



TALLINNA ÜLIKOOL



UNIVERSITÀ DEGLI STUDI DI PARMA



høgskolen i oslo



Education and Culture DG

ERASMUS MUNDUS

Andrea Zanni

Collaboratory Digital Libraries for Humanities in the Italian context

Master thesis
International Master in Digital Library Learning
2010

0.1 Abstract

The study investigates the approach to collaboration in Humanities, within the Italian context, to test the possibility of collaborative digital library for scholars. The research hypothesis is that collaboration can foster innovation and scientific development: therefore, within Humanities, digital libraries can be the collaborative laboratory for research. Thus, understanding perception of scholars towards collaboration, especially online, and comprehending if wiki systems could be the framework of collaboration were the objectives of the study.

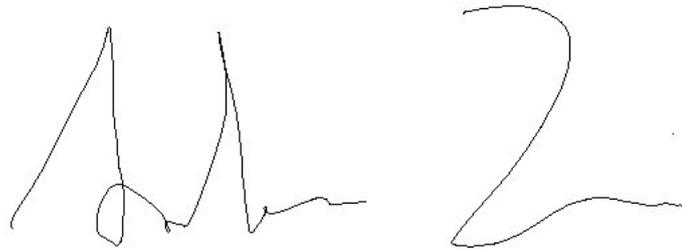
A qualitative approach has been adopted, using case study as research method: five in-depth, semi-structured interviews to Digital Humanities scholars provide data integrated with interviews with two key informants (one of which is prof. Umberto Eco).

The results of the study show that Humanities, within Italian context, do appreciate collaboration and the concept of a collaborative digital library, though several issues need to be solved. In fact, Humanities are still tied with individual work and collaboration is not easy to pursue, for cultural, technical and political reasons. Great effort needs to be done at many different levels to eliminate obstacles and facilitate online collaboration for scholars. The study provides a draft model for a collaborative digital library arisen from gathered data.

Keywords: *collaboration, digital libraries, collaborative, collaborative digital library, wiki, Humanities, Italy*

0.2 Declaration

I certify that all material in this dissertation which is not my own work has been identified and that no material is included for which a degree has previously been conferred upon me.

A handwritten signature in black ink, consisting of a stylized 'A' followed by a 'Z' and a flourish.

Andrea Zanni

0.3 Acknowledgments

I firstly would like to thank Anna Maria Tamaro, who I luckily met by chance two years ago, in a conference in Naples. She was so kind to introduce me to DILL, and help me out before, during and even after my participating to the master course. She believed in me from the very first moment, and I can't really forget that. Furthermore, she actually inspired the present study, and we discussed largely issues and results arise from the research. Her supervision and support were fundamental.

I would also thank interviewees and participants of the study: they contributed greatly and our conversations have been deep, fruitful and inspiring. Professor Umberto Eco was so kind to invite me in his house for two unforgettable hours: I still have in my eyes his unbearable library.

I'm very grateful to my fiancée, Maria Luisa, who waited for me during for a long year, and was there when I returned. I owe her a lot.

I would like to acknowledge my Lecturers at Oslo University College, Parma University and Tallinn University, and all the visiting professors. I want to thank my family, who supported me and always pushed me to follow my aspirations. Needless to say I won't be writing these lines without their unconditioned care.

Finally, I want to thank and hug all my fellow students of DILL-2008 for being so nice friends and comrades in an amazing adventure. Especially Sara, Juan, Mehrnoosh, Monica has been a great support in these long months in the cold. It has been a pleasure to meet you, JAMMS.

0.4 Abbreviation list

- CERN: Organisation Européenne pour la Recherche Nucléaire (originally *Conseil Européen pour la Recherche Nucléaire*)
- CoI: Communities of Interest
- CoP: Communities of Practice
- DL: Digital Libraries
- DH: Digital Humanities
- Fedora: Flexible Extensible Digital Object Repository Architecture
- NSF: US National Science Foundation
- OA: Open Access
- OAI: Open Archives Initiative
- OAI-PMH: Open Archive Initiative Protocol for Metadata Harvesting
- OAI-ORE: Open Archive Initiative for Object Reuse and Exchange
- OWL: Web Ontology Language
- RDF: Resource Description Framework
- RSS: Really Simple Syndication
- STM: Science, Technology and Medicine
- VRE: Virtual Research Environment
- WMF: Wikimedia Foundation
- WMI: Wikimedia Italy
- W3C: World Wide Web Consortium

Contents

| | | |
|----------|---|-----------|
| 0.1 | Abstract | i |
| 0.2 | Declaration | ii |
| 0.3 | Acknowledgments | iii |
| 0.4 | Abbreviation list | iv |
| 1 | Introduction | 1 |
| 1.1 | Collaboration as a new paradigm | 2 |
| 1.1.1 | E-science | 3 |
| 1.1.2 | Wiki, the Web and the <i>wisdom of the crowds</i> | 3 |
| 1.2 | Motivation | 4 |
| 1.2.1 | The CERN experience | 5 |
| 1.3 | Digital Humanities | 7 |
| 1.4 | The Italian context | 9 |
| 1.5 | Humanities, Digital Humanities and Collaborative Digital Li- braries | 10 |
| 1.6 | Research hypothesis, questions and objectives | 11 |
| | References | 13 |
| 2 | Literature review | 15 |
| 2.1 | Introduction | 16 |
| 2.1.1 | Italian literature | 17 |
| 2.2 | Collaboration | 18 |
| 2.2.1 | E-science, cyberinfrastructure, e-research | 19 |
| 2.2.2 | Collaboratories | 22 |
| 2.2.3 | Digital libraries and collaboratories | 24 |
| 2.2.4 | Visions of digital libraries | 25 |

| | | |
|----------|--|-----------|
| 2.3 | Digital humanities | 28 |
| 2.3.1 | Digital classics | 29 |
| 2.4 | Wiki | 32 |
| | References | 37 |
| 3 | Methodology | 41 |
| 3.1 | Research Questions and Objectives | 42 |
| 3.2 | Introduction | 43 |
| 3.3 | Interpretivism | 43 |
| 3.4 | Qualitative research | 44 |
| 3.4.1 | Theoretical and conceptual framework | 45 |
| 3.4.2 | Trustworthiness | 48 |
| 3.4.3 | Limitations | 51 |
| 3.5 | Case study | 53 |
| 3.6 | Target group and sampling | 56 |
| 3.6.1 | Purposive snowball sampling | 57 |
| 3.6.2 | Key informants | 57 |
| 3.6.3 | Interviewees | 58 |
| 3.7 | Data collection techniques | 60 |
| 3.7.1 | Interviews | 61 |
| 3.8 | Computer as a research instrument | 65 |
| 3.8.1 | L ^A T _E X | 66 |
| 3.8.2 | Wiki | 68 |
| 3.9 | Data analysis | 72 |
| 3.9.1 | Open coding | 72 |
| 3.9.2 | Axial coding | 74 |
| 3.9.3 | Selective coding | 74 |
| | References | 75 |
| 4 | Analysis and findings | 77 |
| 4.1 | Introduction | 78 |
| 4.2 | Different types of Digital Libraries | 79 |
| 4.2.1 | Digital library as a <i>repository</i> | 81 |
| 4.2.2 | Digital library as a <i>Virtual Research Environment</i> | 83 |

| | | |
|--------|---|-----|
| 4.2.3 | Focus on VREs | 85 |
| 4.3 | Collaboration | 86 |
| 4.4 | Collaborative editing | 87 |
| 4.4.1 | Consensus and NPOV | 88 |
| 4.4.2 | Tasks and forms of collaboration | 89 |
| 4.4.3 | Lowering costs | 90 |
| 4.4.4 | No “best edition” | 92 |
| 4.4.5 | Neutralization process | 92 |
| 4.5 | Community | 94 |
| 4.5.1 | Control, review and quality | 96 |
| 4.5.2 | Fear | 99 |
| 4.5.3 | Boundaries | 100 |
| 4.6 | Authorship | 101 |
| 4.6.1 | The <i>myth of the lonely scholar</i> | 102 |
| 4.6.2 | Interpretation | 103 |
| 4.6.3 | Attribution | 104 |
| 4.6.4 | Intellectual property | 105 |
| 4.7 | Scientific social networks | 106 |
| 4.8 | Laboratory | 110 |
| 4.9 | Amateur and scholar digital libraries | 112 |
| 4.10 | The Italian gap | 113 |
| 4.10.1 | Political, cultural, institutional issues | 114 |
| 4.10.2 | Italian digital libraries | 116 |
| 4.10.3 | Amateur digital libraries go wiki | 119 |
| 4.10.4 | Wiki is not suitable for scholars | 120 |
| 4.10.5 | Collaborative editing and variations | 121 |
| 4.11 | A Collaboratory Digital Library model | 122 |
| 4.11.1 | Layers | 124 |
| 4.11.2 | From static to dynamic | 124 |
| 4.11.3 | Hypertextuality | 126 |
| 4.11.4 | TEI mark-up | 126 |
| 4.11.5 | Other layers | 127 |
| 4.12 | Open | 128 |
| | References | 129 |

| | | |
|----------|--|------------|
| 5 | Conclusions and recommendations | 133 |
| 5.0.1 | Wiki as a digital library | 135 |
| 5.1 | Different types of digital libraries | 136 |
| 5.2 | Importance of collaboration | 137 |
| 5.3 | Community | 139 |
| 5.3.1 | Quality and reliability | 140 |
| 5.4 | The Italian context | 141 |
| 5.4.1 | Italian digital libraries | 142 |
| 5.5 | Wikis is for amateurs, not scholars | 143 |
| 5.6 | Solving the Authorship problem | 144 |
| 5.7 | Collaboratory Digital Library model | 146 |
| | Appendix 1 | 149 |
| 5.8 | Interview | 150 |
| | Appendix 2 | 167 |
| | Bibliography | 171 |

Chapter 1

Introduction

1.1 Collaboration as a new paradigm

Collaboration has become increasingly important in our world. Collaborative projects are increasing in importance and possibilities, also in fields very different one from each other.

The revolution of the (so-called) “Web 2.0” is making *social* every aspect of the digital world as an ecosystem (i.e. social tagging, social bookmarking, information sharing), but also the physical world is looking at himself as a “globalized village”, in which multiculturalism and multilingualism are increasingly becoming a status *de facto* and collaboration is becoming a necessity to cope with ever-growing complexity of the modern world.

Web 2.0 is a neologism coined by the technological publisher Tim O’Reilly from O’Reilly Inc, in a conference in 2004¹. Although is a controversial expression², it is now widely used for indicating a phenomenon of deeper socialization, openness and empowerment of users. O’Reilly defined it as follows:

a set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet a more mature, distinctive medium characterized by user participation, openness and network effects (Musser, O’Reilly, et al., 2006, p. 1),

Moreover, we have seen other “revolutions” spreading in every sector of human socio-economic life, with the 2.0 brand pasted everywhere: *Business 2.0*, *Enterprise 2.0*, *Office 2.0*, *Library 2.0*, etc. Beside marketing strategies, which are pushing the adopting of new labels to attract customers, it is nevertheless easy to admit that *collaboration*, aside communication and conversation, is a crucial part of this “2.0” revolution, and maybe it is his most hidden but yet most important feature.

¹See O’Reilly (2005)

²Many people have been criticizing the Web 2.0 label, arguing that there is no “new Web”, but just an evolution of previously hidden features of the Web itself. For the sake of simplicity, in the present study Web 2.0 label will be used to indicating a phenomenon of socialization of the Web and user empowerment.

1.1.1 E-science

Science has understood long time ago the power of collaboration, especially when supported by virtual frameworks. *E-science*, *E-research*, *cyberinfrastructure*, *scholarship 2.0*, are some of the labels used to indicate a movement that exploits the digital world for creating a new kind of scientific work. Through the digital environment scientists can work collaboratively succeeding time and space, sharing costs, gathering different perspectives and skills and competences. Although competition is still present in scientific and research world, scientists did reach the conclusions that ambitious goals needed cooperation and collaboration of different individuals, different laboratories, different disciplines, even different nations. International research centers like CERN teach us that deep and huge collaboration is necessary to cope with huge problems, as those that occur in disciplines like *High Energy Physics*, where complex and very expensive facilities are needed. Complex problems require complex solutions, and when complexity arises, different approaches, methods, disciplines, answers are demanded: thus you need *collaboration*.

The *World Wide Web* itself was developed at CERN to be a framework for collaboration among users, namely “a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of a common project” (Wardrip-Fruin & Montfort, 2003).

