-depth look at the FRBR specifications. The U.S. Library of

Congress²¹ offers a lengthier description for Group 1 entities, with some useful insights: a work is

“(a) distinct intellectual or artistic creation. FRBR includes variants—such as revisions, updates, abridgements; added parts, accompaniment, transcriptions and arrangements of music; dubbed or subtitled versions of films, and translations—as expressions of the same original work. Modifications involving a significant degree of independent intellectual or artistic effort—such as paraphrases, rewritings, parodies, musical variations on a theme, dramatizations, adaptations, abstracts, digests, and summaries—are viewed as new works.”

An expression is

“(t)he intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms. Expression encompasses the specific words, sentences, paragraphs, etc. resulting from the realization of a work in the form of a text. Expression excludes aspects of physical form—such as typeface and page layout—that are not integral to the intellectual or artistic realization of the work as such.”

A manifestation is the physical embodiment of an expression of a work. A manifestation

“includes unique embodiments of a work (e.g., an author’s manuscript, an original painting), as well as embodiments produced in multiple copies, either for broad dissemination (e.g., a published book, a commercial sound recording) or for more limited purposes (e.g., a photocopy of an original typescript, a dubbing of a recording of an original piece of music). The set of copies produced in each case constitutes a manifestation.”

And finally an item is a single exemplar of a manifestation.

“The item may be a single physical object (e.g., a copy of a one-volume monograph, a single audio cassette), or it may comprise more than one

21 Library of Congress. (2003). Extending the FRBR model. Definitions from Table 1: FRBR Entities - Entity Definitions. http://www.loc.gov/marc/marc-functional-analysis/original_source/appendixd.pdf. Accessed Jan 26 2010.

physical object (e.g., a monograph issued in two separately bound volumes, a recording issued on three separate compact discs). ”

Complexity

“(C)onsider the categorization of certain highway vehicles as ‘oversized’. This means that they exceed some defined legal dimensional (size) limit. (...) Vehicles that are not categorized as oversized are constrained to specific well-defined limits of certain size characteristics. Conversely, a vehicle that has been categorized as ‘oversized’ may exceed the legal limit by a little or a lot, and may do so in one or more size characteristics. ”²²

Complexity is one of those concepts which through use have come to mean many different things. Mitchell (2009) defines complex systems as

“(l)arge networks of simple interacting elements, which, following simple rules, produce emergent, collective, complex behavior. ”

The salient elements are that these systems are networks, they have simple elements and simple rules, but produce complex behavior, behavior which is emergent, and collective. Although it could be said that the results of correlation between agents, or users, and the document as a system (Francke, 2008) is largely emergent and not controllable from the inception phase, this does not seem to be a definition that can be mapped easily to the specific issues that the handling of historical and juridical manuscripts presents to the designer. Goldstein (1999) maintains that emergence is “the arising of novel and coherent structures, patterns and properties during the process of self-organization in complex systems”. While self-organization is a feature of many social networks and one of the key elements described for folksonomies (Quintarelli, 2005), and hence an important element of applications that build on user interaction²³, the need for authoritativeness and top-down control in managing historical and juridical sources needs to be answered first.

²² (Rosen 1998)

²³ For example, tagging was originally considered a possible feature for the Inerio Archives R2.

Nonetheless, it is interesting to note that correlation is one of the main characteristics that Goldstein associates with emergence:

“(t)he common characteristics are: (1) radical novelty (features not previously observed in systems); (2) coherence or correlation (meaning integrated wholes that maintain themselves over some period of time); (3) A global or macro level (i.e. there is some property of wholeness); (4) it is the product of a dynamical process (it evolves); and (5) it is ostensive (it can be perceived).”²⁴

Rosen (1998) offers an interesting, relational view on complexity. A system is

“(...) simple if all its models are simulable. A system that is not simple, and that accordingly must have a nonsimulable model, is complex.”

Rosen ties his concept of complexity to modeling, and modeling is the act of establishing congruence between the elements and entailment structures of two systems, the object system and its model (Rosen, 1998). A simple model can be simulated, predicted:

“When a single dynamical description is capable of successfully modeling a system, then the behaviors of that system will, by definition, always be correctly predicted. Hence, such a system will not have any ‘complexity’ in the sense above, in that there will exist no unexpected or unanticipated behavior”

Complexity arises when a system behaves in ways that do not match the predictions of the models, and is therefore eminently rooted in comparison²⁵. Another definition, phenomenological in nature, maintains that complexity denotes systems which possess some or all of the following attributes (Johnson, 2007):

1. The number of parts (and types of parts) in the system and the number of relations between the parts is non-trivial – however, there is no general rule to separate

24 Goldstein, J. (1999). Cited in Corning, 2002.

25 To put it otherwise, Rosennean complexity is a relational property, not an intrinsic property, of a system (Rosen, 1978; Rosen, 1985). For more on Rosen and complexity, see also Panmère, http://www.panmère.com/index.php?page_id=16#sec1.

“trivial” from “non-trivial”;

2. The system has memory or includes feedback;
3. The system can adapt itself according to its history or feedback;
4. The relations between the system and its environment are non-trivial or non-linear;
5. The system can be influenced by, or can adapt itself to, its environment;
6. The system is highly sensitive to initial conditions.

Attributes 1, 2 and partially 4 are certainly applicable here. Complexity is also a bridge between dynamism and correlation: it needs to reflect both the evolving nature of a self-contained historical document and the way it relates to other relevant pieces of information which complement it, and these interactions are not easily ascertained.

Historically, standards for manuscripts such as TEI have been aimed at capturing static information: but as dynamic systems such as the Innerio Archives with its annotation engine have clearly showed, these documents can be evolving, living things, for which legal causality and authorship are important indicators.²⁶

Also, complexity as I discuss it here is a manifold issue: on one hand, it is “making the complex understandable” (Wurman, 2001); on the other hand it is the modeling of complex content or information for further consumption and use. While the latter is handled by the language used for modeling, the former is decidedly an IA task for which many of the observations carried out for the Innerio Archives still remain valid. The issue is systemic, but the solution is largely a case-by-case design built on solid IA heuristics and principles. Furthermore, complexity works internally, and externally. Different documents are often related by chains of causality, such as the different pieces that make up the complete picture of the *Authenticum* described by Vano. They are complex because of they are linked entities. Comments, like glosses or catalogs or digital metadata, are necessary pieces of a historical and juridical view of the document but they are clearly not

26 The evolution of TEI P5 ENRICH into a standard capable of managing historical and juridical manuscript sources respecting and modeling dynamism, complexity, and correlation is one of the goals of the FIRB project being carried out at the University of Bologna CIRSIFID together with the Universities of Rome and Naples. See also de Oliveira Lima, Palmirani & Vitali, 2008.

the document itself: this is internal complexity. The heterogeneity brought along by the different topologies of the many sources which can be networked together in a rich document, that is external complexity (Palmirani, Contissa & Rubino, 2009).

Correlation

The scenarios depicted in the examples from the Authenticum and the Manuscript 266 from Cassino are rather common in the reference domain of historical and juridical documents, where complex or compound sources are the rule, and they result in a number of different issues and scenarios that need to be handled. Briefly:

1. same-work correlations. The organization of correlations between the different related sources, or between different editions, or between copies. It is not at all unusual to have different copies of the same text actually become different versions because of the original editing or because of a particular set of glosses;
2. work-to-work correlations. The organization of correlations between different collections, physically or logically separated, but dealing with related works. These often have conflicting or different goals, or simply describe different items with few or none similar characteristics (say, for example, Julius Caesar's *De Bello Gallico* and a Roman vase from the 43 B.C.);
3. in-work correlations or reification. The presence of more than one author on the same page or sheet, where author A quotes author B who interprets author C.²⁷

These seem to be difficult to handle at both the descriptive and conceptual level. If well-structured data interchange standards for the domain, such as the already discussed TEI P5 ENRICH for example, were not built to address these issues, on the other hand information architecture models and tools built to address multi-layered complexity, such as folksonomies or faceted classification for example, still have, as I detailed previously,

²⁷ This is a typical situation in many medieval works and documented in the Irnerio archives as well, where Accursius is actually being commented by scholars through the rendition of his students and collaborators.

no real-world application nor theoretical framing in this context.

Relationships in FRBR

FRBR is built upon relationships between and among entities:

“(r)elationships serve as the vehicle for depicting the link between one entity and another, and thus as the means of assisting the user to ‘navigate’ the universe that is represented in a bibliography, catalogue, or bibliographic database.” (IFLA, 1998).

When dealing with relationships, it must be noted that a number of different terms are used by authors, creators and publishers of intellectual and artistic entities to identify the relationships existing between any given entity. A typical example might be the term “edition”, or the term “version”, but it is also very frequent to encounter phrasal terminology such as “edited by”, “based on”, “translated from”. As these are normally used as the base on which the cataloger will build a relationship to be recorded in the bibliographical metadata, a problem emerges:

“The problem with relying on commonly applied terms as a starting point for analyzing bibliographic relationships is that those terms are neither clearly defined nor uniformly applied.”

In its 2007 Report on FRBR, the International Federation of Library Associations and Institutions (IFLA), comments extensively on relationships. They act as the means through which the links between one entity and another are modeled, and thus represent the primary way to help the user to “navigate the universe that is represented in a bibliography, catalogue, or bibliographic database” (IFLA, 2007).

When querying one or more attributes of a given entity, or by simply browsing an entity’s record, relationships provide an additional layer of “(...) information that assists the user in making connections between the entity found and other entities that are related to that entity.” The IFLA Report describes a number of ways in which relationships can be

reflected in FRBR records: some

“are often reflected simply by concatenating attributes of one entity with attributes of the related entity in a single record. For example, a record will normally couple the attributes of a particular manifestation with the attributes of the expression that is embodied in that manifestation and with the attributes of the work that is realized through that expression.”

Others are referenced implicitly, by simply appending a heading to a record that identifies a related entity:

“The relationship of the work to the person or corporate body (...) is normally reflected implicitly by appending to the record a heading identifying the person or corporate body responsible for the work.”

A third way to reference relationships is “by layering attributes of one entity with those of related entities”, for example in a multi-level record that describes a compound entity and its individual components. Finally,

“(r)elationships are also frequently made explicit through the use of a note or similar device that indicates not only that a relationship exists between the entity described in the record and another entity, but also states specifically the nature of the relationship (e.g., Translated from the English text of the 1891 edition).”

One important feature of relationships is that both entities on each side of the relationship need to be explicitly identified (IFLA, 2008). Stating “based on a novel by John Steinbeck” does not instantiate a work-to-work relationship, but “based on “The Grapes of Wrath” by John Steinbeck does.

Persistence

Unlike hard copy or printed references, Internet references may change or become inaccessible. Close to 20% of Internet addresses in a Web-rich high school science curriculum became inactive between August 2002 and March 2003 (Dellavalle et al,

2003). Also, it seems that addresses in the .com and .edu top-level domains are the ones most frequently becoming inactive²⁸: if that can be somewhat expected of .com names, as they usually are connected with enterprises, companies, and professionals, it cannot but come as a surprise that .edu domains are equally fleeting and moving in and out of the Internet with surprising easiness.