According to his inventor, Tim Berners-Lee:

The basic idea of the Web is that an information space through which people can communicate, but communicate in a special way: communicate by sharing their knowledge in a pool. The idea was not just that it should be a big browsing medium. The idea was that everybody would be putting their ideas in, as well as taking them out. — (Berners-Lee, 1999)

1.1.2 Wiki, the Web and the *wisdom of the crowds*

An extreme form of collaborative projects are *wikis*. Wikipedia³ is obviously the best and best-known example: a titanic encyclopedia, counting 15

³<http://www.wikipedia.org>

million articles (3 millions just in the English version), billions of readers every month, thousand of active users (Meta, 2010). Still an ever-growing effort, Wikipedia is the icon and the symbol of the *wisdom of the crowds* (Surowiecki, 2004) and of the power of cooperation among users. Wikipedia is just one of the many projects involving users all around the Web: we are facing a phase of the Web where users are *empowered* more and more everyday, and where UGC (*User Generated Content*) has become a normal feature of many projects.

Interestingly, the topic of collaboration has been investigated also in pioneering works about the future of libraries (Licklider, 1965), of scholarship (Bush, 1945), of publishing (Nelson, 1987) in the period of the first computers and computational machines.

Returning at the present day, “Library 2.0” is often the label used to indicate the aim of make libraries more participatory (Lankes, Silverstein, Nicholson, & Marshall, 2007), focused on empowering users and foster conversations. On the economics side, some have explored the theme of wiki in enterprises and firms, or in innovation-oriented workplaces (Cammarata, 2007). Other have discussed the more general change of the whole business world, the advent of *wikinomics* (Tapscott & Williams, 2006).

1.2 Motivation

At the end of 2005, I discovered Wikipedia while I was searching something on Google; probably, a very common experience. It was a very exciting discovery: I was thrilled and amazed by the fact that you could edit and contribute to a collective encyclopedia, and that every user got the chance and the power to add some of his knowledge, to donate it to the world. Wikipedia gathers millions users all around the world, and thousand of them got so interested in the project that they become very active users, donating great amounts of their time to write and control articles, patrol recent contributes, check information, etc.

Through Wikipedia, in few months, I stumble upon another project, called *Wikisource*⁴, a wiki repository of public domain texts and books.

⁴<http://www.wikisource.org>

Wikisource is a digital library of free of copyright texts, created and organized by a community of volunteers, exactly as Wikipedia. I soon became an active and enthusiastic user; few things could be more appealing to me than create a digital library, merging both my competences and love for computers and for books. I have been spending on Wikisource few years, learning and getting introduced to the world of digital libraries, even though from an amateur point of view.

At the time, I was studying mathematics: for my bachelor thesis, I decided with my advisor to digitize an ancient book about history of mathematics and to upload it on Wikisource. I also had the chance to create a subproject about mathematical texts, and I coped for the first time with the issues of digitization of books.

After that experience, I met by chance my future professor and advisor Anna Maria Tammaro, within a conference about digital libraries. Thus, I applied for the DILL master, and two years after this is the result of my path through the countless (and exciting) lessons, seminars, conferences experienced.

This master thesis, in a certain way, goes back to the roots, exploring the topic of collaboration within the Humanities scholar community in the Italian context. An underground motivation that boosted this study is in fact the strong belief in collaborative projects and in open access and sharing of knowledge.

1.2.1 The CERN experience

During the master, a month has been dedicated to internship in a chosen institution. I got the chance to stay at CERN, the *European Organization for Nuclear Research*, which is the most advanced research center in particle physics of all over the world, conducting extremely sophisticated experiments to unveil the fundamental mysteries of physics. Since its creation, CERN has made many important discoveries for which scientists have received prestigious awards, including Nobel prizes: one of these inventions is the *World Wide Web*, developed, at the beginning, to improve and speed-up the information sharing between physicists working in different universities

and institutes all over the world. There, I worked in the library institute, where I have explored the CERN digital library system, *CDS Invenio*, an in-house developed information system software they use for providing access to over 1 million between articles, preprints, books, documents.

What is more, CERN is an amazing location where to understand advantages and disadvantages of collaboration. At CERN, I observed the complexity of a huge and multicultural environment focused on e-science and large collaborative projects, a context where people from different disciplines and from different nations gather to unveil some of the mystery of the modern physics. The CERN context is indeed unique: several thousands of students, scientists and researchers from all over the world, who often stays at CERN for less than a year, and come from different disciplines and work in different areas, working on the most complex e-science instrument humans have ever made, the LHC (*Large Hadron Collider*), a 27 kilometers-long particle accelerator.

The open paradigm of HEP

Particle physics, nowadays often called HEP (*High Energy Physics*), is a in fact field of physics that demands incredibly expensive facilities, as a particle accelerator. This meant that different institutions, sometimes different governments (that's the case of CERN) had to collaborate for sharing costs, that were absolutely non affordable by single universities. Thus, from very beginning HEP had to cope with the advantages and disadvantages of collaboration that happened at all the different levels: collaboration among institutions, among disciplines, among researchers.

HEP, and generally particle physics, is then a field where Open Access does work very well: due to the scarcity of important facilities (i.e. Fermilab, CERN), all articles are released in open access, in repositories as arXiv⁵. Particle physics had the first catalogs of pre-print literature in 1974 (Suber, 2009), and spread largely the open access movement in 1991, when Paul Ginsparg started a repository to include pre-prints in physics, mathematics, computer science. arXiv is the most famous and biggest repository of sci-

⁵<http://arxiv.org>

entific papers, and is an authoritative landmark for many fields of physics, mathematics, computer science.

Thus, this specific field is a perfect environment for collaboration. In decades, this has led to an extremely rich and interactive knowledge environment, where information is shared openly and naturally, without the need of mandates or forcing (Voss, 2007).

The sharing culture of the particle physics community makes it a model for other communities of practice, especially in other scientific disciplines. In the last decades, the community has developed tools as arXiv (a disciplinary repository) and SPIRES⁶ (the search engine) to unify and access the whole HEP literature. What is more, they are also trying to develop a new business model to cope with serial crisis in the field, creating a project for making all the HEP literature go open access (SCOAP3⁷).

1.3 Digital Humanities

Also Humanities have found in the digital world a brand new landscape to explore and exploit: interestingly enough, the rise of Digital Humanities coincide with the dawn of the first computers and mechanical machines of the modern era in the middle of the XXth century.

Digital Humanities, also known as *humanities computing*, started in 1949, when an Italian Jesuit priest, Father Roberto Busa, decided to do what even to this day is a monumental task: to make an index of all the words in the works of St Thomas Aquinas (and related authors), totaling some 11 million words of medieval Latin (Hockey, 2004).

In that year Father Busa met Thomas J. Watson, the founder of IBM, in the United States, and found support for his project, with the aid of the first computers of the modern era. Father Busa did not want to make a purely mechanical concordance program, where words are alphabetized according to their graphic forms (just sequences of letters), but produced a more complex and time-consuming *lemmatized* concordance, where words are listed under their dictionary headings, not under their simple forms.

⁶<http://www.slac.stanford.edu/spires/>

⁷<http://scoap3.org/>

Hockey (2004) highlight the originality and vastness of Busa's vision:

His team attempted to write some computer software to deal with this and, eventually, the lemmatization of all 11 million words was completed in a semiautomatic way with human beings dealing with word forms that the program could not handle. Busa set very *high standards* for his work. His volumes are elegantly typeset and he would not compromise on any levels of scholarship in order to get the work done faster. He has continued to have a profound influence on humanities computing, with a vision and imagination that reach beyond the horizons of many of the current generation of practitioners who have been brought up with the Internet. A CD-ROM of the Aquinas material appeared in 1992 that incorporated some hypertextual features (“*cum hypertextibus*”) [Busa 1992] and was accompanied by a user guide in Latin, English, and Italian. Father Busa himself was the first recipient of the *Busa award* in recognition of outstanding achievements in the application of information technology to humanistic research, and in his award lecture in Debrecen, He has continued to have a profound influence on humanities computing, *with a vision and imagination that reach beyond the horizons of many of the current generation of practitioners* who have been brought up with the Internet.

For a more specific definition of DH, Roncaglia (2002) provides a list of features:

- use of digital tools for ecdotic and philology
- use of mark-up languages for structure, formal, semantical representation of texts, and attachment of interpretative and descriptive meta-data
- study of instruments for attachment of multimedial meta-information
- reflection on digital documents and its features

Digital Humanities are thus an interdisciplinary field that meets literary studies and computer science: it has been suggested that nowadays DH are just exploring digital tools for a wider (and sometimes deeper) approach to texts and to textuality. Nevertheless, this new-born and hybrid field, is sometimes ostracized by both the computer science and humanities communities, which tends to be conservative and not consider as a proper field such a blurred area of research.

1.4 The Italian context

Italy is a country with a invaluable cultural heritage, due to his long and noble past. From the Romans to the Renaissance, Italy has been the location of amazing characters who has carried on the art, the knowledge and the culture of humanity. Very few nations can claim a cultural heritage so rich and beautiful as the Italian one... Though, this noble past is sometimes a burden not easy to carry. For example, Italy is one of the few countries which own two national libraries instead of one. Due to the vast amount of documents and cultural objects, and due to its troubled has often privileged a *conservative* approach towards its cultural heritage and traditions, with few openings to innovation. This approach has also been quite closed, in the aim of preservation instead of access.

Italy still lacks institutional digital libraries comparable with great national projects as the French Gallica⁸ or the American Memory⁹, and even in European collective projects as Europeana¹⁰ the presence of Italian documents is minimal (and sometimes provided greatly by non-Italian institutions as Gallica). What is more, even though Italy can boast centuries of culture (in almost every field of human creation), still little has been done to digitize this content and make it available online. Several nations, with far more humble and smaller heritages, spend greater effort in digitizing cultural heritage material.

Thus, in Italy amateur digital libraries (which followed the example of

⁸<http://gallica.bnf.fr/>

⁹<http://memory.loc.gov/ammem/index.html>

¹⁰<http://www.europeana.eu/portal/>

the legendary Gutenberg Project¹¹) still are a landmark for users and scholars too. Liber Liber is by far the most famous digital library in Italy, and it runs an amateur volunteer-driven digitization project since 1993. Wikisource, the wiki digital library whom the researcher is an administrator, largely followed the path of Liber Liber, reusing part of the early digitized collection, which is composed primarily by public domain, out of copyright Italian texts.

Therefore, within the Italian context amateur projects still play an important role for disseminating and providing access to digitized collection. Both these projects are volunteer-driven, and their communities offer a valid and interesting comparison with scholar communities of practice.

1.5 Humanities, Digital Humanities and Collaborative Digital Libraries

Father Busa states that there are only three types of *textual informatics* (Massarenti, 2002):

- **documentary**: it involves databases and repositories online
- **editorial**: it involves the dissemination of the text through different media (CD-ROM, DVD, websites, etc.)
- **hermeneutic**: it involves “statistical, classificatory observation of linguistic facts written on large masses of natural texts, and current representative.” It also involves their markup and their lexicological aspects, etc.

Furthermore, he claims that the hermeneutic is the far more interesting and complex among the others: in fact, this kind of study and observation has been little explored, and introduce us again in a field where corpora and digital libraries are laboratories, framework where to work on texts, provided with helpful tools. Despite Busa’s vision and works, little seem to have been accomplished in this field, in the Italian context.

¹¹<http://www.gutenbergproject.com>

Regarding the same topic, Tammaro (n.d.), recalls the definition of “collaboratories” from e-science, and therefore applies the same definition to this new kind of digital libraries:

[a collaboratory digital library is a] research library, which is proposed as an infrastructure for scholars within Humanities, and which includes the whole cycle of creation, dissemination and use of digital resources to support collaborative cognitive processes. The management of the collaboratory digital library considers a variety of creative collections digital and stimulates new relationships with scholars and experts, using the tools of the participatory Web.

Moreover, Rydberg-Cox (2006) defines the kind of relationships that Humanities scholars and digital libraries should reach:

- integrated access to rare and fragile sources difficult to retrieve, which are distributed in a network of cultural institutions;
- enhance and facilitate wider dissemination and use of classical sources, including also non expert audience;
- make possible new types of advanced research on digital objects, consisting in multimedia documents or large hypertexts;
- ensure long-term preservation of the digital collection.

1.6 Research hypothesis, questions and objectives

Moreover, the experience gained at CERN taught once again the researcher the importance of an open, interactive and collaborative environment for scholars and researchers. Collaborative e-science projects observed at CERN can be used as models also for different disciplines, where can be applied in minor scale.

Therefore, I decided to shift on Humanities.

Libraries, together with archives and the others *memory institutions*, have always been the “laboratories” of Humanities. Thus, a good question

for librarians and scholars should be how to effectively exploit digital libraries for research. Digital libraries could in fact boost their collaborative dimension to become frameworks for scholars and researchers.

More specifically, could be digital libraries collaborative frameworks an aid for researchers? Could Digital Humanities assume the collaborative paradigm of the e-science? Could be digital libraries be more social and collaborative? Could they be based on social softwares as wikis?

These were the background questions that introduced the present research.

Research questions came straightforwardly. It is my aim to show the importance of such collaborative projects, specifically a collaborative digital library (that it will be often referred also as a *collaboratory* digital library) for the Italian Humanities community as the object of my study. The Humanities scholars' perception towards collaboration, in Italy, has not been analyzed yet: in fact, there seems to be a closed approach towards innovation. This is a gap that the present research would like to fill, at least in part, within the limits of the involved resources. Digital Humanities community will be taken as a privileged speaker to investigate these topics. Firstly, due to their competence, expertise and engagement with the digital environment; secondly, because they are probably more interested than other scholars in the existence of collaboratories digital libraries.

The research hypothesis of the present research is thus that collaboration is a facilitator for research and innovation: thus, collaborative digital libraries should help and aid scholars in their activities. Do researchers want such a digital library? This is the main objective of the research.

Three research questions came up:

H1: Can a digital library for the Humanities community within the Italian context be *collaborative*?

H2: Can a digital library for the Humanities community within the Italian context be a *collaboratory*?

H3: Can a digital library for the Humanities community within the Italian context be built upon a *wiki*?

The objectives of the research then are:

- to analyze perception of Humanities scholars (particularly of Digital Humanities) towards collaboration and digital collaborative projects
- to investigate wikis to be a possible framework for a collaborative digital library

References

- Berners-Lee, T. (1999, April 14). *Talk to the LCS 35th Anniversary celebrations*. (Retrieved 10:43, June 13, 2010 from <http://www.w3.org/1999/04/13-tbl.html>)
- Bush, V. (1945). As We May Think. *The Atlantic Monthly*, 176(1), 101–108.
- Cammarata, V. (2007). *Innovation Oriented Workplaces The CERN Case*. University of Lugano.
- Hockey, S. (2004). The history of humanities computing. *A companion to digital humanities*, 3–19.
- Lankes, R., Silverstein, J., Nicholson, S., & Marshall, T. (2007). Participatory networks: the library as conversation. *Information technology and libraries*, 26(4), 17–33.
- Licklider, J. (1965). *Libraries of the future*. Cambridge: The MIT Press.
- Massarenti, A. (2002, August 7). L'umanesimo viaggia sul computer. *Il Sole-24 ore*, 6.
- Meta. (2010, May 12). *List of wikipedias*. (Retrieved 19:04, June 6, 2010 from http://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=1969133)
- Musser, J., O'Reilly, T., et al. (2006). *Web 2.0: Principles and best practices*. Cambridge.
- Nelson, T. (1987). *Computer lib: Dream machines*. Oxford: Tempus Books.
- O'Reilly, T. (2005, September 30). *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. (Retrieved at 18:03, June 14, 2010 from <http://oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>)

- Roncaglia, G. (2002). Informatica umanistica: le ragioni di una disciplina. *Intersezioni*, 22(3), 353–376.
- Rydberg-Cox, J. (2006). *Digital libraries and the challenges of digital humanities*. Oxford: Chandos Publishing.
- Suber, P. (2009, February 9). *Timeline of the Open Access Movement*. (Retrieved 22:23, June 11, 2010 from <http://www.earlham.edu/~peters/fos/timeline.htm>)
- Surowiecki, J. (2004). *The wisdom of the crowds: why the many are smarter than the few*. New York: Doubleday.
- Tammaro, A. (n.d.). Biblioteca digitale co-laboratorio: verso l'infrastruttura globale per gli studi umanistici.
- Tapscott, D., & Williams, A. (2006). *Wikinomics: how mass collaboration changes everything*. New York: Portfolio.
- Voss, R. (2007, June). *Open access publishing in particle physics* (Tech. Rep.). Geneva: CERN.
- Wardrip-Fruin, N., & Montfort, N. (2003). *The NewMediaReader*. Cambridge: The MIT Press.

Chapter 2

Literature review

2.1 Introduction

The purpose of the research is to explore the topic of collaboratories digital libraries and the possibility for the Italian Humanities community to have one, perhaps on a wiki software, considering advantages and disadvantages. The study concentrate on the Italian context, investigating the topic of collaboration within a community of practice as the Humanities community in Italy. Literature searching started therefore from international articles, to understand the state of the art of digital libraries (especially in Italy) with a special focus on collaborative projects and technologically advanced digital libraries.

At the beginning of the study, the researcher was more interested in the technological framework of online scholar collaboration: therefore literature browsing started in databases like ISI Web of Knowledge, Google Scholar, EBSCO. The following keywords has been searched: *wiki*, *digital libraries*, *collaboratories*, with all possible combinations. Then, related terms like *collaboration*, *e-science*, *cyberinfrastructure*, *digital humanities* (and related combinations) were explored as well. Dedicated e-journals like D-Lib and First Monday were browsed too. Unfortunately, an important synonym for cyberinfrastructure projects as VRE (Virtual Research Environment) was discovered during the last interview, and could not be exploited for the literature review.

However, it emerged during the whole research process that different terms were used to indicate similar projects, and they were differently according to the location. For example, “cyberinfrastructure” is a term used mainly in US, when “e-science” is utilized in UK. “Grid” is another keyword that is referred mainly to *grid technologies* and computing, and is used largely in European projects. “VRE” is then a term indicating single projects of “e-science”, and it can be regarded as a broad synonym of collaboratories (used mostly for STM projects). Literature review will go deeper in the analysis.

Documental analysis also included specific web-sites focused on Human-

ities and Computer Science as Griselda Online¹, Informatica umanistica², Engramma³. Wikipedia has been used as a source only for overviewing topics which were not covered by academic literature, and that often regarded digital culture (i.e. technology, net culture, wikis, Wikipedia).

Research also included cyberinfrastructure and digital humanities projects like Perseus Digital Library⁴, Stoa⁵, Suda⁶, Alpheios⁷, HRIT⁸, TextGrid⁹, that has been very helpful for defining functionalities and comprehending the *state of the art* of digital humanities collaborative projects. Several of the digital libraries and mentioned came up during interviews.

Focus was on collaboration in digital libraries and general cyberinfrastructure: aside from some historic articles, emphasis was not on STM (*Science, Technology and Medicine*) e-science projects, Only literature on digital humanities and humanities collaborative projects has been explored. Wiki projects in science and humanities were searched as well.

2.1.1 Italian literature

Unfortunately, the search of documents and articles in Italian language was almost a total failure. Very few articles were found on topic and other few were closely related. Almost no bibliography is available on collaborative digital libraries and Digital Humanities collaborative projects.

What is more, searching for English articles regarding the Italian context was unsuccessful too. Few articles were found on topic, written by Italian scholars, but they were too technologically oriented and did meet only partly the scope of the present research. This is quite significant about the state of the art of digital libraries in Italy, and will be discussed in the **Chapter 5**.

Digital Humanities do have a long and successful history in Italy: Father

¹<http://www.griseldaonline.it/>

²<http://www.digitalvariants.org/>

³<http://www.engramma.it>

⁴<http://www.perseus.tufts.edu/>

⁵<http://www.stoa.org>

⁶<http://www.stoa.org/sol/>

⁷<http://alpheios.net/>

⁸<http://ctsdh.luc.edu/?q=node/24>

⁹<http://www.textgrid.de/>

Busa, an Italian Jesuit, is one of the pioneer of one of the pioneers in the usage of computers for linguistic and literary analysis, which started the first computational lemmatization of the works of of the works of Saint Thomas Aquinas, the *Index Thomisticus*, in 1949. He is worldwide considered one of the founders of DH.

Thus, DH are a field where great successes has been achieved by the Italian community. However, in DH the approach of the single scholar remains individual: very few projects can be regarded as collaborative, and scholars tend to use digital tools for individual researches. Interviews and data analysis will highlight much better this cultural aspect of the Humanities Italian community.

Therefore, the researcher did have to seek for literature in the international context, mainly the English one. Due to iterative phases of research process, purposes were tweaked during the first part of the work, so did the scope and the coverage of the review. Suggestions from *key informants* and from the first interviews gave new and helpful directions for the research, adding important authors and articles to the bibliography. For example, a focused and exhaustive bibliography was suggested by Alison Babeu, librarian of Perseus Project.

The following review will be organized in topics and related subtopics, following keywords emerged from literature (methodology will be discussed further in the next chapter).

2.2 Collaboration

In a world of ever increasing complexity, brand new issues need brand new answers. Often, what is really needed are brand new approaches to find answers. Multidisciplinarity has often become a necessity to cope with entangled questions, multiculturalism a default approach for a globalized world. With the rise of the Web and the so-called Web 2.0, we have seen users of the Net using the ubiquity of the digital world for collaborating in any possible direction, from organizing and gathering photos, websites, even books

(Flickr¹⁰, del.ici.ous¹¹, LibraryThing¹², aNobii¹³) to create entire ecosystem of information with blogs and wikis.

Shirky (2008) provides several examples of new forms of organization through the Web, where auto-organization of normal citizen has lead to projects, actions and events unthinkable only few years ago. User empowerment is boosting innovation in many fields and giving users a new consciousness of the power of groups, communities and crowds.

Following a similar path, Surowiecki (2004) states that when a group of people reaches *diversity of opinion, independence, decentralization and aggregation*, it might happen what he calls the *wisdom of the crowds*: the empirical evidence that, statistically, a disorganized crowd can reach goals and answer questions faster and better than single but expert individuals. Examples are everywhere: from prediction markets to Delphi methods, to massive collaborative projects as Wikipedia.

Tapscott and Williams (2006) enthusiastically state that the world is facing a brand new economics they call *Wikinomics*, where *openness, peering, sharing* and *global action* are the new principles.

In this brand new world many projects are bottom up, peer produced, collaborative. Science and scholarship are coping with this new world in many ways, discussing issues and proposing solutions regarding *collaboration* (e.g. e-science, e-research), *intellectual property* (e.g. Science Commons), *scholarship communication* (e.g. Open Access) and many more. The present literature review will focus on collaboration regarding digital libraries for humanities, covering briefly literature regarding e-science and collaboration.

2.2.1 E-science, cyberinfrastructure, e-research

A number of terms are in vogue that describe the transformation of science through utilization of Grid computing, Internet-based instrumentation, and global collaboration. *E-science* (standing for "Enhanced science", used mostly in UK) and *cyberinfrastructure* (used mostly in US) are probably the

¹⁰<http://www.flickr.com>

¹¹<http://www.del.ici.ous.com>

¹²<http://www.librarything.com>

¹³<http://www.anobii.com>

most common terms used for describing this transformation in the scientific enterprise (Jankowski, 2007). Both terms will be used as synonyms in the present review. E-science is associated with “computationally intensive science carried out in highly distributed network environments”, often using “immense data sets that require *grid computing*; the term sometimes includes technologies that enable distributed collaboration, such as the Access Grid.” (Wikipedia, 2010). It is used for simulations and experiments, especially in disciplines like high-energy physics, earth science, bio-informatics. A report by the US National Science Foundation, known as the *Atkins Report* (2003) and commonly quoted as a manifesto of e-science, introduced the term *cyberinfrastructure* as “the infrastructure of distributed computer, information and communication technologies” (Jankowski, 2007) needed for innovation and the knowledge society.

If infrastructure is required for an industrial economy, then. . . cyberinfrastructure is required for a knowledge economy. — (Atkins, 2003, p. 5)

Though the visionary language and the lack of scholarly concerns, the Atkins report found great resonance: NSF established the Office of Cyberinfrastructure¹⁴, suggesting a form of institutionalization.

Another definition states:

Cyberinfrastructure integrates hardware for computing, data and networks, digitally enabled sensors, observatories and experimental facilities, and an interoperable suite of software and middleware services and tools. Investments in interdisciplinary teams and cyberinfrastructure professionals with expertise in algorithm development, system operations, and applications development are also essential to exploit the full power of cyberinfrastructure to create, disseminate, and preserve scientific data, information, and knowledge.—(Council, 2006, p. 7)

Including also humanities and social sciences, *e-research* is also a term often used to extend the original meaning of e-science and cyberinfrastructure.