Identifiers as a Social Construct

“What makes the bridge identifier persistent? It’s essentially a social construct. It’s not a technical thing (primarily). It’s not the paint the number is written in, or the bricks of the bridge itself, or the computer system at head office that maps the number to a map reference. These things help... but it’s mainly people that make it persistent.”

In a post on his blog on February 3, Andy Powell (2010) wrote these notes on how the understanding of the way URLs work is a fundamental node in the way persistence can be addressed on the Web.

Powell maintains that URIs are powerful tools because they are widely understood and makes a point towards their re-elaboration as the basis for a persistent addressing model. In a way, Powell maintains that the human-centered element should precede the technological, or formal, element.

Besides being the general point this research tries to make, and the general instance that an information architecture approach brings to the table, this is a view supported for example by Nielsen (2007).

28 See Dellavalle et al (2003). The study examined the frequency, format, and activity of Internet references in three high-circulation U.S. journals with scientific impact in the top 1% of all journals as rated by the Institute for Scientific Information (ISI) Journal Citation Reports (Science Edition of 2001). References were defined as a numbered citation appearing at the end of an article. The study proved that the percentage of inactive Internet references increased from 3.8% at 3 months to 10% at 15 months and to 13% at 27 months after publication. For articles 27 months old, the journals varied in a range from 21% inactivity to 11%. As said above, inactive Internet references were most commonly .com addresses (46% lost after 27 months) followed by .edu (30%), with .gov providing an interesting 10%. In terms of content, book reviews had the greatest loss (17%) and opinion and news articles the least (8%).



Figure 4: Andy Powell, Identifier for a railway bridge, UK. (<http://www.blipfoto.com/view.php?id=465380&month=2&year=2010>)

Nielsen is very well known for his heuristic principles for usability engineering of Web pages, and for favoring recognition over recall (Nielsen, 1997), as human beings are generally better at recognizing things they have previously experienced than at recalling those same things from memory. Nielsen maintains that task-based structures endure better than structures organized on content or organization (like departments inside an enterprise or university). Task-based navigation tends to facilitate ease-of-learning. In this light, many professionals in the IA-related field of Search Engine Optimization (SEO), the practice that works towards the maximization of a web site search impact, support the idea that URL structures should work as interfaces to the web site information architecture (Thurow, 2006)²⁹.

Interestingly enough, URLs and their relevance to information architecture is a vastly under-explored area in scientific literature. Rosenfeld and Morville (2002) mention URLs when introducing the importance to build context, but always from a single-website perspective. This might be a reflection of the social and technological reality of the Internet roughly ten years ago, but nonetheless the Polar bear book does not consider

²⁹ See also the chapter discussing the Innerio Archives.

outbound (or inbound) linking or referencing an important enough subject³⁰. Similarly, Kalbach (2007) only mentions URLs in reference to issues with navigation labeling.

Persistence in the Scientific Internet

The persistence through time of a document's location has been considered a strategic issue inside the scientific community as soon as the Internet became a viable tool for publishing and dissemination. In 1990 Engelbart, the father of the mouse and a pioneer of networking, maintained that in principle in an open hyperdocument system

“every object that someone might validly want or need to cite should have an unambiguous address.”

One year later, in 1991, Tim Berners-Lee wrote that the act of naming resources

“(...) is probably the most crucial aspect of design and standardization in an open hypertext system. It concerns the syntax of a name by which a document or part of a document (an anchor) is referenced from anywhere else in the world.”

It is plain to see why this necessity is important. If Internet references occur frequently and become often inaccessible within months after publication, even when published in high-impact U.S. medical and scientific journals (Dellavalle et al, 2003), then the general good and progress of the Internet as a viable scientific channel is hindered.

Short of a ban³¹ or severe restrictions to the use of references on the Internet, Dellavalle (2007) suggests that new systems like Digital Object Identifiers (DOIs)³², Uniform

30 For further support of this stance, see the chapter on Search, where again the entire design perspective is directed towards in-site search.

31 “Although examples of journals prohibiting the use of URLs in endnotes do exist (e.g., Cancer Research), the vast majority of medical and scientific journals, including those in this study, currently have no such restrictions. Prohibiting URLs is undesirable given that rapidly developing fields, such as human genetics, rely on extensive information available only on the Internet.” (Dellavalle et al, 2003).

32 See further on in the chapter. Correctly Dellavalle maintains that a further barrier to DOIs widespread, global use and acceptance is the cost for small publishers and individual authors.

Resource Names (URN), and Persistent Uniform Resource Locators (PURL) could be used to address the issue. More in general, Berners-Lee (1991) believes that a “hypertext link to a document ought to be specified using the most logical name as opposed to a physical address. This is (almost) the only way of getting over the problem of documents being physically moved. As the naming scheme becomes more abstract, resolving the name becomes less of a simple look-up and more of a search”.

On the other hand, this necessity of persistence does not reach to the business and professional communities: as much as reliable sources go, there seems to be a certain assumption in practitioners and managers alike that the very nature of the World Wide Web is somewhat transient and anyhow sufficiently resilient to provide alternatives to resources which get lost³³.

The very problem of persistence itself arises as URLs serve as both the identifier (name) and the location (address) of the information: but what are names and addresses?

For the purpose of research, and for the sake of simplicity, we will use the generic term of identifiers to address the different naming schemes, methods, and protocols which might yield at least some partial answer to this mostly academic problem. It must be noted though that in technical literature *identifier* bears a much more specific meaning which is tied to the URI – URL – URN discussion. Very briefly, it can be said that while a *name* is a logical way of referring to an object in some abstract name space, and *address* usually specifies a location, *identifier* or *unique identifier* generally refers to something which unambiguously references a resource, but has little significance in every other respect. That is, it is an arbitrary code, and what is called a name server, a resolver, is required to convert identifiers into meaningful addresses: for example, the string `1551166ABDF141AF56` might certainly be unique, but to be an useful identifier it requires a middle-layer which effectively couples it to some accessible resource (Berners-Lee,

33 Supporting this idea of fast access, extreme redundancy, and little preservation, a number of URL-shortening services have bloomed in recent years. These applications or web sites, for example <http://bit.ly/> and <http://ow.ly>, allow to transform cumbersome, long URLs into minimal compact strings. Then they provide the resolution service so that accessing the shortened URLs users get redirected to the original resource. For a brief overview of the issues involved, the Wikipedia entry on URL shortening is a good starting point (http://en.wikipedia.org/wiki/URL_shortening).

1991)³⁴.

For the purpose of communication, the basic infrastructure of the World Wide Web can be described as based on data formats and protocols, with identifiers coupling them in the middle (Connolly, 2005). The common example would be to think of XML or XHTML (data format), served via FTP or HTTP (protocols), by means of URLs (identifiers)³⁵.

URLs, URIs, and URNs

A Request for Comments (RFC) is a document or memorandum published by the Internet Engineering Task Force³⁶ (IETF), the volunteer body promoting and enforcing Internet standards, which describes “methods, behaviors, research, or innovations applicable to the working of the Internet and Internet-connected systems”. The IETF itself describes the development process of a typical Request for Comments and its complications in RFC 2026, sect. 1.2:

In outline, the process of creating an Internet Standard is straightforward: a specification undergoes a period of development and several iterations of review by the Internet community and revision based upon experience, is adopted as a Standard by the appropriate body... and is published. In practice, the process is more complicated, due to (1) the difficulty of creating specifications of high technical quality; (2) the need to consider the interests of all of the affected parties; (3) the importance of establishing widespread community consensus; and (4) the difficulty of evaluating the

34 It must be noted that this distinction becomes fuzzier and fuzzier as systems become larger distributed entities.

35 Connolly believes a certain fuzziness exists around the precise definition of URL, as these can be absolute, that is precisely pinpoint a single resource on the Internet like in <http://www.domain.com/path/to/resource>, or relative, concerned only with resources pertaining to a specific Internet domain, like in [path/to/resource](#). As this is primarily a terminological issue in communication between users, as no relative URL can be reached without an assessment of its position in respect to a given absolute URL, this research simply assumes all URLs to be either absolute URLs or irrelevant to its goals and scope.

36 From the mission statement at <http://www.ietf.org/>: The Internet Engineering Task Force (IETF) is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual.

utility of a particular specification for the Internet community.

RFCs go back to the end of the 1960s and Arpanet, and are today a conspicuous formal document series that is the backbone of the IETF standards process. Of the many thousand RFCs which have been published through the years only few have been accepted as full Internet Standards (IS). Among these, RFC793, defining TCP, RFC821 – RFC822, on the Internet Mail format, and RFC3986, *Uniform Resource Identifier (URI): Generic Syntax*, describing URIs, which was advanced to IS status in 2005.

Since URIs were designed to serve as both names and locations, when the IETF made them into a standard they became known as Uniform Resource Locators (URL), and a separate effort was dedicated to the standardization of Uniform Resource Names (URN) (Connolly, 2005). We will use URI and URL interchangeably.

According to RFC3986 an URI has three components: scheme name, domain name, and path. The scheme name corresponds largely to the protocol in use, the domain name to the registered domain, the path to an actual (or virtual) file pathname on a given server. If for example our URI is `http://www.domain.com/files/download/somefile`, *http* is the scheme name, *www.domain.com* is the domain name, and */files/download/somefile* the path.

These three parts are managed by different entities: the IETF manages schemes³⁷, ICANN³⁸ and national associates worldwide manage domain names, and single domain maintainers manage paths. Currently, all three elements are limited to using the basic ASCII character set³⁹.

While schemes and domains have to follow rules and obtain approval, URI paths are treated like typical file pathnames on a computer and fundamentally have no restrictions

37 The Official IANA Registry of URI Schemes includes current working schemes, like `http`, `https`, `ftp`, and `mailto`.

38 ICANN, the Internet Corporation for Assigned Names and Numbers. <http://icann.org/>.

39 It might be worthy noting, if passingly, that RFC3987 describes Internationalized Resource Identifiers (IRIs). IRIs are fundamentally URIs that use the whole range of Unicode characters, not just ASCII. RFC3987 states that any IRI has to have a corresponding encoding as a URI, in case an IRI needs to be used in protocols such as HTTP that only accepts URIs. Even though not an Internet Standard, this specific technology is maturing fast and actively being deployed.

on them if not those pushed forward by the underlying operating system. At the time the URI specifications were designed, in the late 1980s, different notations existed to access files remotely. One of them, *Angie-ftp*, combined host names, user names, and pathnames to allow for quick access: in a largely technical and engineering-oriented sub-culture, *user@host.domain.com:/path/to/file* was a much efficient way to access *file* than going through a number of iterations to specify host first, user second, path third, and file last.

The URI syntax that was developed for the Internet used the double-slash notation (*//*) for cross-machine naming, at the time used on a UNIX variant and allowing quick referencing of files on different servers, and introduced schemes, effectively opening up to a standardization and common use of naming conventions from different protocols. Common examples include the use of *http:* to access web pages⁴⁰, *ftp:* to access file-transfer protocol archives, and *mailto:* to access mailboxes.