¹⁴<http://www.nsf.gov/dir/index.jsp?org=OCI>

Despite the variety of terms, Jankowski (2007) suggest a list of main features of e-science, claiming they “contribute *added value* to the scientific enterprise, when combined” (p. 5):

1. International collaboration among researchers;
2. Increasing use of high-speed interconnected computers, applying Grid architecture;
3. Visualization of data;
4. Development of Internet-based tools and procedures;
5. Construction of virtual organizational structures for conducting research;
6. Electronic distribution and publication of findings.

Importance of the features differs for disciplines, but they can be regarded as the core features of e-research and cyberinfrastructure. Present research will focus especially on the feature 4 and 5, considering them within the specific field of digital libraries.

Digital libraries are an crucial component of cyberinfrastructure. Atkins (2003) introduces them in his famous report with a particular focus on data and information:

An information-driven digital society requires the collection, storage, organization, sharing, and synthesis of huge volumes of widely disparate information and the digitization of analog sensor data and information about physical objects. The digital library encompasses these functions, and research and development are needed for the infrastructures to mass-manipulate such information on global networks. Digital libraries also provide powerful tools for linking and relating different types of information, leading to new knowledge. These capabilities require new paradigms for information classification, representation (e.g., standards, protocols, formats, languages), manipulation, and visualization.

Digital libraries provide not only reference and documentation for scholars and researchers, but begin to cope with information stored in *data*. For librarians employed in research centers and universities, the shift from *libraries of documents* to *libraries of data* is disruptive. With huge amounts of data, new professionals with both librarianship and computer science skills are needed to cope with increasing complexity. The specific topic is broad and will be not covered by the present research: it will suffice to say that librarians need to rethink their roles to cope with complex issues of data acquisition, management, storage, preservation (Williams, 2009).

2.2.2 Collaboratories

While he was working for the US National Science Foundation, William Wulf coined the term *collaboratory*, merging the words *collaboration* and *laboratory* (Kouzes, Myers, & Wulf, 1996). He defined it as a

center without walls, in which the nations researchers can perform their research without regard to geographical location—interacting with colleagues, accessing instrumentation, sharing data and computational resource, and accessing information in digital libraries. — (Cerf et al., 1993)

Collaboratories are shared virtual environments providing computing and communication technologies for scientists to access, manage and analyze data, and work collaboratively.

They are strictly bounded with e-science, because are another form of collaboration. While e-science is global and tries to link and reach different scholars and disciplines around the world through digital infrastructure, collaboratories affect a *cluster* (Albert & Barabási, 2002) of researchers of the same field. Even though the principle and goal is the same, the *scale* is different.

Kouzes et al. (1996) explore issues of collaboratories, identifying 4 broad categories of collaboration among researchers:

- *peer to peer*
- *mentor-student*

- *interdisciplinary*
- *producer-consumer*

Authors suggest that

although we present these classifications as distinct types, a single collaboration may actually contain elements from several styles [...]. Nevertheless, these categories do help to show the varying communications needs researchers have as they work in different modes and how an individuals needs may change as the task or nature of the collaboration changes. The fact [...] implies that an electronic collaboratory environment should not impose a particular mode. It should instead provide a wide range of capabilities that can be quickly and easily selected and configured for the task at hand. Such flexibility addresses some of the social barriers inhibiting collaboration. —(Kouzes et al., 1996, p. 43)

Collaboration is one of the main keyword of the research, and has been questioned during all the interviews conducted. Outcomes and a deeper analysis of the term itself will be provided in **Chapter 4**.

Thus, Kouzes et al. (1996) provides an analysis of a variety of communications tools, from e-mails to Web-browser synchronization to shared whiteboards.

Collaboratories are part of e-science: the main difference is the scale of the projects. Collaboratories started in early 90's to exploit the digital world for scientific research with small teams, when cyberinfrastructure came later to gather huge resources for computing and data storage.

Using previous features proposed by Jankowsky, both of them have the goals of the

- Development of Internet-based tools and procedures;
- Construction of virtual organizational structures for conducting research.

Meanwhile, we need to explore the relationship between digital libraries and collaboratories.

2.2.3 Digital libraries and laboratories

Regarding digital libraries, Agosti, Ferro, Frommholz, and Thiel (2004) compare two different definitions, the former from computer scientists and the latter from librarians:

Digital Libraries are concerned with the creation and management of information resources, the movement of information across global networks and the effective use of this information by a wide range of users. — (Neuhold & Frommholz, 2004)

Digital Libraries are organizations that provide the resources, including the specialized 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 and economically available for use by a dened community or set of communities. — (Digital Library Federation, 1998)

Thus authors argue that definitions above share the same common view with Wulf's definition of laboratory, the view that information resources have to be accessed. So, as digital libraries are part of e-science, they are also part of laboratories, because participants need to access information to analyze and interpret it. Then (Agosti et al., 2004) list features of digital libraries:

- Authoring/editing
 - creation
 - interpretation
- Management
- Access retrieval
- Effective use

comparing them with features of laboratories:

- Interaction
- Sharing
 - data
 - resources
- Access
 - instrumentation
 - information

Authors continue their article proposing a system for annotation, arguing a mayor involvement from users and a improvement for learning and interaction with texts. Although the proposal is going towards user empowerment, this vision of digital libraries is still quite traditional, thinking of libraries only for the reference services and as a documentation access point. Still, vision of digital libraries is generally really traditional, not going far from the paper library model revived on a digital environment.

Following Tammaro (2009), we can define the collaboratory digital library as:

a research library, which is proposed as an infrastructure for scholars within Humanities, and which includes the whole cycle of creation, dissemination and use of digital resources to support collaborative cognitive processes. The management of the collaboratory digital library considers a variety of creative collections digital and stimulates new relationships with scholars and experts, using the tools of the participatory Web.

2.2.4 Visions of digital libraries

Many authors agree with the fact that vision of digital libraries is still anchored to the image of the book and the organization of traditional libraries collections (for example Tammaro, 2009; Crane, 2009). Many potentialities of digitization are not yet explored (for example, hypertextual digital libraries), ebooks are still paper document-like, too often in formats that

are proprietary and non editable from users, thus avoiding manipulation of text and annotation. As Ted Nelson says, our digital documents are mostly “paper under glass” (Nelson, 2008).

Interesting enough, the first pioneer vision of digital libraries were focused in reaching a revolution in the human-information interaction, in enhancing human mind: in systems as Vannevar Bush’s *memex* or Licklider *procog* computers were prosthesis for “augmenting human intellect” (Engelbart, 1963) for gaining a direct and disintermediated human-machine interaction. All these pioneers, from Bush to Nelson, followed the idea of building systems that aided humans to think more naturally and fitting human mind, in which (they thought) ideas and concepts come for association, non linearly. Hauben (2004) quotes from Bush:

“With one item in its grasp, [the brain] snaps instantly to the next that is suggested by the association of thoughts”. This is “in accordance with some intricate web of trails carried by the cells of the brain.”— (p. 101[10])

The term “hypertext”, however, was coined only in 1963 by Nelson, inside the amazing, ideal and impossible design of Xanadu. Meanwhile, in 1960 Licklider published *Man-computer symbiosis* and in 1965 *Libraries of the future*, in which he explored the idea of the “Procog systems”, computers able to answers defined commands that helped users to think and avoid mechanical actions. Licklider recalls a personal experience:

Throughout the period I examined, in short, my “thinking” time was devoted mainly to activities that were essentially clerical or mechanical: searching, calculating, plotting, transforming, determining the logical or dynamic consequences of a set of assumptions or hypotheses, preparing the way for a decision or an insight. Moreover, my choices of what to attempt and what not to attempt were determined to an embarrassingly great extent by considerations of clerical feasibility, not intellectual capability. — (Licklider, 1960)

Thus, computer can help humans save time and resources accomplishing mechanical tasks. Licklider continues:

The information-processing equipment, for its part, will convert hypotheses into testable models and then test the models against data (which the human operator may designate roughly and identify as relevant when the computer presents them for his approval). The equipment will answer questions. It will simulate the mechanisms and models, carry out the procedures, and display the results to the operator. It will transform data, plot graphs (“cutting the cake” in whatever way the human operator specifies, or in several alternative ways if the human operator is not sure what he wants). The equipment will interpolate, extrapolate, and transform. It will convert static equations or logical statements into dynamic models so the human operator can examine their behavior. In general, it will carry out the routinizable, clerical operations that fill the intervals between decisions. — (p. 7)

In 1965, Licklider merged this idea of a “real-time, cooperative thinking” machine with the concept of library, in *Libraries of the Future*, in 1965. Is in this book that he mention and deeply explores the *pro-cognitive utility net*, also labeled *procog*. He foresaw a digital library as a learning framework, providing documents and collections, but along with complex tools for information retrieval, data and text mining, statistical analysis, printing. Similar to Bush’s Memex, the procog was digital and went in the direction of what we call *Artificial Intelligence*. Later, he funded the legendary Douglas Engelbarts Knowledge Augmentation Laboratory at the Stanford Research Institute, the place were the computer mouse and the ability to interact with documents displayed on a computer screen would be invented; his original idea would never be accomplished, but at the same time Internet and personal computing as we know it are germinated from his work.

Douglas Engelbart too explored his whole life the idea of augmenting

Moreover, digital libraries for collaboratories and e-science share the same context, that is the scientific one: for humanities, thinks are quite different, as the interpretation of texts is the core of the discipline. If a library provides tools and services, the library becomes a framework where

to study, analyze and experiment. For example, for philologists the text itself is the data, so a library where texts are accessed can be a laboratory. Recursively, the digital library for humanists can be the collaboratory itself. This new digital library must providing services, beyond the texts. Services for text analysis, text mining, annotation. This vision of digital libraries as laboratories has been explored from the beginning, when humanities first met the digital worlds.

2.3 Digital humanities

Regarding digital humanities projects, L. Cunningham (2010) cite Christiansen, Stomblor, and Thaxton (2004), confirming the collaborative nature of libraries:

Libraries are collaborative in nature, seeking out relationships with students, encouraging sharing and cooperation.

They thus put this collaborative approach in contrast with the cultural environment of academic faculties, more closed to information sharing and collaboration:

In contrast, the culture of faculty has traditionally been “generally more isolated and proprietary”, and inter-faculty relationships are informal and difficult to document (p. 119). Digital humanities projects may be seen as changing this traditional research culture in the humanities, and librarians and the research library need to seize this opportunity to foster more collaborative relationships with faculty.

Moreover, Roncaglia (2002) provides a list of features of DH:

- use of digital tools for ecdotic and philology, particularly aimed to the realization of electronic editions of texts
- use of mark-up languages for structure, formal, semantical representation of texts, and attachment of interpretative and descriptive metadata

- study of instruments for attachment of multimedial meta-information
- reflection on digital documents and its features

2.3.1 Digital classics

Inside Humanities, classical philology can claim a long tradition in engagement with the digital environment. Father Busa himself is a member of this community of practice, and several innovative projects have been developed over few generations, such as the Perseus Project. Perseus founder, Gregory Crane, reflected for years upon the relationships between classics and the digital environment. In Crane, Bamman, Babeu, and Schreibman (2008), he recalls a vision of digital libraries and collaboratories very similar to Engelbart's works on intellect augmentation:

The great challenge for the rising generation of scholars is to build a digital infrastructure with which to expand our intellectual range. We seek to advance two effects already enabled by the digital infrastructure at hand. On the one hand, we are extending the intellectual range of individual scholars, enabling them to pursue topics that require analysis of more primary sources or more linguistic materials than was feasible with print. [...] At the same time, we want to increase the complementary effect and further extend the audiences that the products of particular cultures can reach. Machine translation is one technology that aims to advance this goal, but even the simple translation-support systems already provided in environments such as the Perseus Digital Library have for years made foreign language texts intellectually more accessible to students than print resources alone.

— (p. 2)

The authors are in fact interested in projects providing *services*, because they are aware of the fact that digital libraries will grow incrementally:

On a practical level, what will happen to print collections, large or small, if large Digital libraries become larger, more accessible and more flexible than any university library in history?

On a broader level, we are facing a shift that may eclipse the significance of print and approach the impact of writing itself: for writing first stored human ideas outside of our brains and print enhanced the effect. Digital libraries, where books read one another in however a rudimentary fashion, have already begun to separate intelligence and action from the human brain.
— (Crane, 2006)

Crane, Bamman, Babeu, Breuel, and Cerrato (2009) actually thinks of a “million book library”:

What can we do with a million books with the tools now at our disposal and which we could build? What are the research questions that emergent huge collections raise for the historians, literary critics, and other humanists who study their contents and for the computer and information scientists who develop methods with which to process digital information in general?
—(p. 2)

Among the potential services and features of this ideal library, the authors suggest:

Multitexts: Scholars have grown accustomed to finding whatever single edition a particular collection has chosen to collect. In large digital collections, we can begin to collate and analyze generations of scholarly editions, generating dynamically produced diagrams to illustrate the relationships between editions over time. We can begin to see immediately how and where each edition varies from every other published edition.

Chronologically deeper corpora: We can locate Greek and Latin passages that appear any where in the library, not just in those publications classicists are accustomed to reading. We can identify and analyze quotations of earlier authors as these appear embedded in texts of various genres.

New forms of textual bibliographic research: We can automatically identify key words and phrases within scholarship,

cluster and classify existing publications, generate indices of particular people (e.g., Antonius the triumvirvs. one of the many other figures of that name, Salamis on Cyprus vs. the Salamis near Athens). Such searches can go beyond the traditional disciplinary boundaries, allowing students of Thucydides, for example, to analyze publications from international relations and political philosophy as well as classics. — (p. 5)

In fact, infrastructure (in this case, cyberinfrastructure) is becoming more and more important for Humanities and their goals. According to Crane, Seales, and Terras (2009):

Infrastructure provides the material instruments whereby we can produce new ideas about the ancient world and enable other human beings to internalize those ideas. Infrastructure includes intellectual categories [...], material artifacts [...], buildings [...], organizations [...], business models [...], and social practices [...]. — (p. 5)

Yet, even in this pioneer discipline cultural barriers are yet to be eliminated, and they seem the most difficult obstacles. Authors continue:

The greatest barrier that we now face is cultural rather than technological. We have all the tools that we need to rebuild our field, but the professional activities of the field, which evolved in the print world, have only begun to adapt to the needs of the digital world in which we live hardly surprising, given the speed of change in the past two decades and the conservatism of the academy. — (p. 7)

Moreover, authors confirm the importance of collaboration and openness in cyberinfrastructure:

Collaboration: While the final form of the papers in this collection may be familiar, their production and content reflects a fundamental change in scholarly practice: the majority of the

papers published here have multiple authors, while the single-author papers either report on group projects or on general methods whereby classicists can create interoperable data. **Open access and Open source production:** All of the scholars who have contributed to this collection depend upon open access and open source production. [...] In cases where authors are making particular arguments at a particular point in time, open access allows third parties to locate and automatically analyze what they have produced: search engines such as Google can index and then deliver their arguments to any one online; more specialized text mining systems could analyze what has been written to search for trends in scholarship or to apply specialized services designed for classics. — (p. 17)

2.4 Wiki

Ward Cunningham, the *wiki* creator, originally defined wiki as “The simplest online database that could possibly work.” (W. Cunningham, 2003)

A wiki is a user-editable website that allows the “easy creation and editing of any number of interlinked web pages via a web browser”, using often a simplified markup language (Wikipedia, 2010). Wikis are often used to create collaborative websites, for personal note taking, in corporate intranets, and in knowledge management systems.

Cunningham created the first wiki, the *WikiWikiWeb*, in 1995, for the *Portland Pattern Repository*. He was in part inspired by Apple’s *HyperCard* program, which allowed users to create virtual “card stacks” of information with a host of connections, or links, among the various cards.

HyperCard in turn drew upon the by Bush’s *memex*: that machine would allow readers to annotate and create links between articles and books recorded on microfilm. “HyperCards’ stacks’ implemented a version of Bush’s vision, but the program relied upon the user to create both the text and the links. For example, one might take a musical score of a symphony and annotate different sections with different cards linked together.” (Britannica, 2010)

Then Koblas (2006) analyzes deeply wikis and their structure, history and usages.

Wiki is a *social software*, software that enhances social interaction, collaboration and information sharing: furthermore, it fosters the growth of communities based on user groups activities.

Wikis are tools designed to be driven by user groups and written by more than one author, therefore the communication model is *all-to-all*, and not *one-to-all*. Unlike blogs, they are structured page by page, and divided by content.

It is meaningful that the term “wiki” is referred both to wiki web-sites and to the software frameworks behind them: wikis are designed for collaboration and tend to gather a community to populate them. Thus, there is a deep link between a wiki and his community: one can’t live without the other. Interaction between authors, wiki web sites and wiki softwares is so high that often, when we refer to “a wiki”, we intend a blurred unit of all of them.

The term *wiki* comes from Hawaiian “wiki wiki”, that means “quick, fast” (W. Cunningham, 2003): wikis are environments in which is very quick and easy to edit and contribute. You click on the **Edit** button of a page, you change the text, you click on the **Save** button and your edit is saved. The new version of the page is just online and available to everyone, ready for another edit. Pages are saved *incrementally*, every version of the same page is saved, readable and restorable directly from the **History**, a special system page that is the archives of all the saved version of a specific wiki page. Moreover, every edit of a user is saved in a specific log: the system, called *audit trail*, allows the *tracking* of all the contributions and edits of a particular user. Both these tools (*Pages history* and *Audit trails*) grant the community the possibility to limit errors and vandalism and restore clear version of pages, under the principles of *soft security* (Wiki, 2008), which tends to trust social forces for keep the order in the project. Hard security instead is codified and trust more technology than humans, and cannot discriminate errors from attacks, and does not allow discussion and it’s not responsible.

Wagner (2004) summarize features of wikis as follows:

Table 2.1: Wagner, Wiki Design Principles

| Principle | Explanation |
|--------------------|--|
| <i>Open</i> | If a page is found to be incomplete or poorly organized, any reader can edit it as he/she sees fit. |
| <i>Incremental</i> | Pages can cite other pages, including pages that have not been written yet. |
| <i>Organic</i> | The structure and text content of the site is open to editing and evolution. |
| <i>Mundane</i> | A small number of (irregular) text conventions will provide access to the most useful (but limited) page markup. |
| <i>Universal</i> | The mechanisms of editing and organizing are the same as those of writing so that any writer is automatically an editor and organizer. |
| <i>Overt</i> | The formatted (and printed) output will suggest the input required to reproduce it. (E.g., location of the page.) |
| <i>Unified</i> | Page names will be drawn from a flat space so that no additional context is required to interpret them. |
| <i>Precise</i> | Pages will be titled with sufficient precision to avoid most name clashes, typically by forming noun phrases. |
| <i>Tolerant</i> | Interpretable (even if undesirable) behavior is preferred to error messages. |
| <i>Observable</i> | Activity within the site can be watched and reviewed by any other visitor to the site. |
| <i>Convergent</i> | Duplication can be discouraged or removed by finding and citing similar or related content. |

Wikis, therefore, are much more than social softwares. They are spaces for social construction of knowledge. (Koblas, 2006, p. 13)

Issues

- Intellectual property: as previously said, collaboration is easier when licenses are open. Wiki contributors are collectively owner of its content.

- No authority, bounded with an open and informal peer review, in many different dimensions (grammar, project coherence).

Following a summary developed by Koblas (2006, p.7-10)

- **Pages are accessible by a standard web browser:** this guarantees a maximum of participation and the possibility to collaborate everywhere and with every system;
- **Easiness of writing:** editing pages is simple like writing using a common word processor software. The WYSIWYG is guarantee for users that dont know or dont want to use the mark-up language.
- **Easiness of linking:** internal or external links can be made simply, it permits authors to create a structure from a personal point of view.
- **Real time updating:** the technical time of publishing (table 2.0) is reduced to a simple passage Edit ↵ Save. The authors can manage personally the publishing phase and can immediately correct errors or mistakes.
- **Collective editing:** remains the base concepts of Wiki revolution. One common document or collection of documents can be created or edited by many authors.
- **History and tracking:** is always possible to revert to a previous version and find the authors of each one. This represent a marginal guarantee on work quality.
- **Visualization of latest version:** is useful to compare near versions of the same documents and to highlight differences.
- **Changes notifications:** using the RSS (*Really Simple Syndication*) feed is possible to be updated on what happens on the Wiki and which kind of activities users do.
- **Search and navigation:** is simple to find information and topics published by other users.

- **Simple permission structure:** there are only three levels of users: the readers, the editors and the administrators.

Wiki and DL

Frumkin (2005) suggests uses of Wikis in digital libraries as *virtual reference*, *content management system* and *annotation tool*.

Krowne (2003), co-founder of the mathematical peer-produced encyclopedia *PlanetMath.org*¹⁵, explains present the Noosphere system¹⁶ as a " *case study to demonstrate common based peer production digital library system design.*"

Krowne introduces the Yochai Benkler's notion of *commons-based peer production*:

The term "commons-based peer production" (CBPP) was introduced by Yochai Benkler in his theoretically grounded explanation of a recently recognized Internet-based phenomenon. This phenomenon includes the production of the Linux kernel by a worldwide and shifting team of volunteers, as well as web sites like Kuro5hin and Wikipedia. The defining characteristic of CBPP is the voluntary and community-regulated production of an intellectual work. Benkler notes that the rise in CBPP is because the Internet lowers certain communication and collaboration barriers, allowing CBPP to flourish and to serve as a viable alternative to produce a large and complex intellectual work. Eric Raymond also observed (and participated in) CBPP at an earlier time, but called it the "bazaar model", which he contrasted with the "cathedral model" of traditional production (Raymond, 1999). In Benkler's economically grounded exposition, the latter would be called "firm-based production". Benkler also discusses the open market as another vehicle of production,

¹⁵<http://planetmath.org/>

¹⁶Noosphere, the software that relies behind Planetmath, is a complex and interesting framework similar to a wiki, but technically more defined and advanced: nonetheless, for the purpose of the present research, it will be considered along as a wiki because of the high similarity of collaboration procedures and mechanisms involved.

contrasting it with firm-based and ultimately commons-based production.

The present definition lead us to the same principles and mechanisms that originates wiki projects as Wikipedia, that root in early days of the free software/open source phenomenon. The “bazaar model”, the concept introduced by Eric Raymond in his well-known essay “The Cathedral and the Bazaar” indicates the new way of peer production, when different contributes by different people create content that is often freely available on-line. Krowne give also specific motivations for building a digital library the CMPP-based way:

Table 2.2: Krowne, Building a DL the CMPP-way

| Category | Motivations |
|----------------------|--|
| <i>Philosophical</i> | Spirit of camaraderie, democracy, altruism; aversion to hierarchy/command; rejection of IP regimes; dislike of mixing money and learning. |
| <i>Logistical</i> | Knowledge distributed unevenly and/or widely; inflexibility of centralized effort; experts too busy doing work to commit entirely; DL builders not experts. |
| <i>Fiscal</i> | No or low funding; inability to provide major coverage using works for hire. |
| <i>Optimal</i> | More material; more peer review; more up-to-date content; greater involvement of readers/learners; new treatments. |

References

- Agosti, M., Ferro, N., Frommholz, I., & Thiel, U. (2004). Annotations in digital libraries and laboratories—facets, models and usage. *Research and Advanced Technology for Digital Libraries*, 244–255.