In XHTML documents and resources, URIs are referenced by means of the anchor (*a*) element and its *href* attribute. URIs starting with the scheme and a colon are absolute; all others are relative. Relative URIs are relative to the current path of the document they are in. This is an important part of the discussion concerning persistence, and one that has more than one side to it: providing absolute URIs offers unique identifiers, while relative URIs might have more than one resource in different domains associated to them. In other words, while *http://www.website.com/file.html* is a unique resource⁴¹, */file.html* is not, as it might point to different files if moved to different domains.

As the problem of broken links (or link rot, as it is commonly called in Internet parlance) grew with the growth of the sheer mass of documents available and with the changes brought on by the passing of time, the academic community was faced with a new challenge. Initial approaches were prescriptive and down-to-earth, calling in to layman

40 Modern Internet browsers usually allow to skip the *http://* part of the URI, but that is simply a comfortable shorthand to alleviate typing tasks to users and not the correct way to reference a URI.

41 Of course, nothing is said about the real nature of *file.html* on *www.website.com*. The file could change every second, for example being a process log detailing access to server resources, or be totally different, unrelated files being served under the same address along a time line.

strategies to avoid the dreaded HTTP Error 404⁴². These went from practical instructions for naming, to considerations on the necessity of periodical checking, to the use of external resources like the Internet archives or even Google cache (Veronin, 2003)⁴³. We will deal with some of these in the section on URLs and Information architecture.

The IETF community on the other hand felt that the model was somewhat being misused, with web addresses behaving more like locations than names, and started to come up with solutions which were formalized in 1994 in RFC1630, Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as used in the World-Wide Web, RFC1738, Uniform Resource Locators, and RFC1737, Functional Requirements for Uniform Resource Names. Of these three, only RFC1737 carried some consequences and was followed three years later by Proposed Standard RFC2141, URN Syntax, which specified a new scheme, *urn:*, to be added to *http:*, *ftp:*, *rsh:*, and the rest.

RFC3986 clarifies the distinction between URI, URL, and URN in Sect. 1.1.3:

“A URI can be further classified as a locator, a name, or both. The term “Uniform Resource Locator” (URL) refers to the subset of URIs that, in addition to identifying a resource, provide a means of locating the resource by describing its primary access mechanism (e.g., its network “location”). The term “Uniform Resource Name” (URN) has been used historically to refer to both URIs under the “urn” scheme [RFC2141], which are required to remain globally unique and persistent even when the resource ceases to exist or becomes unavailable, and to any other URI with the properties of a name.”

It is well explained there that a scheme does not follow a either / or logic and instances of URIs from schemes may “have the characteristics of names or locators or both, often depending on the persistence and care in the assignment of identifiers by the naming authority”. Special care is also placed in the document to highlight that the term general

42 This is the HTTP error message for resource not found.

43 Interestingly enough, the long term perspective in Veronin is mostly tied to basic user-side procedures and to the acknowledgement of the role of the webmaster (sic) as a gatekeeper. There is no accounting in this vision for any inter-organizational rules from external bodies like the IETF.

URI should be preferably used in place of the more restrictive terms URL and URN.

PURLs

Because of the struggle over widespread acceptance of URNs, a mid-term strategy has been elaborated via the deployment of Persistent Uniform Resource Locators (PURLs). A PURL is defined by the OCLC as⁴⁴ a

“Uniform Resource Locator (URL) (i.e. location-based Uniform Resource Identifier or URI) that does not directly describe the location of the resource to be retrieved but instead describes an intermediate (more persistent) location which, when retrieved, results in redirection (e.g. via a 302 HTTP status code) to the current location of the final resource.”

An important point about PURLs is that they are considered to be an interim measure while Uniform Resource Names (URNs) are being mainstreamed⁴⁵ to help solve the problem of transitory URIs in location-based URI schemes like HTTP. As persistence problems are caused by the practical impossibility of every user having their own domain name, and the inconvenience and money involved in re-registering domain names, that results in authors putting their documents in “rather arbitrary locations of questionable persistence” (OCLC, 2004).

Technically, PURLs resolve exactly like SEF or SEO URLs⁴⁶. The official PURL HTTP server can be reached as purl.oclc.org as well as purl.org, purl.net, and purl.com. PURLs are organized into so-called *domains* like directory paths, e.g. [/net/scape](http://purl.net/scape) is the *subdomain* *scape* of *domain* *net*, and has itself *subdomains* like “about”. These *domains* are unrelated

44 OCLC is a nonprofit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world’s information and reducing information costs. More than 71,000 libraries in 86 countries and territories around the world use OCLC services to locate, acquire, catalog, lend and preserve library materials. From <http://www.oclc.org/about/default.htm>. Accessed Jan 19 2010.

45 The OCLC maintains that existing official PURLs on purl.org will probably be mapped to a URN namespace at a later date.

46 Many Mediawiki Wikis support PURLs in the net domain by shorthands like `[[purlnet:scape]]` for purlnet:scape as shown above.

to Internet domains; their purpose is to define one or more *maintainers*. The maintainers can grant write access to all or to other registered users: for example the domain *net* is open for all registered users to use.

The PURL server, or *resolver* in PURL terminology, first tries to match a request directly to a defined PURL. If the PURL exists the reply is a redirect to the last URL associated with it as specified by its maintainer. This can be another PURL, any http-URL, or, in fact, any URL. It's solely the job of the maintainer to guarantee that target URLs do in fact still exist.

Because PURLs are designed to be persistent, deletion is not supported. Nonetheless, PURLs can be disabled. A special parser operates when no direct match returns positive results on a given PURL: truncation is performed right to left to try and get partial redirects, the longest one of which wins. URLs of partial redirects do not necessarily end with “/”.

PURLs also adhere more strictly to RFC 3986 than ordinary HTTP servers⁴⁷: “/path/to/file” and “/path/to/file/” are usually treated as two different URLs pointing to different PURLs, with the latter being typically a partial redirect.

Practicalities of Persistence

A natural tension exists between persistence and availability. A file made available to someone via the World Wide Web relies on a complex infrastructure of layers. A web server runs on the host and provides a URI which is tied to the host name, the local path, and the filename. Different actors act on the different parts of the identifier. And for identifiers to work, there has to be some level of agreement about what is being identified, at least by all the parties that need to make use of it (Powell, 2010). This is usually achieved by way of dereferencing, that is, a “widely agreed way of moving from the

⁴⁷ It is customary to silently add a final trailing “/” to URLs or to strip it away according to server policies, effectively making them one single URLs. This is not required by RFC 3986, which allows these URLs to refer to different resources.

identifier to the thing being identified”.

This is why persistence is of the utmost importance in the cultural and digital heritage areas, as one might expect to be able to refer to things for very long periods of time, usually much longer, in the hundreds of years, than what new technologies and new tools like the Internet consider an affordable time-span. Thinking about making identifiers work persistently for more than a hundred years is a non-trivial task.

Think of a typical Web URL like this one: <http://www.domain.com/get.php?id=1455>. As I explained, we have three different parts, two of which lead to two different issues:

1. the scheme part (<http://>)
2. the domain name part (www.domain.com)
3. the filepath part ([get.php?id=1455](http://www.domain.com/get.php?id=1455))

While 1 is a given standard, 2 and 3 are problematic. Part 2 is problematic as it is socially connected to a resource which might change for reorganization purposes, like a single host changing its name from www to www2, or for business purposes, like a domain not being renewed and left to expire, with the consequence of making an entire tree of documents unavailable. Part 3 is problematic as it directly refers a programming language, PHP in this case, for the resolution of the filepath into a meaningful resource. It is difficult to imagine such a specification to be lasting enough to actually guarantee persistence. For example, apart from taking into account notions of *urls-as-interface* as previously discussed, URLs like those encoded for the *Irnerio Archives R2* get rid of the second problem⁴⁸, but not of the first, which is tied in that case to the naming conventions for servers and at the University of Bologna.

As a final note, it is interesting to remark that once more the OCLC describes its URI model as part of an overall “information architecture”:

“Uniform Resource Identifier (URI) provides a simple means for identifying

48 See Chapter III for details.

a resource within the Web global information architecture. Each URI begins with a scheme name that refers to a specification for assigning identifiers within that scheme.”

Again, it seems we are in presence of that “second” meaning I discussed in Chapter I, concerned with the information systems approach, and not with the overall user experience. Nevertheless, be it either considered from IS or IA, URL design is a largely underestimated practice in current digital library design, as documented in the case studies in Chapter II. On the other hand, it was a large part of the place-making strategies of the Innerio Archives R2 application, and the empirical results from the development show that static intervention on traditional, classical information architecture touch-points is not enough.

Open Issues

Despite an evident tension between persistence and availability, a good URI should actually work on both levels: it should be a persistent name, and an available location. Proponents of the *urn:* scheme argue that this tension is irreconcilable within the distributed framework of HTTP and DNS. The world changes continuously, and keeping things in sync takes effort. Most of the time, the hierarchical nature of DNS naming is convenient, but it concentrates a lot of power in one place and raises challenging governance issues. Peer-to-peer designs such as distributed hash tables may eliminate some of the centralization issues with DNS, but they introduce new governance issues.

Nonetheless, of extreme practical and theoretical importance in this issue is the fact that a number of cutting-edge developments show how new protocols can be used to service existing *http:* names, adding value to the existing hypermedia network: this seems more likely to succeed than the deployment of entirely new schemes for anything that is similar to the core HTTP’s GET/PUT/POST/DELETE operations.

Shortcomings of FRBR

Some authors maintain that the FRBR model might provide to be more satisfactory in some areas than in others, probably reflecting its declared status as “a useful starting point” for “further analysis” (Le Beuf, 2005).

Although it might be argued that FRBR’s theoretical model of relationships offers an interesting starting point that opens up digital library information design to information architecture methodologies and principles, some maintain that this very model is still to be interpreted and applied to cataloging rules. As such, each implementation will be dependent on system design. That means that unless there is a clear agreement on what FRBR represents in cultural sense, which is unlikely to happen, the model is open to different usages within different communities (Pisanski et al, 2009).

This is of course not bad *per se*, and in a certain way is the natural outcome of the more heuristic, design-related part of the process. But on the other hand if the quality of bibliographic records is inconsistent or bibliographic data is missing, as it frequently happens with manuscripts and generically speaking with historical sources, these authors consider extracting any structure, FRBR-like or other⁴⁹, an impervious task. As many FRBR prototypes are of union catalogs, consistency of data and, consequently, usability of these prototypes is further diminished (Pisanski et al, 2009).

I will point out that Pisanski seems to move inside a classical information architecture paradigm, where a structure, one structure, is there and has to be “extracted”. If the data is not good data, this is not possible. He also maintains that this practically impairs the “usability” of the digital library: I disagree. Imprecise or missing data might render designing the information architecture of a digital library complicated, but nowhere near

49 Not surprisingly, Pisanski et al also report that the first experiments in FRBR re-cataloging were conducted on Finnish and Norwegian national bibliographies (Hegna and Murtomaa, 2002). National bibliographies offer the best possible catalographic conditions for a move, being usually complete and thorough. Still, there were issues related to the correct identification of entities, as even though the elements of the FRBR model are usually recorded in bibliographic data, the rules of cataloging do not favor them, and “important information is often recorded in a way that is intelligible to humans but not computers”.

to impossible. This is simply because the process certainly relies on the data, but it is neither automatic nor *in re*: the cataloger makes the library (Resmini & Rosati, 2009a).