- Albert, R., & Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47–97.
- Atkins, D. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. National Science Foundation.
- Britannica, E. (2010). *wiki*. (Retrieved June 24, 2010, from Encyclopedia Britannica Online: <http://www.britannica.com/EBchecked/topic/1192819/wiki>)
- Cerf, V., Cameron, A., Lederberg, J., Russell, C., Schatz, B., Shames, P., et al. (1993). *National collaboratories: Applying information technologies for scientific research*. Washington: DC: National Academy Press.
- Christiansen, L., Stomblor, M., & Thaxton, L. (2004). A report on librarian-faculty relations from a sociological perspective. *The journal of academic librarianship*, 30(2), 116–121.
- Council, N. (2006, July). *NSF's Cyberinfrastructure Vision for 21st Century Discovery*.
- Crane, G. (2006). What do you do with a million books? *D-Lib Magazine*, 12(3), 1.
- Crane, G., Bamman, D., Babeu, A., Breuel, T., & Cerrato, L. (2009). Classics in the Million Book Library. *Digital Humanities Quarterly*, 3(1).
- Crane, G., Bamman, D., Babeu, A., & Schreibman, S. (2008). ePhilology: when the books talk to their readers. *A Companion to Digital Literary Studies*.
- Crane, G., Seales, B., & Terras, M. (2009). Cyberinfrastructure for Classical Philology. *Digital Humanities Quarterly*, 3(1), 1–27.
- Cunningham, L. (2010). The Librarian as Digital Humanist: The Collaborative Role of the Research Library in Digital Humanities Projects. *Faculty of Information Quarterly*, 2(2).
- Cunningham, W. (2003, November). *Correspondence on the etymology of wiki*. (Retrieved 10:43, June 21, 2010 from <http://c2.com/doc/etymology.html>)
- Engelbart, D. (1963). A conceptual framework for the augmentation of man's intellect. *Computer-supported cooperative work: a book of read-*

- ings*, 52.
- Frumkin, J. (2005). The wiki and the digital library. *Perspectives*, 21(1), 18–22.
- Hauben, M. (2004). *History of ARPANET*. (Retrieved 23:13, June 23, 2010 from <http://www.dei.isep.ipp.pt/~acc/docs/arpa.html>.)
- Jankowski, N. (2007). Exploring e-science: An introduction. *Journal of Computer-Mediated Communication*, 12(2), 549–562.
- Koblas, J. (2006). *Wikis: tools for information work and collaboration*. Oxford: Chandos.
- Kouzes, R., Myers, J., & Wulf, W. (1996). Collaboratories: Doing science on the Internet. *Computer*, 29(8), 40–46.
- Krowne, A. (2003). Building a digital library the commons-based peer production way. *D-Lib magazine*, 9(10).
- Licklider, J. (1960). Man-Computer symbiosis. *IRE Transactions on Human Factors in Electronics*, 1, 4–11.
- Nelson, T. (2008). Geeks Bearing Gifts.
- Neuhold, E., & Frommholz, I. (2004). Digital libraries as experienced by the editors of the journal. *International Journal on Digital Libraries*, 4(1), 1–2.
- Raymond, E. (1999). The Cathedral and the Bazaar. *Knowledge, Technology & Policy*, 12(3), 23–49.
- Roncaglia, G. (2002). Informatica umanistica: le ragioni di una disciplina. *Intersezioni*, 22(3), 353–376.
- Shirky, C. (2008). *Here comes everybody: the power of organizing without organizations*. New York: The Penguin Press.
- Surowiecki, J. (2004). *The wisdom of the crowds: why the many are smarter than the few*. New York: Doubleday.
- Tammaro, A. (2009). Biblioteca digitale co-laboratorio: verso l'infrastruttura globale per gli studi umanistici.
- Tapscott, D., & Williams, A. (2006). *Wikinomics: how mass collaboration changes everything*. New York: Portfolio.
- Wagner, C. (2004). Wiki: A Technology for Conversational Knowledge Management and Group Collaboration. *Communications of the AIS*, 13, 256–289.

- Wiki, M. (2008, September 7). *Soft Security*. (Retrieved 10:49, June 21, 2010 from <http://usemod.com/cgi-bin/mb.pl?SoftSecurity>)
- Wikipedia. (2010, June 18). *E-science*. (Retrieved 17:33, June 24, 2010 from <http://en.wikipedia.org/w/index.php?title=E-Science&oldid=368766879>)
- Wikipedia. (2010, June 24). *Wikipedia*. (Retrieved 17:39, June 24, 2010 from <http://en.wikipedia.org/w/index.php?title=Wikipedia&oldid=369721268>)
- Williams, K. (2009). A Framework for Articulating New Library Roles. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, 265, 3–8.

Chapter 3

Methodology

3.1 Research Questions and Objectives

The core hypothesis of the present research is the importance of collaborative projects in scholar communities, also in the Humanities context. Italian digital humanities community has been contacted to be asked about approach of Humanities scholars towards collaboration, particularly in digital projects. To explore issues and potentialities of collaboration in digital humanities project, especially regarding Italy, the researcher chose three correlated questions:

H1: Can a digital library for the Humanities community within the Italian context be *collaborative*?

H2: Can a digital library for the Humanities community within the Italian context be a *collaboratory*?

H3: Can a digital library for the Humanities community within the Italian context be built upon a *wiki*?

Research questions are here proposed as a hierarchy, because they are dependent in different ways. *Collaborative* is a broader term than *collaboratory*: it is a more general question that though served perfectly the aim to investigate potentialities of collaboration in digital humanities projects. Then the term *collaboratory* specifically refers to the e-science and cyberinfrastructure area, which was important for indicating the scholar context. Furthermore, questions provided in the presented sequence did help interviewees to understand and slowly understand topics proposed. In fact, not all interviewees did have a clue of what e-science and cyberinfrastructure is, although they were prepared on the single topics or examples.

Finally, the researcher proposed a *wiki* system a collaborative digital library to investigate a concept of “strong collaboration”, defined as *collaborative editing*.

Therefore objectives of the study are the following:

- to analyze perception of Humanities scholars (particularly of Digital Humanities) towards collaboration and digital collaborative projects

- to investigate wikis to be a possible framework for a collaborative digital library

3.2 Introduction

Following the *research hierarchy* model introduced by Lincoln and Guba (1985), the research **paradigm** is *interpretivism*

the research **methodology** is *qualitative*

the research **method** is *case study* and *focus group*

the research **techniques** are *interviews* and *focus group*

the research **instruments** are *human* and *computer*.

3.3 Interpretivism

Adopting the major view assumed in information science, empirical interpretivism and specifically *constructivism* will be the chosen paradigm for the present research, as it deals with “investigation in natural settings of social phenomena” (Pickard, 2007, p. 11). This study explored opinions and insights of a community of practice, gaining data to understand cultural assets and possible directions of online collaboration between scholars and researchers.

Although it would not be accurate to state that “critical theory” has been used as a principal paradigm, it is correct to argue that this study goes in a parallel direction. In fact, “collaboration” is an crucial topic that has been investigated by the research and it touches the real nature and core of disciplines here involved: especially in the humanities, collaboration (regarding authorships, attribution, intellectual property and related matters) is an extremely *sensitive* topic, that would need further interdisciplinary and multidisciplinary research to be understood and comprehended. The present study offers just a glimpse of approach towards collaboration (especially online) in the Humanities. Investigating facilitators and barriers for collaboration was a related objective of the research and fundamental for answering research questions.

Pickard (2007, p. 13) paraphrases a definition by Greene (1990), saying

that “positivists and positivists can be thought of as social engineers, interpretivists as storytellers and critical theorists as catalysts of social change”: though the present research does not dare to be a *catalyst of social change*, it sincerely suggests a precise direction for the future of digital libraries and digital scholarship.

3.4 Qualitative research

Therefore, perspective could only be *qualitative*. The research reached a defined community of practice to ask if a collaboratory digital library could be useful and helpful. The main data collection technique would be *interviews*, and using as the main research instrument the “the researcher himself” (Pickard, 2007, p. xvii). Nonetheless, as great effort was spent in developing a software framework for the study, it shall be stated that the computer itself has been a research instrument too. A complete description of the system used will be further provided.

As the object of study are people, the interpretivist paradigm takes place. Following Heisenberg (1958), and the succeeding post-positivist approach, researchers and respondents are mutually influenced and interact in many ways.

Researcher must consider himself as an “instrument” (Maykut, Morehouse, & Morehouse, 1994), and can apply his tacit knowledge to cope with the complexity of people and human relations in order to “produce meaning from data and using that meaning to develop theory” (Pickard, 2007, p. 14).

Pickard (2007, p. 15) adapts the research design model from Kumar (1999) and Lincoln and Guba (1985; 1998, p. 104) which illustrates the whole process “within the boundaries of trustworthiness: transferability, credibility, dependability and confirmability” (p. 14). The essential components of a qualitative research design are “literature review, theoretical framework, fieldwork in a natural setting, using a human instrument, purposive sampling, appropriate data collection techniques, inductive analysis, emergent design, iteration of activities, grounded theory, negotiated outcomes” (*ib.*).

Interaction with participants is the core of qualitative research: there-

fore, for establishing trustworthiness and credibility, every outcome of the research needs verification with respondents. Both data and interpretations must be reviewed by involved persons (Erlandson, 1993). Stake (1995, cited in Pickard, 2007) emphasizes the importance of “member checking” stating that it always led his report to improvement.

Moreover, the present research roots its design in what Lincoln and Guba (1985) call “emergent design”, namely that in qualitative research “the research design must therefore be ‘played by ear’; it must unfold, cascade, roll, *emerge*” (p. 203, cited in Pickard, 2007, p. 14).

Emergent design is specific to social sciences, where is very unlikely to have clear and defined guidelines, procedures and boundaries at the beginning of the study. The researcher is in front of the unknown and must unfold it. Therefore, the design itself must emerge as the research process goes. This is exactly what happened for the present research, where aims and goals of the researcher remained still when techniques and tools changed and were tweaked to fit the purposes.

Indeed, Warden and Wong (2007) mention that qualitative analysis is *iterative*. In the context of this study, the researcher used such approach in order to verify facts or fill gaps that has been forgotten or uncovered during the initial investigation. This happened especially with interviews: websites and projects suggested in the process of interviews were visited to verify information and gain in-depth understanding of the data.

3.4.1 Theoretical and conceptual framework

As the analysis of the literature progressed, a concept map was started to illustrate the theoretical and conceptual framework of the research. Keywords were extracted from literature with a procedure that will be further explained.

However, for now it will suffice to say that from literature concepts arose that draw a rich picture illustrated below in fig.1. The map was drawn with open source software Cmap¹, and it is the first attempt of the researcher with that software².

¹<http://cmap.ihmc.us/>

²It may be helpful to state that this was not the first concept map software used by

The theoretical framework is the overall picture of the domain here presented: the different color represent topics belonging to different scopes and with different level of importance for the present studies. Links between nodes are labeled and describe the relation between topics. The conceptual framework can be considered the set of the red, orange and yellow topics, which for a subnet of the whole theoretical framework graph. At the center of the map it will be noticed the two main topics (colored in red) of the present study (*Collaboratory digital library* and *Digital Humanities community*), while around them the net stand the crucial topics of *e-science*, *collaboratories*, *digital libraries* and *collaboration* (colored in orange). Further, stand the related important topics of digital *humanities* and *wiki* (still colored in orange): after those, more peripheral, stay yellow colored nodes representing the *features* of the main topics (e.g. *services* as a feature of *collaboratories*, *collaborative editing* as the core of *wikis*, etc.)

The map must be read as an incomplete portrait of the research theoretical structure; yet it is an helpful guide to understand and comprehend the scopes and the domain of the study.

the author, who had other experiences in drawing maps. This program is by far the best tool used for this kind of work, and it is highly recommended for similar tasks.

3.4.2 Trustworthiness

There are several methods for establishing trustworthiness in qualitative research. Lincoln and Guba (1985) proposed four concepts to gauge the value of research, being

1. truth value
2. applicability
3. consistency
4. neutrality

Accordingly, Pickard (2007) adapted the model for qualitative research, proposing criteria as

1. credibility
2. transferability
3. dependability
4. confirmability

Credibility

Credibility in qualitative research is proved by everlong engagement with participants, triangulation of data collection techniques and member checking. Due to the specific design of the research, which needed in-depth investigation of participants thoughts and insights, triangulation of techniques has not been possible. Aside from interviews, the researcher did not find an appropriate technique to triangulate data gathered with interviews. Thus, he concentrated on member checking.

All transcriptions of the interviews have been submitted to interviewees for checking. Member checking revealed helpful for both interviews (that cleared some concepts and insights) and the researcher, who gained a better understanding of topics and discussion. The researcher discussed and submitted insights and data gathered to his advisor during the whole research process. Furthermore, he explained some of the results of the outcomes directly to interviewees to have feedbacks.

Transferability

Context of the study is specific: a community of practice in the Italian context focusing his study and attention on ancient texts was asked about collaborative digital libraries and collaboration facilitator and issues. Although some general results of the study could be extended to the whole field of Humanities, transferability of the outcomes of the research could only be guaranteed if important factors of the present context would be controlled.

The most important factors appear to be *cultural environment* and *approach towards collaboration*. Humanists generally are considered to be quite individualistic and relatively closed to collaboration: this attitude may change between nationalities, cultures, subdisciplines, communities of practice. Data analysis seems to suggest that there are few differences between the Italian and the English context towards collaboration and authorship in Humanities. However, data gathered are not enough to make sure statements: thus, it will be better to leave approach to collaboration as a variable.

In fact, the present research was based on the Italian context, which will show to be relatively closed to collaboration and communication between scholars. Further and interdisciplinary research is needed to enlight collaboration issues and approaches of determined communities, and this was not the scope of the present study. Anyway, interviews suggested that each community of practice has his specific way of dealing with collaboration. Furthermore, collaboration through Internet and general digital tool vary dramatically between subdisciplines and singular individuals.

Nonetheless, data gathered suggests that, generally, “*attention to textuality*” and focus on texts is the fundamental core of different humanistic disciplines. One of interviewees stated, regarding collaboration within different Humanities disciplines (and communities of practice):

Scientific communities change because it the *object of their study* changes, but their *methodologies* are more or less the same, they converge. [In our community] different aspects of language do change, of course. But from the perspective of “management of textuality”, it does not change much. There are huge

differences [between communities], even on issues of techniques, procedures and languages, but it's always *attention to textuality*.