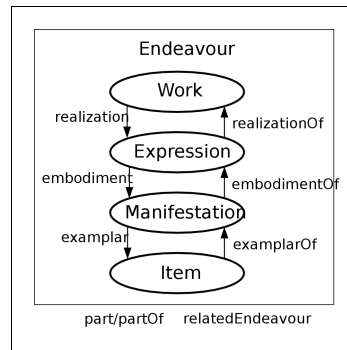


Figure 5: Voss, J. Basic Group 1 entities and relations of the FRBR model

To all extent, adopting a conceptual model that has the potential to change the nature of national bibliographies and other bibliographic databases and make them more user-friendly, brings in plenty of complications, mostly stemming from two different perspectives: technically, from the way existing bibliographic data is recorded; design-wise, from the lack of a real information architecture process and awareness in the field of digital libraries in general. Nevertheless, even those more skeptical of the standard consider that the only feasible solution to making legacy data compatible with any new data is frbrisation (Pisanski et al, 2009).

Conceptual issues with FRBR and URIs

Sauermann and Cyganiak (2007) provide an extended discussion of the issue concerning where the boundary between web documents and real-world objects might lie. If the HTTP response to a GET is 2xx, then the resource identified by the URI is an information resource. This is the only response code which allows the user to make a positive inference about the resource type being sought after. The URI owner, on the other hand, needs to make the choice, for each resource, whether to provide a representation or not, based on their understanding of the nature of the resources they are exposing on the Web.

What does this come up to in the case of FRBR Group 1 Entities? When exposing on the

Web a set of resources based on the FRBR model, are there any general rules for which of these resources are representable in one way or another?

In the case of a FRBR Work, it seems fairly clear that this is a conceptual resource, like an idea. For Sauermann and Cyganiak, it is a *real-world object*, albeit an abstract one, and not a *Web document*. On this basis, while a FRBR Work may be identified by an http URI, an HTTP server should not return a representation and a 200 status code in response to a GET for that URI, though the server may provide access to a description of the Work, a Web document itself identified by a distinct URI (Sauermann & Cyganiak, 2007).

A similar argument can be made for Expressions, as they too are an abstraction, albeit a more specific one than Works. Again, they fall into the category of real-world objects, and again, while an Expression may be identified by an http URI and an HTTP server may provide access to a description of an Expression, it should not provide a representation of an Expression.

In considering the other two FRBR Group 1 entity types, Manifestation and Item, it is perhaps easiest to consider the application of FRBR to physical resources and to digital resources as two separate cases, where Items can help understand Manifestations. The FRBR definition of Item is very much grounded in the physical, mirroring those real world objects libraries have been concerned with managing.

From the perspective of Cool URIs, then, these real-world objects may be described by Web documents, but they are not themselves Web documents. So a physical Item may be identified by an HTTP URI, and an HTTP server may provide access to a description of such an Item, but it can not provide a representation of it. Then, Manifestations deal with physical form, but they are still an abstraction: their role in FRBR is to capture characteristics that are true of a set of individual Items which exemplify that Manifestation (even in the case where a unique Item which is the sole exemplar of a Manifestation). Seen in this light, then, a Manifestation also falls into the category of things which may be described by one or more Web documents, but is not itself a Web document (Sauermann & Cyganiak, 2007).

It is probably worth observing that although the FRBR specification contains some references to electronic resources, the standard does not address the digital specifically:

“the coverage of digital resources in the text (is) very limited, and indeed the introduction acknowledges that the dynamic nature of entities recorded in digital formats is one of the areas that require further analysis”.

The step required to adopt such concepts as Work and Expression into an all-digital view is small, as both are independent of the form in which content is embodied. Items are more difficult to pin down, though, since FRBR focuses exclusively on the physical in its discussion of an Item⁵⁰. If we are treating a FRBR Item as a Web resource, is that resource an information resource, a representation of which can be provided? From a Web perspective, it seems that in this case “all of their essential characteristics can be conveyed in a message”.⁵¹

Finally, the concept of a digital Manifestation is the most difficult to categorize. On one hand FRBR states that a Manifestation is “an abstraction corresponding to a set of objects with the same characteristics of both form and content”.

On the other hand, it seems that one could argue that for Manifestations in digital form, it is true that “all of their essential characteristics can be conveyed in a message”, since the notion of Manifestation encapsulates that of specific intellectual content “embodied” in a specific form (Sauermaun & Cyganiak, 2007).

50 See Floyd, I. & Renear, A. What Exactly is an Item in the Digital World?. Proceedings of ASIST 2007. The authors speculate that in the physical world the thing which carries information is the same thing for which information managers typically describe characteristics such as provenance, condition, and access restrictions - the attributes of the Item in FRBR. In the digital world, this is no longer true: information is carried by the physical state of some component of a computer system, something the authors call an instance of patterned matter and energy (PME) - but information managers rarely concern themselves with managing such entities and recording their attributes. Entities such as a file, however, are the focus of management and description - but a digital file isn't really the concrete entity that FRBR calls an Item. Two approaches to the Item are possible, then: the Item-as-PME approach, which maintains that a fundamental aspect of being an item is being a concrete physical thing, or the Item-as-file approach which adopts the pragmatic position that items are the things, whatever their nature (physical, abstract, or metaphorical), which play the role in bibliographic control that FRBR assigns to items.

51 The Scholarly Works Application Profile takes this approach: the copy of a PDF document available from an institutional repository server, or the copy of an mp3 file constituting an episode of a podcast, is

Table 1: Correspondences and mappings between FRBR concepts and Web / HTTP behaviors and events. Sauermann & Cyganiak, 2007

FRBR Entity Type		HTTP Behavior
Work		Identify using HTTP URI Provide description of Work. Either use a hash URI, or respond to GET with 303 See Other and http URI of description (generic or using content negotiation).
Expression		Identify using HTTP URI Provide description of Expression. Either use a hash URI, or respond to GET with 303 See Other and http URI of description (generic or using content negotiation).
Manifestation	Physical	Identify using HTTP URI Provide description of Manifestation. Either use a hash URI, or respond to GET with 303 See Other and http URI of description (generic or using content negotiation).
	Digital	Identify using HTTP URI Provide description of Manifestation. Either use a hash URI, or respond to GET with 303 See Other and http URI of description (generic or using content negotiation).
Item	Physical	Identify using HTTP URI Provide description of Item. Either use a hash URI, or respond to GET with 303 See Other and http URI of description (generic or using content negotiation).
	Digital	Identify using HTTP URI Provide representation of Item. (Respond to GET with 200 and representation).

Finally, there is one issue which might arise in connection with Content-negotiation. Content-negotiation is a mechanism defined in the HTTP specification⁵² and used by HTTP servers to serve different versions of a resource from the same URI, so that user agents can specify which versions work best for them. It is typically used to serve different image formats (JPG or PNG, for example), or to serve content in different languages. The use of HTTP content negotiation on the Web introduces a dimension which does not sit

the FRBR Item.

52 Content-negotiation is described in RFC 2616, HTTP Protocol, in Section 12. For more on this see <http://www.w3.org/Protocols/rfc2616/rfc2616-sec12.html>.

very easily within the FRBR model. Using Content-negotiation, a single resource on the Web can be exposed using one single URI, but the server can be configured so that, depending on whatever factors are relevant, that URI returns different representations of the resource (Johannesen, 2009). These representations, as I said above, can vary for example by media type or language, and from an FRBR perspective, such variations might be considered entirely different Manifestations (media) or different Expressions (language)⁵³.

DOIs and PURLs

Both the PURL project and the Digital Object Identifier⁵⁴ (DOI) systems present different approaches to the persistence problem. As I outlined, a Persistent URL (PURL) is an ordinary HTTP URI in a domain backed by a strong persistence policy. An example of this is the purl.org domain itself, directly run by the OCLC. At purl.org the basic mechanism is that anyone can apply for an account and administer their own set of PURLs as content published on an ordinary web server and then redirected via HTTP to PURL.

This indirection level from PURLs to less-persistent HTTP URIs is very much like the indirection provided by DNS in coupling host names with IP addresses, except that the source and the destination of the redirection have a similar nature, ie that of URLs. A PURL such as <http://purl.org/net/persistent> can be used and accessed like any other URI. Even more important, from a user-centered point of view, the people who receive it see it and use it as just like a run-of-the-mill HTTP URI.

The DOI system on the other hand uses its own scheme, in the format `doi:11.145/556`. To support this scheme, ordinary Web browsers need a software plug-in. The DOI foundation provides policies, registration services, and HTTP redirection services similar to PURL

53 In a message to the NGC4LIB list, Alexander Johannesen pointed out that “content-type has a significant role in what the meaning of the URIs resolved resource is. If the content-type is something more like XML, the anchor - being at the mercy of the server - is a different resource all-together. The only reason HTML content-types returns the *same* as with an anchor is because we have decided that special case for HTML and browsers.” See References, Johannesen, A. (2009).

54 See Open Issues, further on.

providers like OCLC. Nonetheless, while the DOI foundation supports an alias for each DOI in the form <http://dx.doi.org/11.145/556>, the DOI Handbook states that this system has “significant disadvantages when compared with the resolver plug-in.”

A Few Final Notes on the DOI

“The only good long term identifier is a good short term identifier” – A. Powell

As I explained above, a Digital Object Identifier (DOI) is an alphanumeric string used to uniquely identify a given electronic resource which remains consistent and fixed as long as that resource exists, even if its URL changes. To achieve this, the DOI System provides a mechanism⁵⁵ for locating an up-to-date URL for a document from its DOI, and for associating other forms of metadata with an object. This way, naming a document by its DOI provides a more stable mechanism than URLs for linking to online content. DOIs are especially used for the referencing of scientific publications on the Web. Unfortunately, DOIs work using their own scheme, [doi:](http://doi.org/), which is not compatible with HTTP.

This is a major drawback. Coupled with the endorsement of multiple ways of encoding the DOI as a URI, and in their scarce capability to be good short term identifiers, regardless of their diffusion in scientific communities DOIs are ill suited to resolve correlation (Powell, 2010). While Powell maintains that persistence is “largely a social issue rather than a technological one”, it is plainly obvious that bad or badly used technology can get in the way of persistence (Nielsen, 1997). Powell (2010) believes that the DOIs could benefit from a step back and a good look into the principles of Linked Data and its vision that [http:](http://) URIs should be used.

For the scope of this research, this would be an important contribution to the ongoing discussion, after all when Berners-Lee imagined the Semantic Web he imagined a world of links explorable by persons, something decidedly not easy with DOIs, and machines, and

⁵⁵ The DOI System is implemented through a federation of DOI Registration Agencies coordinated by the International DOI Foundation.

would allow to reframe the base question from “Since they are broken, what do we use in place of URIs?” to “How do we make http URIs work as well and as persistently as possible?” (Powell, 2010).

From an information architecture point of view, there is one important distinction to be made in the way this web of data differs from the web of hypertext (Hinton, 2009): in this latter web, links are relationship anchors in hypertext documents written in XHTML; in the former, they link between arbitrary things described by semantic languages like RDF, and URIs identify any kind of object or concept. For both webs though, the same scientific expectations apply: growth, findability, and accountability.

Developing TEI Further

The multi-disciplinary research group coordinated by Monica Palmirani and comprising CIRSIFID at the University of Bologna, the University of Naples, and the University of Rome, is currently working at an expansion of the capabilities of TEI P5 ENRICH to make it capable of handling historical - juridical manuscripts. The group has defined an abstract model which includes all the conceptual categories for metadata representing a manuscript (Palmirani, Contissa & Rubino, 2009), building what they call a meta-model. A first version of this ontology template drafted by legal informatics was validated by historians and domain experts and further compared to the results of a literature run on current XML schemes for manuscripts. On the basis of these preliminary steps, Palmirani (2009) maintains that although TEI P5 ENRICH can be considered the best attempt so far at providing the researcher with a viable tool, it still falls short on a number of important indicators such as the capability to follow the flow and discursive nuances of historical description, and the ability to correctly model the complex relationships that often weave layer on layer in century-old documents. Palmirani and her group are currently working then on extending TEI P5 ENRICH in accordance with the results from their initial qualitative assessment of medieval manuscripts structural and conversational requirements to provide the widely-used standard with the historical-juridical modeling formalisms it is

currently lacking⁵⁶.

Integrating Approaches

How the PURL approach could benefit from visions like that of Linked Data as outlined by Berners-Lee (). Linked Data is a method for the exposure, sharing, and connection of data via dereferenceable URIs on the Web.. data.gov.uk was launched at the end of January 2010.

Unlike a lot of government IT projects, the web site and its underlying infrastructure was developed rapidly and at a low cost by a small team based on open source software

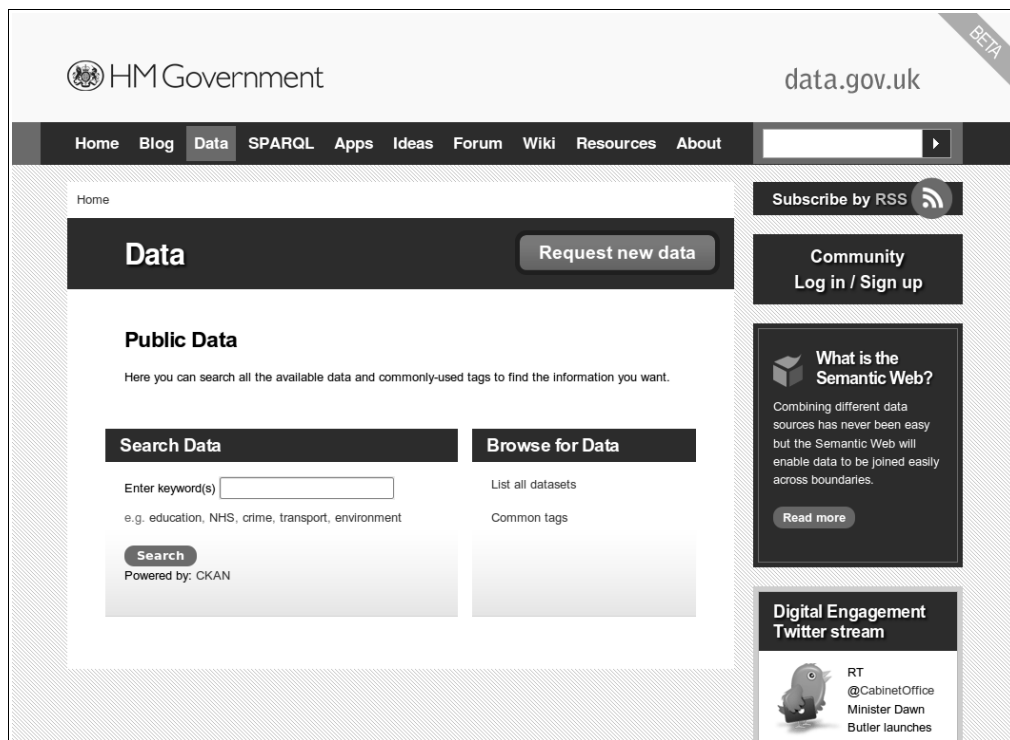


Figure 6: The data search page at data.co.uk, UK's open data for reuse web site.

56 An official partnership has been established between the two groups for this specific purpose, with CIRSFID becoming an associate member of ENRICH (Palmirani, 2009): “Bisogna infine segnalare l'apertura di un collegamento formale con il progetto ENRICH (European Networking Resources and Information concerning Cultural Heritage), coordinato dalla Biblioteca Nazionale di Praga e che 'ha come obiettivo primario quello di fornire un accesso diretto ai beni documentari antichi disponibili in formato digitale'. Nell'ambito di questo progetto il CIRSFID, delegato dal consorzio, (è) divenuto associate member del progetto firmando un agreement formale”.

Tennison (2010). And it uses the Linked Data approach. The reasons behind this choice are eminently technical and programmatical: the UK government is committed to publishing data as linked data because “they are convinced it is the best approach available for publishing data in a hugely diverse and distributed environment, in a gradual and sustainable way” (Tennison, 2010). Linked Data uses RESTful APIs (DEFINE), and this allows to skip any advance planning on what data to publish, how much data to publish, and how to develop the system, and to exploit the natural infrastructure of the Internet by:

- ◆ using HTTP URIs to identify resources: naming *things* with URIs rather than actions on those things (which are carried out using the standard set of HTTP verbs);
- ◆ recognizing the distinction between resources and representations of those resources: the same URI might return a different representation of the resource (XHTML, XML, JSON);
- ◆ returning self-descriptive messages: being able to process representations in a manner that is obvious from the MIME type;
- ◆ using hypermedia as the engine of application state: being able to locate additional resources through the use of (typed) links (Tennison 2010).

Linked data is about following these rules for publishing data. It is about using URIs to identify things, providing information at the end of those URIs that is self-descriptive, and linking those things to other things through typed links. Again, as with emergent systems and with the FRBR standard, correlation seems to be a key word.

Chapter V

Conclusions

“You see, I don’t believe that libraries should be drab places where people sit in silence, and that’s been the main reason for our policy of employing wild animals as librarians.”

– Monty Python’s Flying Circus, The Gorilla Librarian

In Chapter I, I introduced information architecture (IA). I traced the history of human-information interaction back to its roots in design, human-computer interface science, and information retrieval. I briefly outlined how in the late 1990s the traditional hard boundaries between disciplines started to be eased out by the global phenomenon we call World Wide Web, and how in the 2000s mobile or ubiquitous computing gave it finally a tilting push. In the wake of these changes, a number of new fields of practice emerged to produce a design of the digital environment more in sync with the needs of the users and with the absence of boundaries: information architecture among these. I tried to illustrate how actually information architecture, although the very concepts of it go way back to the Xerox Labs in Palo Alto in the 1970s, has been framed as it is today by the work of a few people, including Richard Wurman, Lou Rosenfeld, and Peter Morville.

I documented how four different approaches can be clearly identified in the field, and I labeled them after White as information design, information science, information as a resource, and architecture of information, and I connected IA to the other digital design fields and user experience in the light of the Big IA – Little IA debate. Finally, I considered the long debate on how to define what information architecture is: I registered the views of the community of practice, and I added the results of academic research on the subject matter into the discussion. Finally, I maintained that the definition of information architecture provided in three parts by the Information Architecture Institute is the one that best reflects the recent changes and the mindset and works of the community

at large. I then moved on to consider how information architecture relates in detail to the related practices of interaction design, usability engineering, and user experience design, and provided an explanation as to how this impacts a discourse on digital libraries.

Finally, I explained how the focus of this research was directed to the “information design” and “architecture of information” issues of navigation, categorization, and general findability, as these have the most impact on the structure of those dynamic conversation which this research wants to offer a way to model, and how the idea of context, or place-making, is of the utmost importance for users to successfully understand and use a digital library or digital archive. To better assess how digital libraries perform in light of these requirements, a basic heuristic methodology based on pragmatical indicators such as navigation, grouping, labeling, and place-making was introduced to evaluate the IA of a given system.

In Chapter II, I detailed what a digital library (DL) is. As much as information architecture itself, digital libraries have been around for a long time. If on the one hand it was not until the mid 1990s, when the Digital Library Initiative was founded in the US, that digital libraries came to be the subject of recognized active research, on the other hand Licklider envisioned a computer-based library he called the “library of the future” in 1965. Even so, the framing of what a digital library is has proved to be an elusive subject. I accounted for Harter’s three-profiles view, narrow-broader-broadest, and for his early call for taking the design of digital libraries out of the hands of librarians.

I also reported some of the more mainstream definitions coming from scholars and institutions such as the National Science Foundations, or the EU-funded DELOS project. I detailed how all of these definitions present or frame digital libraries from a number of different perspectives, normally rooted in library and information sciences but reaching out to information systems and business and organization modeling, and how some of them reference “information architecture” even though the concept behind the wording seems to picture a rather different approach from the one currently endorsed by the IA community. I eventually provided a working definition of digital library within the scope

of this research, where the digital library or digital archive is simply the means to the end of improving the overall information architecture and scientific fruition of complex information in the historical and juridical domains on-line: a digital library as a World Wide Web-enabled abstract artifact which renders organized, correlated digital content pertaining to one or more physical or logical collections accessible and findable by users. I briefly outlined the difference between digital libraries and digital archives, and made a point for using them as interchangeable terms in the course of this research.

I tailored a compound difference table from Harter's table of properties to highlight those elements of a DL that better respond to an in-research view of the requirements for dealing with historical and juridical collections, and I then analyzed if and how the information architecture community of practice and research had took part to the ongoing conversation on the design of DLs, and found cursory evidence of very little interest and involvement in both some of the most respected international conferences, and in journals. I documented how this is true the other way around as well. On the other hand, though, a certain long-standing relationship with usability engineering and to some extent with user experience is well documented in digital libraries literature. I explored the ISO standards supporting usability, and argued how they provide metrics for the optimization of human performance and not for the optimization of user satisfaction, and how this still remains a largely unexplored field.

I concluded by reporting four case studies which were analyzed during development of the Innerio Archives R2 using the heuristic evaluation methodology outlined in Chapter I. These included the Early Manuscripts at Oxford DL, the Laurentius DL, the Catalogue of Illuminated Manuscripts, and the Manuscriptorium. For all of them I provided a textual overview of the relevant issues based on the heuristics, their shortcomings, and points of excellence, and a synthetic overview recapping both these indicators and touch-points from the compound table derived from Halter's narrow-broader-broadest set of properties.

In Chapter III, I documented the development and deployment of the Innerio Archives release 2 (R2) for CIRSFID. Innerio is a web application / digital archive that serves 288

medieval manuscript codices of utmost historical and juridical interest from the collection of the Royal College of Spain in Bologna, Italy, amounting to more than 135000 digitized images of single pages in high resolution and full color.