Within the same digital humanities community different disciplines share little in common but focusing their studies on ancient texts with a particular attention to digital tools. Tools, approaches and aims can therefore be different, but if the approach remains the same outcomes of the research would be still valid.

Likewise for difference of *language*: Italian is not much different from others Latin languages, and the same approach for Italian digital libraries could fit Chinese or English digital libraries as well. Therefore, language seems no to be a real barrier for transferability of the outcomes of the present study.

Dependability

The researcher has been in constant contact with his advisor for the whole length of the research process. The advisor got a report of the work at least every week, and guided the researcher regarding every aspect of the methodology. The advisor and the researcher had frequent conversation via e-mail, chat and often telephonic meetings, which were recorded and re-listened by the researcher not lose advisor's comments and feedbacks. The advisor received every transcription of the interviews and followed the study at each stage.

Confirmability

Confirmability is crucial to limit the bias of the researcher. The raw data gathered by the researcher (as notes, TAGs and the research diary) are all contained in the wiki system, that acted as both as a case database and a framework during the whole research process. Transcriptions of interviews are in Italian and they will be available under explicit request to whom may be interested.

3.4.3 Limitations

Limitations of the present study are physical and conceptual. Time and economic resources obviously limited the scope of the research; at the same time, the study suffers from some methodological flaws and bias.

Time and economic limitations

The main limitations of the study are concerned with time and resources. The researcher had four months to design, execute and complete the research, and several improvements could have been applied with a major amount of time available. Furthermore, with no economic resources, the researcher could not travel much to meet his interviews (it happened only with Eco), and thus interviews were conducted online. This resulted to be much more easy and cheap, but avoided possibility of direct observation and a real-life interaction with participants.

Due to the same scarcity of economic and temporal resources, the researcher could not conduct a parallel but complementary study on the community of Liber Liber, the volunteer driven digital library of Italian public domain texts, (the Italian analog of the Gutenberg project). The comparison between a scholar and a volunteer community of practice, both working with digital libraries, was foreseen at the beginning of the study (some of the scholar interviewees of the study were also part of the Liber Liber community). The researcher contacted the Liber Liber community several times, through their mailing list, to organize an online focus group, but the attempts were not successful. The researcher spent a lot of time to find a good and stable framework for having an online focus group, but the Liber Liber community revealed to be more busy and “digital divided” than expected. Few members of the community replied to the request, but each of them had problems to find a day for the meeting or declared issues in using a computer for a video-conference.

It is the opinion of the researcher that holding a real focus group in Rome (the city where several members of the Liber Liber community live) could have been much more successful. Unfortunately, there were absolutely no time and economic resources to organize it.

The limited amount of time did lead the author to slowly abandon the complex framework created on the wiki³, for the last weeks of the process. The wiki framework revealed as very helpful at the beginning, but too time-consuming at the end, when the presentation of results had to be written.

Methodological flaws

As this is the first research of the author, expert reviewer could find some methodological flaws in the research process. The author used a full constructivist approach that lead him to construct the research day by day, gathering data wherever possible and tweaking the process as the theory was emerging. This may have caused some flaws in the methodological structure, due to the inexperience of the researcher. To avoid these flaws, the researcher have been in constant contact with the advisor, who helped him trough the whole process.

Experiences, qualifications and possible biases

Qualitative research and analysis is shaped by both the subjects and researchers characteristics such as experiences, qualifications and even biases (Warden & Wong, 2007)(Warden and Wong, 2007). The researcher of this study has been working as a volunteer in the Italian version of Wikisource since 2005. Wikisource, a wiki digital library, is a sister-project of the well-know free encyclopedia Wikipedia. As Wikipedia, Wikisource relies on the volunteer work of a community of active users, who upload, format and proofread public domain texts. Therefore, the researcher developed a personal (and perhaps biased) view about collaboration of communities of practice, grounding his experience in a volunteer environment. Moreover, in the last 5 years he has been exposed to many discussions, readings, conferences, workshops and lectures about collaboration, digital libraries and wikis.

Thus, the choice of collaboration, collaborative digital libraries and communities as core topics of the research is significant and not neutral for the researcher, who is deeply convinced of the goodness of knowledge and information sharing. Furthermore, the author is involved in a controversial

³The wiki framework will be explained further.

project as Wikipedia, which in few years destroyed the well-known idea that only experts and scholars can write an encyclopedia. Even though there is an ongoing debate on issues quality, review and evaluation, it is nonetheless a fact that this peer-based bottom-up approach is a brand new way of knowledge production, which is still far from being understood.

In a case study approach, the investigator cannot be an outside observer. In such cases, biases are unavoidable. It is however patent that bias should not affect the result of the research significantly.

On the other hand, it is hoped that engagement with literature, observation and different opinions decreased researcher's bias and increased credibility of the study.

3.5 Case study

The researcher selected to investigate the topic of collaboration among the Humanities community in the Italian context, to discover the possibility of the existence of collaborative digital libraries: therefore, case study was the most obvious choice.

Experimental research, ethnography, action and historical research were suddenly excluded for obvious reasons. Experimental research is rarely used in a study that involves human subjects opinions and insights, and ethnography demanded prolonged engagement within the context, as well a more sociological approach. Action research is used for investigating "organizational functions" and "improving service provision, encouraging reflective practice and structuring and disseminating experience to the wider community" (Pickard, 2007, p. 134). Historical research was not helpful for exploring contemporary issues that do have little history.

Surveys too were estimated not to fit the objectives of the study: they would aim to "study relationships among specific variables, which are identified at the outset of the research and stated as either an hypothesis or a research question, or to describe certain characteristics of the population" (Pickard, 2007, p. 95). Surveys also need a probability sampling that was not suitable for the design of the research, because research questions needed members of the Humanities community expert in digital libraries and

collaborative projects. This was due to the fact that it is very unlikely to gain meaningful data about a specific and advanced topic as “collaborative digital libraries” if participants are not aware of that possibility and have no experience in the field.

Theoretically, Delphi group was a possible choice as a method: nonetheless, it was considered too structured and highly complex to organize for a first-time researcher as the author. Furthermore, lack of economic and time resources, as well as the experience of the researcher, did automatically exclude a possible application.

On the other hand, case study is an immersive research method that allows researcher to study a whole context and develop an in-depth analysis of a single case. It allows purposive sampling and a very constructivist approach, as well as engagement with participants. Therefore, *Case study* was the research method chosen for the present research.

Pickard (2007) reports a quote from Yin (2002, p. 23) as a definition of *case study*:

[case study is] an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used — (Yin, 2002, p. 23)

A case study is a well-suited strategy to empirically investigate “contemporary phenomena”, such as collaboration among humanists and potentialities of collaborative digital libraries for researchers and scholars.

Case study is a term used for different methods, although is very specific. It should operate in well delimited boundaries: for the present study, the Humanities community in the Italian context was therefore chosen for several reasons. It can be regarded as a broad *community of practice* (CoP), following definition by Wenger and Snyder (2000, p. 140)

[CoP] are groups of people informally bound together by shared expertise and passion for a joint enterprise — engineers engaged in deep-water drilling, for example, consultants who specialize in strategic marketing, or frontline managers in charge

of check processing at a large commercial bank. Some communities of practice meet regularly for lunch on Thursdays, say. Others are connected primarily by e-mail networks. A community of practice may or may not have an explicit agenda on a given week, and even if it does, it may not follow the agenda closely. Inevitably, however, people in communities of practice share their experiences and knowledge in free-flowing, creative ways that foster new approaches to problems.

And Wenger (2006)

Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly.

As a CoP, the Humanities community in the Italian context:

1. geographically circumscribed
2. naturally defined
3. easily reachable
4. sharing the same mother language
5. sharing the same object of study

Operating in the same context, the community shares the advantage of working with texts that are mainly out of copyright, allowing any sort of derivative work or analysis. Avoiding all the legal issues related with intellectual property make the community free to operate and work with texts in an open and potentially innovative environment. This reason made it perfectly fit the objectives of the present study: the environment was therefore suitable for a delimited and in-depth study about collaboration and potentialities of laboratories.

Case study is an “iterative process: once in the field the researcher will allow the design of the study to develop as he gains into the salient issues” (Pickard, 2007, p. 87). The whole research process followed as much as possible phases identified by Lincoln and Guba (1985):

1. orientation and overview
2. focused exploration
3. member checking

In the first phase, a research focus had been established, namely “collaboratories digital libraries”, and “issue subquestions” (Stake, 1995) followed naturally through the process.

The digital humanities community has been chosen to be the *unit of analysis*.

A personal wiki system has been used during the whole research process, acting also as a *case database*, storing notes and transcriptions of interviews. Customized folders archived the audio files of the online interviews.

Second phase, focused exploration, comprehend data collection and iterative analysis, which means reacting properly to emerging themes. As Pickard (2007) suggests, “one of the greatest case study benefits is the ability to respond as [the researcher’s] knowledge of the case increases”.

Third phase include member checking and verification of analysis: exiting the field is then the final stage of the research process, when all data cease to reveal new information.

3.6 Target group and sampling

Interviews participants has been chosen following purposive snowball sampling. All of them (but one) were Humanities scholars working in advanced digital projects. The research questions needed data that would have been answered only by people with great knowledge of online projects, In fact, all the participants worked in the field of Digital Humanities: the main parameter for the choice was the in-depth knowledge of these scholars in Digital Humanities and the digital environment. Though their expertise were different disciplines of Humanities (mostly were philologists), every participant did have competence in Digital Humanities related project and experience with online texts and digital libraries.

3.6.1 Purposive snowball sampling

Pickard (2007, p. 64) cites Patton (2002, p. 169) about purposive sampling

The logic of purposeful sampling lies in selecting *information-rich cases* for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the research

Purposive sampling is the preferred practice for case study method, and can be run in two different ways, *a priori criteria* sampling and *snowball* sampling.

As the present study is the first attempt of the researcher to qualitative research, it demanded some boundaries to the sample, and Flick (2002) suggests *a priori criteria* to be more helpful for “analyzing, differentiating and perhaps testing assumptions about common features and differences between groups” (p. 63, cited in Pickard, 2007, p. 64).

Nonetheless, Pickard (2007) also highlights that “in any bounded system there are “key informants” who will have a great deal of knowledge about the case as a whole and what goes on at a variety of levels within the case”. As the research aimed to understand and investigate opinions of the Humanities community Italian context regarding collaboration and collaborative digital libraries, help of key informants was estimated to be extremely useful and snowball sampling seemed to fit perfectly the purpose and the interactive design of the research. Thus, snowball sampling was actually chosen: commonly, its exit strategy for is *redundancy* of information gathered (Lincoln & Guba, 1985). Moreover, literature suggests also that even though the choice of the first key informant is likely to be biased, the subsequent gathering of the participants reduces the bias (Ford, 1975).

3.6.2 Key informants

Key informants of the study were two, one from the US context (who suggested an extensive bibliography to meet the lack of Italian references on the topic) and the other one was Umberto Eco.

Professor Eco was interviewed for a project called “Wiki@Home”, supported by the Italian chapter of the Wikimedia Foundation⁴, Wikimedia Italia⁵. Wikimedia Italia (WMI) is an association for open culture and open knowledge, and it supports Wikimedia projects as Wikipedia and Wikisource. Wiki@Home is a subproject of WMI which is aimed to interview important members of the cultural and entertainment world. In this context, the researcher contacted and gained an appointment with professor Eco to interview him about Internet, collaboration and Wikipedia. The researcher thus exploited this unique occasion to utilize professor Eco as a key informant for the study, investigating topics as collaboration between scholars, especially in the Italian context. Interview revealed to be extremely rich and helpful to gain in-depth information about Humanities in the Italian area and their approach to Internet and collaboration. The transcript of the interview is available (in Italian) under **Creative Commons-Attribution-Share Alike** license at the URL: http://it.wikinews.org/wiki/Intervista_a_Umberto_Eco

The English translation will be available in **Appendix 1**.

3.6.3 Interviewees

Interviewees were 5 members of the Digital Humanities Italian community, each with a different specialization (greek philology, modern literature, digital libraries, even history of logic), but all involved in digital humanities and digital libraries. They are working in academic and research institutions in Italy, UK and US.

The first interviewee was contacted after an overview of the literature and a conversation with the advisor: the others were too selected from the literature and with the method of *snowball sampling*. At the end of each interview, scholars were requested to suggest other people to be interviewed, and often participants converged proposing the same names. People who were named by more than one participant were contacted: every name has been discussed with the thesis advisor.