I briefly introduced R1 to better highlight the foundations of the Archives, both from a technical point of view and from a design point of view, and then proceeded to illustrate in detail the refactoring and redesign that was carried out with R2. I detailed how the different IA heuristics were considered, and how issues of navigation, labeling, grouping, and place-making were addressed or partially addressed. I documented the new features and concepts that were introduced in R2, and I concluded by arguing that if R1 was perfectly fitting the concept and idea of a traditional DL / DA, R2 moved the vision a little further on, accounting for a more dynamic and complex vision of the manuscripts themselves and for an improved information architecture and user experience, and the Annotation Engine for glosses factually opened up the vision and challenges brought along by a possible future R3 and its accounting of a living, authoritative juridical and historical document.

Finally, in Chapter IV, I outlined some of the current challenges digital libraries and digital archives face in the on-line modeling of complex, binding scientific information connected to historical and juridical manuscripts such as those from the Irnerio Archives.

I argued that the traditional library science-based model of archival for on-line consumption which is a large part of the search experience in large catalogs is also the *de facto* model for the traditional digital library or digital archive, and that this prevents digital libraries to preserve, model, and render the dynamic authoritative, or legally valid, chain of causalities of artifacts in the historical and juridical domains.

Examples from a current FIRB were presented to illustrate how these documents easily break out of the narrow boundaries of static bibliographical records: the Authenticum and the Montecassino 266 manuscripts both show a complex structure, where different documents are recalled, over a span of time, space, and hyper-space, to build a comprehensive scientific view over the subject. The Montecassino 266 manuscript extends

this view to comments and metadata, effectively reconstructing them as independent, related works which need to be accounted for.

On these grounds, a number of priority tasks for dealing with such documents were identified: the necessity to properly understand the specific requirements for the modeling of information systems capable of handling digital collections for historical and juridical documents; the preservation of the legal axiology of the documents; the modeling of juridical knowledge on said documents; the design of new capabilities for data and metadata correlation; the design of new interfaces for improved user access to these documents.

From these, four basic goals, or capabilities to be implemented, were preliminarily outlined for information architecture modeling of historical and juridical manuscript collections: dynamism, for moving past the static bibliographical record to preserve the historical and juridical conversation commenters, historians and the flow of time have weaved around the artifact; complexity, to preserve the multi-layered structure of complex and compound documents and artifacts, where the metadata is a work *per se*; correlation, for dealing with the heterogeneity of descriptions, sources, formats, and connect across separated digital collections; and persistence, to ensure findability and reliability of URLs through time.

I then proceeded to discuss these goals or capabilities. *Dynamism* required an introduction to current modeling standards, and I described TEI, TEI / MASTER, and TEI ENRICH P5 briefly illustrating their history and their relationships. I then discussed FRBR as a possible step further in the process, allowing a far more abstract control on how documents are modeled and how their dynamic relationships are represented in digital space. *Complexity* brought two different issues to the table: the need to model complex chains of causality and correlation, and the need to present these complex information in the simplest possible way. I argued how these two are connected, and how they cannot be solved on an individual basis or with a simple, technical, language-driven approach. They require a deep understanding of both the structure of the document as conversation, and of

the way information has to be conveyed to be useful. For *correlation*, I introduced a distinction between same-work correlations, work-to-work correlations, and in-work correlations or reification. I outlined how these are considered by the literature to be difficult to handle at the descriptive and conceptual level, for reasons both connected to intrinsic limitations in the standard modeling languages, and to the absence of a coherent information architecture discourse on this very topic. I explained how FRBR, although as argued by some still primarily rooted in first-order logic, addressed many of the technical constraints with a new terminological clarity and with a flexible relationship model which can be easily applied to digital artifacts.

Finally, I discussed *persistence* and detailed how the World Wide Web has a fleeting nature that does not spare academic institutions, web sites, and documents, and that does not go well with the scientific stance of traceability of the sources and solidity of references required from a legal informatics point of view. Before moving on to discuss a number of technical solutions, including PURLs, URNs and DOIs, I explained the way URIs and URLs work, and I reported how this issue is seen by many scholars and commenters as part of a larger problem that has to do with the fact that identifiers are by a great deal social constructs. I connected this view to the larger user-centered stances brought on by information architecture and user experience design, and suggested how a holistic approach which does not neglect the technical side but works with the users could probably produce the best results in the long run.

I then presented some open issues connected to shortcomings in the FRBR language, to the architecture of the ubiquitous DOIs, and to the way FRBR and URIs have friction points mostly related to the identification of where the boundary between web documents and real-world objects might lie. Finally, I briefly described how Linked Data seems to be one possible way to make some of these different approaches work together to obtain an advancement in the way we handle the problems of dynamism, complexity, correlation, and persistence.

Final Considerations

In the end, I personally see this research and these observations as the light switching on in a house far away when dusk is coming, barely showing a treaded path out of the woods. Or as the pile of roughly hewed stones ready for the building of a tower before the workers arrive. So many answers and possibilities still stand unchecked or unexplored, but some of the questions have been asked. And, after all, a few conclusions on the modeling of historical and juridical manuscript collections can certainly be drawn:

1. As the development of the Innerio Archives through its R1 and R2 releases testimonies, handling, modeling, and communicating historical and juridical manuscripts content through on-line digital libraries presents unique challenges.
2. These challenges are technical, concerned with the formalisms used to model complex manuscripts data which configures conversations and not just bibliographical metadata; user-related, concerned with the overall information architecture and user interaction models that allow unencumbered fruition of those conversations; and historical-juridical, connected to specific issues of flow, authoritativeness, accountability, and authorship, so that those conversations properly configure and present their content in a scientifically accurate way.
3. The answer to these challenges is not technological, or not just technological. While standards and formalisms are an important piece, a successful approach needs to consider the legal informatics requirements, which mostly put constraints on the content and the way it is represented and presented, as well as the information architecture requirements, which mostly emphasize the navigational, place-making, and user-related aspects of the system.
4. The answer to these challenges is also multi-disciplinary: they cannot be solved by legal informatics principles, they cannot be solved by digital library research, they cannot be solved by information architecture approaches if these are used in isolation. A holistic approach is required.

5. The disciplines and fields that should be involved in the problem solving discussion have no mainstream conversation happening between them.

I sincerely hope this research provides enough unpolished and roughly hewed stones to build a common understanding of the problem space upon, and that we finally have this long due conversation started. Indeed, it is time we have it.

“As I’ve said many times: the future is already here; it’s just not very evenly distributed.” – William Gibson, 1999

References

Bibliography

- Abrams, S. L. (2004). The role of format in digital preservation. *The Journal of Information and Knowledge Management Systems*. Vol. 34. No. 2. Pp. 49-55.
- Abrams, S.L., & Seaman, D. (2003). Towards a global digital format registry. 69th IFLA General Conference and Council. [http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf]. Accessed Feb 12 2010.
- Alexander, C. (1979). *Timeless Way of Building*. Oxford University Press.
- Anderson, P. W. (1972). More is different: broken symmetry and the nature of the hierarchical structure of science. *Science*. Vol. 177. No. 4047. Pp. 393-396.
- Barabasi, L. (2002). *Linked: The New Science of Networks*. Basic Books.
- Baroni, M. R. (1998). *Psicologia ambientale*. Il Mulino.
- Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition*. Vol. 11. No. 3. Pp. 211– 227.
- Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. [<http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>]. Accessed 5 November 2009.
- Bates, M. (2002). Toward an integrated model of information seeking and searching. Keynote Address, 4th International Conference on Information Needs, Seeking and Use in Different Contexts. [http://www.gseis.ucla.edu/faculty/bates/articles/info_SeekSearch-i-030329.html]. Accessed 5 November 2009.
- Beall, J. (2006). Some Reservations About FRBR. *Library Hi Tech News*. No. 2. Pp. 15-16.

- Berners-Lee, T. (1998). Cool URIs Don't Change.
[<http://www.w3.org/Provider/Style/URI>]. Accessed Jan 17 2010.
- Berners-Lee, T. (2006). Design Issues: Linked Data. W3C.
[<http://www.w3.org/DesignIssues/LinkedData.html>]. Accessed Jan 16 2010.
- Berners-Lee, T. (1991). Document Naming.
[<http://www.w3.org/DesignIssues/Naming.html>]. Accessed Jan 18 2010.
- Bertram, M. (2009). Goffredo Tranensis Apparatus Decretalium, Riproduzione digitale del Manoscritto Montecassino 266. FIRB.
- Bevan, N. (2009). What is the difference between the purpose of usability and user experience evaluation methods? [<http://www.nigelbevan.com/>]. Accessed Dec 30 2009.
- Boersma, P. (2005). Shoulder IA: t-model extended with business layer.
[<http://www.peterboersma.com/blog/2005/03/shoulder-ia-t-model-extended-with.html>]. Accessed Dec 22 2009.
- Boersma, P. (2004). T-Model: Big IA is now UX.
[<http://www.peterboersma.com/blog/2004/11/t-model-big-ia-is-now-ux.html>]. Accessed Dec 22 2009.
- Bonsiepe, G. (1995). Dall'oggetto all'interfaccia. Feltrinelli.
- Brancheau, J. C. & Wetherbe, J. C. (1986). Information architectures: methods and practice. *Information Processing & Management*. Vol. 22. No 6. Pp. 453-463.
- Bournard, L. (2008). The ENRICH Schema – A Reference Guide.
[http://tei.oucs.ox.ac.uk/ENRICH/Deliverables/referenceManual_it.pdf]. Accessed Dec 12 2009.
- Bussolon, S. (2009). Card Sorting, Category Validity, and Contextual Navigation. *Journal of Information Architecture*. Vol. 1. No. 2. [<http://journalofia.org/0102-bussolon/0102-bussolon.pdf>]. Accessed Feb 2 2010.

- Candela, L, Castelli, D., & Pagano, P. (2007). A Reference Architecture for Digital Library Systems: Principles and Applications. In Thanos, C., Borri, F., & Candela, L. (Eds). Digital Libraries: Research and Development. Springer.
- Carroll, J. M. (2009). Human Computer Interaction (HCI). [http://www.interaction-design.org/encyclopedia/human_computer_interaction_hci.html]. Accessed Feb 2 2010.
- Carter, H. (1999). Information architecture. Work Study. Volume 48. No. 5. Pp. 182-185.
- Centre for Technology and the Arts. (1999). MASTER: A Gentle Introduction. [<http://xml.coverpages.org/masterGentintr.html>]. Accessed Feb 2 2010.
- Chowdhury, G.G. (2004). Access and usability issues of scholarly electronic publications. In Gorman, G.E., & Rowland, F. (Eds). Scholarly Publishing in an Electronic Era. International Yearbook of Library and Information Management 2004/2005. Facet Publishing. Pp. 77-98.
- Chowdhury, S., Landoni, M., & Gibb, F. (2006). Usability and impact of digital libraries: a review. Online Information Review. Vol. 30. No. 6. Pp. 656-686.
- Clerwall, C. (2003). Information Architecture – a Descriptive Overview. [http://www.humanit.org/pdf/HumanIT_2003_Ch6_Clerwall.pdf]. Accessed Jan 12 2010.
- Cohen, L. (2004). Issues in URL Management for Digital Collections. Information Technology and Libraries. Vol. 23. No. 2. Pp. 42-49.
- Connolly, D. (2005). Untangle URIs, URLs, and URNs - Naming and the problem of persistence. [<http://www.ibm.com/developerworks/xml/library/x-urlni.html>]. Accessed Dec 30 2009.
- Cooper, A. (1999). The Inmates Are Running the Asylum (1st ed.). SAMS.
- Corning, P. A. (2002). The Re-Emergence of Emergence: A Venerable Concept in Search of a Theory. Complexity. Vol. 7. No. 6. Pp. 18-30.

de Oliveira Lima, J. A., Palmirani, M., & Vitali, F. (2008). Moving in the Time: An Ontology for Identifying Legal Resources. *Computable Models of the Law, Languages, Dialogues, Games, Ontologies*. Pp. 71-85.

Dellavalle, R.P., Hester, E.J., Heiling, L.F., Darke, A.L., Kuntzman, J.W., Graber, M., & Schilling, L.M. (2003). Going, going, gone: lost internet references. *Science*. Vol. 302. No. 5646. Pp.787-8. [<http://dx.doi.org.bibl.proxy.hj.se/10.1126/science.1088234>]. Accessed Dec 30 2009.

Dillon, A. (1994). *Designing Usable Electronic Text: Ergonomic Aspects of Human Information Usage*. Taylor and Francis.

Dillon, A. (2000). Understanding and Evaluating the User Experience with Information Spaces. *Proceedings of the ACIA Information Architecture 2000 Conference*. [http://argus-acia.com/acia_event/dillon_session.html]. Accessed Feb 19 2010.

Ding, W., & Lin, X. (2009). *Information Architecture – The Design and Integration of Information Spaces*. Morgan & Claypool Publishers.

Diringer, D. (1982). *The Book Before Printing: Ancient, Medieval and Oriental*, Courier Dover Publications.

Drabenstott, K. (1993). *Analytical Review of the Library of the Future*. Council on Library Resources.

Engelbart, D. (1990). Knowledge-domain interoperability and an open hyperdocument system. [<http://www.doungengelbart.org/pubs/knowledge-domain-interoperability-and-ohs.html>]. Accessed Jan 11 2010.

ENRICH. (2007). *European Networking Resources and Information concerning Cultural Heritage*. [<http://enrich.manuscriptorium.com/>]. Accessed Feb 1 2010.

Evernden R., & Evernden, E. (2003). *Information first: integrating knowledge and information architecture for business advantage*. Elsevier Butterworth-Einemann.

- Fenn, T., Hobbs, J., & Resmini, A. (2010). *Maturing a practice: Challenges and opportunities for a functioning practice in the absence of a discipline*. Working paper. Presented at the HCI Colloquium 2009. University of Johannesburg.
- Garrett, J. J. (2002). *IA/Recon*. [<http://www.jjg.net/ia/recon/>]. Accessed July 30 2009.
- Garrett, J. J. (2002b). *The Elements of User Experience*. Peachpit Press.
- Geisler, G. (2004). *The Open Video Project: Redesigning a Digital Video Digital Library*. Proceedings of the ASIS&T IA Summit 2004. [www.open-video.org/papers/geisler_ia_summit_04.pdf]. Accessed Feb 8 2010.
- Gibson, W. (1999). *The Science in Science Fiction*. Talk of the Nation. NPR. [<http://discover.npr.org/features/feature.jhtml?wfId=1067220>]. Accessed Feb 28 2010.
- Gilchrist, A., & Mahon, B. (2004). *Information Architecture*. Facet Publishing.
- Goldstein, J. (1999). *Emergence as a Construct: History and Issues*. *Emergence: Complexity and Organization*. Vol. 1. No. 1. Pp. 49–72.
- Gomes, D. & Silva, M. J. (2006). *Modeling Information Persistence on the Web*. Proceedings of the 6th International Conference on Web Engineering (ICWE06). [<http://xldb.di.fc.ul.pt/daniel/docs/papers/gomes06urlPersistence.pdf>]. Accessed Jan 5 2010.
- Gover, D. (2006). *EuroIA Summit Wrap-up*. [http://www.bboxesandarrows.com/view/euro_ia_summit_wrap_up]. Accessed Nov 11 2009.
- Greenfield, A. (2006). *Everyware*. New Riders Publishing.
- Greenstein, D. I., & Thorin, S. E. (2002). *The Digital Library: A Biography*. Digital Library Federation. [<http://www.clir.org/PUBS/reports/pub109/pub109.pdf>]. Accessed Dec 1 2009.

Gregotti, V. (1986). *Il disegno del prodotto industriale in Italia (1860-1980)*. Mondadori Electa.

Grossman, J. (2006). *Designing for Bridge-experiences*. UXMatters. [<http://www.uxmatters.com/mt/archives/2006/06/designing-for-bridge-experiences.php>]. Accessed May 19 2009.

Guenther, K. (2000). *The Evolving Digital Library*. *Computers in Libraries*. Vol. 20. No. 2. Pp. 48-50.

Hakala, J. (2006). *The Sevel Levels of Identification*. *Electronic library and information systems*. Vol. 40. No. 4. Pp. 361-371. DOI 10.1108/00330330610707935.

Harter, S. P. (). *What is a Digital Library? Definitions, Content, and Issues*. *Proceedings of KOLISS DL 96. International Conference on Digital Libraries and Information Services for the 21st Century*. [<http://php.indiana.edu/~harter/korea-paper.htm>]. Accessed Feb 16 2010.

Haverty, M. (2002). *Information Architecture Without Internal Theory: An Inductive Design Process*. *Journal of the American Society for Information Science and Technology*, Vol. 53. No.10. Pp. 839-845. [<http://dx.doi.org/10.1002/asi.10096>]. Accessed Jun 21 2009.

Hearst, M., Yee P., Swearingen, K., & Li, K. (2003). *Faceted Metadata for Image Search and Browsing*. *Proceedings of ACM CHI 2003*.

Hedstrom, M., & Lee, C.A. (2002). *Significant properties of digital objects: definitions, applications, implications*. Pp. 218-27, *Proceedings of the DLM-Forum 200*. [http://europa.eu.int/historical_archives/dlm_forum/doc/dlm-proceed2002.pdf]. Accessed Feb 2 2010.

Hill, S. (2000). *An Interview with Louis Rosenfeld and Peter Morville*. [http://www.oreillynet.com/pub/a/oreilly/web/news/infoarch_0100.html]. Accessed Oct 16 2009.

Hinton, A. (2009). The Machineries of Context. *Journal of Information Architecture*. Vol. 1. No. 1. [<http://journalofia.org/volume1/issue1/04-hinton/jofia-0101-04-hinton.pdf>]. Accessed Jan 18 2010.

IFLA. (2007). *Functional Requirements for Bibliographic Records: Final Report*. [http://archive.ifla.org/VII/s13/frbr/frbr_current5.htm]. Accessed Jan 31 2010.

IFLA Study Group on the Functional Requirements for Bibliographic Records. (2008). *Functional Requirements for Bibliographic Records. Final Report*. IFLA.

Information Architecture Institute. (2005). *Our Mission*. [http://www.iainstitute.org/en/about/our_mission.php]. Accessed Nov 12 2009.

Inalhan, G., & Finch, E. (2004). Place attachment and sense of belonging. *Facilities*. Vol. 122. No. 5/6. Pp. 120-128.

ISO. (1998). ISO 9241-11. *Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on Usability*.

ISO. (2009). ISO FDIS 9241-210. *Human-centred design process for interactive systems*.

Jeng, J.H. (2005). Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability. *Libri: International Journal of Libraries and Information Services*. Vol. 55. No. 2/3. Pp.96-121.

Jenkins, H. (2006). *Convergence Culture*. New York University Press.

Johannesen, A. (2009). Re: FRBR WEMI and identifiers. Message to the NGC4LIB mailing list. [<http://serials.infomotions.com/ngc4lib/archive/2009/200911/1913.html>]. Accessed Feb 2 2010.

Johnson, N. F. (2007). *Two's Company, Three is Complexity: A simple guide to the science of all sciences*. Oneworld.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. University of Chicago.

- Johnson, S. (2001). *Emergence: the connected lives of ants, brains, cities, and software*. Scribner.
- Johnston, P. (2009). *httpRange14, Cool URIs & FRBR*.
[<http://efoundations.typepad.com/efoundations/2009/02/httprange14-cool-uris-frbr.html>]. Accessed Feb 1 2010.
- Jonsson, E. (2002). *Inner Navigation*. Scribner.
- Kahn, P., & Krzysztof, L. (1998). *Website Information Architecture*. New Riders.
- Kalbach, J. (2007). *Designing Web Navigation*. O'Reilly.
- Katic, T. (2001). Application of the FRBR to rare book cataloguing. In *ELAG 2001: Integrating Heterogenous Resources, 25th Library Systems Seminar, Prague*.
[<http://old.stk.cz/elag2001/Papers/TinkaKatic/TinkaKatic.html>]. Accessed Jan 17 2010.
- Katic, T., & Klarin, S. (2003). Application of the FRBR to older books and electronic resources cataloguing. In *Libraries in the Digital Age, Dubrovnik*.
- Kruger, J. (2008). *Complexity: Why Simple Things Become Complex (and How Complex Things Can Be Made Simple)*. Hyperion.
- Lakoff, G. (1990). *Women, Fire, and Dangerous Things*. University of Chicago Press.
- Law, E. L., Roto, V., Hassenzhal, M., Vermereen, A. P. O. S., & Kort, J. (2009). Understanding, scoping and defining user experience: a survey approach. *Proceedings of the 27th international conference on Human factors in computing systems*. Pp. 719-728.
- Le Beuf, P. (2005). *Functional Requirements For Bibliographic Records (FRBR): Hype, Or Cure-all?* Routledge.
- Levy, D. M. (2000). Digital Libraries and the Problem of Purpose. *D-Lib Magazine*. Vol. 6. No. 1. [<http://www.dlib.org/dlib/january00/01levy.html>]. Accessed Nov 11 2009.
- Licklider, J. C. R. (1965). *Libraries of the Future*. MIT Press.

- Madsen, D. (2009). Shall We Dance?. *Journal of Information Architecture*, Vol. 1, No. 1. [<http://journalofia.org/volume1/issue1/01-madsen/jofia-0101-01-madsen.pdf>]. Accessed July 22 2009.
- Marbeck, D. (2008). URL Design for Information Architects. Slideshare. [<http://www.slideshare.net/deanna.marbeck/url-design-for-information-architects-presentation>]. Accessed Jan 6 2010.
- Marchionini, G. (2004). From Information Retrieval to Information Interaction. 26th Annual European Conference on Information Retrieval. [<http://ils.unc.edu/~march/ECIR.pdf>]. Accessed Oct 18 2009.
- McCullough, M. (2005). *Digital Ground: Architecture, Pervasive Computing, and Environmental Knowing*. The MIT Press.
- Miksa, F. L., & Doty, P. (1994). Intellectual Realities and the Digital Library. Proceedings of the First Annual Conference on the Theory and Practice of Digital Libraries. [<http://www.csd.tamu.edu/DL94/paper/miksa.html>]. Accessed Feb 17 2010.
- Mitchell, M. (2009). *Complexity: A Guided Tour*. Oxford University Press.
- Morrogh, E (2003). *Information Architecture: an Emerging 21st Century Discipline*. Prentice Hall.
- Morville, P. (2005). *Ambient Findability*. O'Reilly.
- Morville, P. (2000). Big Architect, Little Architect. [http://argus-acia.com/strange_connections/strange004.html]. Accessed Jan 3 2010.
- Morville, P. (2004). *User Experience Design*. [<http://semanticstudios.com/publications/semantics/000029.php>]. Accessed Oct 16 2009.
- Morville, P., & Rosenfeld, L. (2006). *Information Architecture for the World Wide Web* (3rd ed.). O'Reilly.

Munro, A. J., Höök, K., & Benyon, D. (1999). *Social Navigation of Information Space*. Springer.

Nielsen, J. (2007). *Intranet Information Architecture*. Alertbox. [<http://www.useit.com/alertbox/ia.html>]. Accessed Jan 16 2010.

Nielsen, J. (1997). *Measuring the Usability of Reading on the Web*. Alertbox. [<http://www.useit.com/alertbox/readingmetrics.html>]. Accessed Jan 12 2010.

Nielsen, J. (2007). *Top Ten Mistakes in Web Design*. [<http://www.useit.com/alertbox/9605.html>]. Accessed Dec 8 2009.

Nielsen, J. (1993). *Usability Engineering*. AP Professional.

Nielsen, J., & Levy, J. (1994). *Measuring usability preference vs performance*. *Communications of the ACM*. Vol. 37. No. 4. Pp. 66-76.

Nielsen, J., & Loranger, H. (2006). *Prioritizing web usability*. New Riders Press.

Norman, D. (1999). *Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex and Information Appliances Are the Solution*. MIT Press.

OCLC. (2004) PURL Web site. [<http://purl.org>]. Accessed Feb 1 2010.

Pake, G. E. (1985). *Research at Xerox PARC: a founder's assessment*. *IEEE Spectrum*. Cited in Hearst, M. A. (1996)

Palmirani, M. (2007). *Diritto dell'informatica e beni culturali: il progetto IRNERIO*. *Proceedings of the 2nd DELOS Conference on Digital Libraries*. Pisa.

Palmirani, M. (2009). *Legal Resources Modelling in the Semantic Web for implementing eGov*. Presentation for CodeX: Stanford Center for Computers and Law, Stanford Law School, Stanford University.

Palmirani, M. (2009). *Un modello semantico eterogeneo e armonizzato per la classificazione dei manoscritti*. *Proceedings of the Erice FIRB Meeting*. Erice.

Palmirani, M., Contissa, G., & Rubino, R. (2009). Fill the Gap in the Legal Knowledge Modelling. *RuleML 2009*. Pp. 305-314.

Petrucciani, A. (2007). The FRBR model application to Italian cataloguing practices: general problems and its normative use. ICCU Report for the RICA Commission. [<http://www.iccu.sbn.it/FRBReng.doc>]. Accessed Jan 21 2010.

Pieri, B. (2002). About the Innerio Project. [<http://irnerio.cirsfid.unibo.it/about/>]. Accessed Dec 11 2009.

Pilgrim, M. (2002). Dive Into Accessibility: Not opening new windows. [http://diveintoaccessibility.org/day_16_not_opening_new_windows.html]. Accessed Feb 1 2010.

Pisanski, J., Zumer, M., & Ålberg, T. (2009). Frbrisation : towards a bright new future for national bibliographies. *Proceedings of the World Library and Information Congress: 75th IFLA General Conference and Council*. Milan.

Powell, A. (2010). More Famous than Simon Cowell. [<http://efoundations.typepad.com/efoundations/2010/02/more-famous-than-simon-cowell.html>]. Accessed Feb 4 2010.

Quintarelli, E. (2005). Folksonomies: Power to the People. ISKO Meeting. [<http://www.iskoi.org/doc/folksonomies.htm>]. Accessed Jan 22 2010.

Quintarelli, E., Resmini, A., & Rosati, L. (2008). The FaceTag Engine: a Semantic Collaborative Tagging Tool. In *Browsing architecture: Metadata and beyond*. Zambelli, M. et al. (Eds.). Fraunhofer IRB Verlag, Pp. 204-217.

Ranganathan, S. R. (2006). *The five laws of library science*. Ess Ess Publications.

Reiss, E. (2003). Web Dogma. [<http://www.fatdux.com/how/our-web-dogma/>]. Accessed Feb 25 2010.

- Resmini, A. (2009). Pervasive IA. Closing Plenary at the German IA Konferenz 2009. [<http://prezi.com/73087/view>]. Accessed Sept 11 2009.
- Resmini, A., Madsen D., & Byström, K. (2009). IA Growing Roots. Bulletin of the American Society for Information Science and Technology, Vol. 35. No. 3. Pp 31-33. [http://www.asis.org/Bulletin/Feb-09/FebMar09_Resmini_Bystrom_Madsen.html] (Accessed June 2 2009).
- Resmini, A., & Rosati, L. (2009a). Information Architecture for Ubiquitous Ecologies. Proceedings of the International Conference on Management of Emergent Digital EcoSystems. Lyon. [<http://doi.acm.org/10.1145/1643823.1643859>] (Accessed December 12 2009).
- Resmini, A., & Rosati, L. (2008). Semantic Retail: Towards a Cross-Context Information Architecture. Knowledge Organization. Vol. 35 No. 1. Pp. 5-15
- Resmini, A., & Rosati, L. (2009b). The Semantic Environment: Heuristics for a Cross-Context Human-Information Interaction Model. In *The Engineering of Mixed Reality Systems*, Dubois et al. (Eds.), Springer.
- Roggio, A. (2009). Ecommerce Know-How: Information Architecture to Improve SEO and Usability. [<http://www.practicalecommerce.com/articles/921-Ecommerce-Know-How-Information-Architecture-to-Improve-SEO-and-Usability>]. Accessed Jan 24 2010.
- Ronda León, R. (2008). Arquitectura de Información: análisis histórico-conceptual. *No Solo Usabilidad*. No. 7.
- Rosati, L. (2007). *Architettura dell'informazione: Dagli oggetti quotidiani al Web*. Apogeo.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*. No. 7. Pp. 532–547.
- Rosen, R. (1985). *Anticipatory Systems*. Pergamon Press.
- Rosen, R. (1998). *Essays on Life Itself*. Columbia University Press.

Rosen, R. (1978). *Fundamentals of Measurement*. Elsevier Science.

Rosenfeld, L. (2009). Stop listening to people like me.

[http://www.louisrosenfeld.com/home/bloug_archive/2009/03/stop_listening_to_people_like.html]. Accessed Jan 2 2010.

Sauermann, L. & Cyganiak, R. (2007) Cool URIs for the Semantic Web.

[<http://www.w3.org/TR/2007/WD-cooluris-20071217/>]. Accessed Jan 22 2010.

Shirky, C. (2005). *Ontology is Overrated: Categories, Links, and Tags*.

[http://www.shirky.com/writings/ontology_overrated.html]. Accessed Nov 12 2009.

Singhal, A. (2001). *Modern Information Retrieval: A Brief Overview*. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*. Vol. 24. No. 4. Pp. 35–43.

[<http://singhal.info/ieec2001.pdf>]. Accessed Nov 3 2009.

Sletten, B. (2007). What's in a URI? [<http://www.devx.com/semantic/Article/34994>].

Accessed Jan 6 2010.

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Harvard University Press, Cambridge, Massachusetts.

Sterling, B. (2005). *Shaping Things*. The MIT Press.

Streitz, N. A. (2005). *From Human-Computer Interaction to Human-Artifact Interaction: Interaction Design for Smart Environments*. In *From Integrated Publication and Information Systems to Information and Knowledge Environments: Essays Dedicated to Erich J. Neuhold*. Hemmje M., Niederee C., & Risse T. Eds. Springer.

Tennison, J. (2010). *Why Linked Data for data.gov.uk?*

[<http://www.jenitennison.com/blog/node/140>]. Accessed Feb 1 2010.

Thurow, S. (2006). *SEO and URL Structure*. [<http://www.clickz.com/3622997>]. Accessed Jan 16 2010.

- Tognazzini, B. (2003). It's time we got respect. [<http://www.asktog.com/columns/057ItsTimeWeGotRespect.html>]. Accessed Oct 03 2009.
- Toms, E., Dufour, C., & Hesemeier, S.. (2004). Measuring the user's experience with digital libraries. International Conference on Digital Libraries. Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries. Pp. 51-52.
- Troll Covey, D. (2002). Usage and Usability Assessment: Library Practices and Concerns. Digital Library Federation - Council on Library and Information Resources.
- Tufte, E. (1990). Envisioning Information. Graphics Press.
- Vanderwal, T. (2001). Comment. In Merholz, P. Thoughts on the definition and community of information architecture. [<http://www.peterme.com/archives/00000091.html>]. Accessed Feb 5 2010.
- Van Dijck, P. (2007). Global Information Architecture. Proceedings of the 2nd Italian IA Summit.
- Van Dijck, P. (2003). Information Architecture for Designers: Structuring Websites for Business Success. Rotovision.
- Van Dijck, P. (2008). Social networks going global. Proceedings of the 4th ASIS&T European Information Architecture Summit. [<http://www.euroia.org/2008/>]. Accessed Oct 16 2009.
- Van House, N.A., Butler, M.H., Ogle, V., & Schiff, L. (1996). User-centred iterative design for digital libraries. D-Lib Magazine. [www.dlib.org/dlib/february96/02vanhouse.htm]. Accessed Jan 12 2010.
- Vano, C. (2009). Elementi per una catena di più livelli d'informazione relative a un ms medievale e a un ms ottocentesco collegato. University of Napoli. FIRB.
- Veronin, M. A. (2003). URL 404 File Not Found - Dealing with the transient nature of the Web. Journal of Audiovisual Media in Medicine. Vol. 26. No. 4. Pp. 153-155.

- Zang, Y., & Salaba, A. (2007). Major Issues Facing FRBR Research and Practice Identified in a Delphi Study. [http://frbr.slis.kent.edu/publications/delphi_issues.pdf]. Accessed Jan 22 2010.
- Zeldman, J. (2003). *Designing with Web Standards*. New Riders.
- Weinberger, D. (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder*. Times Books.
- Weinberger, D. (2002). *Small Pieces Loosely Joined*. Perseus Publishing.
- Weng C., Mi J. (2006). Towards accessibility to digital cultural materials: an FRBRized approach. *OCLC Systems and Services*. Vol. 22. No. 3. Pp. 217-232.
- Weitzmann, F. (1995). *The architecture of information: interpretation and presentation of information in dynamic environments*. MIT Press.
- White, M. (2004). Information architecture. *The Electronic Library*. Vol. 22. No. 3. Pp. 218-219.
- Wodtke, C. (2002). *Information Architecture – Blueprints for the Web*. New Riders Press.
- Wurman, R. S. (2001). *Information Anxiety 2*. Que.
- Wurman, R. S. (1996). *Information Architects*. Graphis Inc.
- Wyllys, R. E. (2000). *Information Architecture*. [<http://www.ischool.utexas.edu/~l38613dw/readings/InfoArchitecture.html>]. Accessed Oct 22 2009.