⁴<http://wikimediafoundation.org/wiki/Home>

⁵http://www.wikimedia.it/index.php/Who_we_are

Participants have been labeled by numbers, from I1 to I5. From interviewee I1 (who served also as a pilot for interviews), I6 and I3 were individuated. The researcher and the advisor agreed to interview I2, who then suggested also I3 and I4.

After I5, the information gathered started to be redundant, which according to Pickard (2007) is a condition for exit fieldwork; thus, data collection stopped, with the consent of the advisor.

All but one the participants were Italian: therefore, all but one of the interviews was conducted in Italian.

Italian audience had been privileged for several reasons:

- the core interest of the research was on collaborative digital libraries for the Humanities Italian community
- speaking Italian allowed both the researcher and participants to be more comfortable during interviews and helped in-depth data analysis

The researcher spent great effort in translating correctly interviews, not to misspell or provide a wrong interpretation of interviewees' words. The transcription of Eco's interview provided in [Appendix 1](#) has been translated collaboratively among members of Wikimedia Italy: one of them is actually a professional translator and guaranteed a correct and formal approach for the translation.

Although the research question was focused on the Humanities community in the Italian context, sampling did not choose only Italian philologists for purpose: in fact, context suggested that only people previously involved in digital humanities and digital libraries could understand and have motivated opinions about a specific and in-depth question as the research questions of the present studies. Sometime, they did work in related but different communities (i.e. the Greek philology community). However, difference of language does not seem an obstacle for transferability of both the structure and even the outcomes of the study. A crucial aspect of the present case study is obviously environmental culture of the community, and everyone but one of the interviewees are Italian. This choice was made to explore the Italian environment of humanities but not being subjected only to a single community of practice.

3.7 Data collection techniques

To explore the digital humanities community views, opinions and feelings about collaboratories digital libraries, the main data collection technique used were *semi-structured interviews*.

The research question and objectives demanded an in-depth investigation in participants' opinions and tacit knowledge and experience about collaboration. As the topic touched the entire community culture and psychology, the data collection technique needed to explore deeply and broadly participants thoughts. As people, when asked a question, do not want to write much to carefully explain their insights and thoughts, semi structure, spoken interviews seemed the best technique to collect sensitive data.

This technique resulted to be the correct one: data gathered were rich and opened different questions and paths to be explored.

Interviews were semi-structured, with few open-ended and broad questions, in order to keep the conversation between the researcher and the interviewee consistent but open to new directions and topics related to the research question. Interaction revealed to be fundamental: every interview informed the following, and every conversation were gained experience from the past one.

Interviews were conducted online, with the VOIP software Skype⁶. They were synchronous and spoken, but computer-mediated and with no webcam (practically, they were *telephonic interviews*). The procedure has been chosen for several reasons:

- interviewees live in different cities in Italy and therefore it would have been really time and money consuming traveling to meet all of them.
- most of interviewees had high computer skills and interviewing online did not constitute a problem.
- doing *telephonic interviews* with the computer allows taking notes directly on the personal wiki and real-time interaction with interviewees, browsing websites and sending links or files. Furthermore, it allowed automatic recording of the speeches and chat logs.

⁶<http://www.skype.com>

There were some exceptions where interviews had been conducted live or by e-mail.

3.7.1 Interviews

Conversation is a basic mode of human interaction. Human beings talk to each other... Through conversations we get to know other people, get to learn about their experiences, feelings, and hopes and the world they live in. (Kvale, 1996, p. 23)

Interviews are frequently used data collection techniques in information research (Pickard, 2007). They are applied to “access what was in, and on, the interviewee’s mind” (Stenhouse, 1984).

As the present research is focused in exploring researchers and users opinions about possibilities and potentialities of a *collaboratory digital library*, interviews allows them to express complex and articulated thoughts. Lincoln and Guba (1985) suggest that interview permit respondents to go back and forth in time, interpreting past and present and predicting the future: the explicit request of foreseeing developments in digital libraries and scholarship collaboration made interviews perfectly fitting data collection techniques.

Furthermore, “qualitative, descriptive, in-depth data [...] specific to the individual” (Pickard, 2007, p. 172) were requested by the purposes of the research.

Kvale (1996) lays out seven stages of the interview process, (*thematizing, designing, interviewing, transcribing, analyzing, verifying, reporting*) but Pickard (2007) argues process is often not as linear as suggested. She hints to use the stages as a broad outline of the process, but being “prepared to iteration” (p. 173). Moreover, interviewing as a whole is a practice that is *learned by doing* and often beginners start being confident with their personal procedure after several interviews.

Thematizing

Is necessary to be very clear about the purpose of the research and the topic that is being investigated. Themes and questions are needed to be

appropriate for each interviewee, because it is likely to ask different people different things.

It is suggested to structure themes of the interview in a natural order to allow “interviewee to follow a logical thought process and allow[ing] the interviewer to gain a growing understanding of feelings, behavior and beliefs” (p. 173).

Designing

The purpose of using an interview guide is to “ensure that each interview covers basically the same ground but gives the interviewer considerable discretion in the conduct of the interview” (Ellis, 1993, p. 475). The guide controls the interview and its dependent to the chosen type of interview: it can range from a strict script (actually similar to a questionnaire) to a general list of topics to be covered during the conversation.

Interviews accomplished were more of the second type: they were semi-structured, and only few questions and topics (i.e the research questions) were necessary and did recur in each conversation. The guide of the interview was the following:

- self presentation of the interviewee, explaining his academic background and interests
- the research questions
 1. **H2**: Can a digital library for the Italian studies community be *collaborative*?
 2. **H2**: Can a digital library for the Italian studies community be a *collaboratory*?
 3. **H3**: Can a digital library for the Italian studies community be built upon a *wiki*?
- an opinion about facilitators and barriers to collaboration in humanities and digital humanities
- an opinion about features and services of possible collaborative frameworks for humanists

Interviewing

Interviewing is heavily dependent on the rapport between interviewer and interviewee: conversation has to be as much as relaxing and comfortable as possible, and is responsibility of the interviewer to make it happen. Furthermore, interviewer needs to be prepared on topics covered by the interview and react properly to answers: questions has to be asked and answer can lead to unforeseen but meaningful paths.

Role of the interviewer is “to *listen, reflect and respond*” (Pickard, 2007, p. 177).

Due to the non-existence, in the Italian context, of collaborative and collaborative digital libraries, it was sometime difficult to provide interviewees with practical examples of features and characteristics of such projects. In those cases, the interviewer had to discuss hypothetically and with scenarios on what kinds of services, functionalities and features a collaborative digital library may offer. Discussions were therefore theoretical and focused on potentialities, barriers and facilitators.

The consent of interviewees was asked in order to voice record the responses and discussions arisen out of the interview.

Audio recording revealed fundamentally important to keeping all the data and listen interviews several times: furthermore, it provided security and allowed interviewer to be totally engaged in conversation not taking care of writing every important quote or note. Interviews were audio recorded with Skype Call Recorder⁷, a free/open source software for Linux that automatically records Skype conversations and save them in different formats. Audio recordings were mp3 files stored in different folders, divided by authors.

Transcribing

Transcription is suggested to be done as soon as possible after the interview. Qualitative research needs engaging with interviews as their done, to look for emerging patterns and important data. Transcribing informs researcher and provides insights of the interviewing process, making him aware of possible

⁷<http://atdot.ch/scr>

flaws of his procedure. Transcribing were therefore done commonly the second day after the interview, because most of the interviews were recorded in the evening.

Analysis

Analysis “is a constant, ongoing element of the research process” (p. 178). It will be described later on.

Verifying

Kvale (1996) intend the verification as the stage to understand if the interview covered the extent of the research question. Pickard (2007, p. 178) adds that verification is “a form of *member check*”, a concept introduced by Lincoln and Guba (1985). Returning to the interviewees after transcription and analysis can confirm or deny if the interviewer had understood what they really meant, and can also be an opportunity to harvest new data. Websites and other literature were used to verify some of the issues that were mentioned in the process of the interview; mutually, projects and articles suggested in interviews were explored for in-depth understanding of the interviews themselves. The process also helped to verify the names of initiatives and projects.

Reporting

Evidence from interviews must form the foundation of the new researcher’s emerging theory, and this must be present in the final report. Pickard (2007) highlights that “spoken word is evidence” (p. 179), hence important verbatim quotes important were reported.

A pilot, unstructured interview has been conducted with the first key informant, who is also a friend of the researcher. The interview has been useful also to get acquainted with the interview process and ways of exploring the defined topics. After the pilot and a discussion with the thesis advisor, a set of open-ended questions had been defined to be the guide of further interviews.

3.8 Computer as a research instrument

In interpretivist research, the main research instrument is the human himself, who explores, analysis and interpret a phenomenon.

However, it has previously stated that the *computer* has been used as a research instrument too. The statement wanted to emphasize the used of softwares as crucial tools of the writing and the design of the present study.

Several programs has been used for developing this work:

- a wiki software (**MediaWiki**)
- a typesetting software (**L^AT_EX**)
- a concept map software (**Cmap**)
- a VoIP software (**Skype**)
- a recording software (**Skype Call Recorder**)
- an automatic translator (**Google Translate**)
- an English thesaurus (**OpenOffice.org**)

Developing the technical framework has been very natural for the researcher, who is used to work with different softwares and has an expert knowledge in computers. Yet, a good amount of time has been invested in selecting and developing tools: this however is not seen as a wasting of time, but as a good opportunity to gain deeper comprehension of useful tools for writing and research and, moreover, an experiment to confirm the idea of *laboratory* proposed by the present study.

It is in fact the opinion of the researcher that a good technological framework is a fantastic boost for innovation, especially when the framework is collaborative and *collaboratively improvable*.

Eventually, the whole framework was as valuable aid the researcher: the developing was both time-consuming (sometimes) and time-saving (often), but it was definitely worth the effort. Furthermore, experience gained will surely help the researcher in future studies.

Softwares and their uses for the scope of the present study will be explained in the following sections.

3.8.1 L^AT_EX

L^AT_EX is a document markup language used for typesetting. Although quite complex, is widely used in hard sciences and academia for scholarship communication, due to the huge flexibility and publishing feature. LaTeX documents can achieve great level of complexity, containing tables, images, mathematical and scientific formulas. A fundamental feature of LaTeX is to apply a WYSIWYM (*What You See Is What You Mean*) approach instead of the more common WYSIWYG (*What You See Is What You Get*) approach, used by normal word processors. This allows a much more comprehension and control over document structure and layout. Latex technically *compiles* the source text as a *code*, automatically taking care of a bug part of the structure, layout, reference, settings. Moreover, while LaTeX at the beginning is complex to understand and time-wasting, eventually it reveals extremely time-saving for organizing the text and the layout completely in automatic. All chapters, sections, figures and quotes are automatically counted and thus indexes are automatically generated, with the precise chapter and page. No time is wasted to generate table of contents or list of figures and tables. Very little time too is needed for generating bibliographies. Being a document markup language, LaTeX is highly customizable, providing hundreds of packages for different needs.

In LaTeX, it is possible to organize and compute directly the code of the document. With a reference management system called BibTeX, it is possible to directly *compute* all quotations and references of the bibliography. Therefore bibliographies are not written by the author, but by the program itself who reads the tags put by the author and correlates them with proper metadata stored in an additional file also created by the author. This allows to maintain distinct the database of metadata related to the bibliography and the paper that is written, which will harvest the data when needed. An example will be provided. The metadata of a book, as

```
@book{pickard2007,  
  title = {Research methods in information},  
  author = {Pickard, A.J.},  
  year = {2007},
```

```
address = {London},  
publisher = {Facet Publishing},  
}
```

are contained in a file called `Methodology.bib`. Then, in the *Methodology* chapter that book can be cited in various ways, recalling author and year, with or without parenthesis, etc. For example, with the two codes:

```
\cite{pickard2007}  
\citeA{pickard2007}
```

they will respectively lay out like this:

(Pickard, 2007) Pickard (2007)

Thus, the bibliography of the present study is available as a separate file, fully interoperable with normal standards and software using BibTeX. Needless to say, the work is reusable for further documents, with no other effort needed.

For this document a package as `apacite` is being used that follows the APA bibliographic style⁸. This means also that the `Bibliography` section is following APA guidelines. More styles are available, supported by related packages.

Moreover, the researcher found extremely helpful and time-saving that scholar search engines like Google Scholar⁹ or CiteSeerX¹⁰ could export bibliographic data directly in the BibTeX format.

Thus, LaTeX allows great flexibility in bibliography styles and formatting.

⁸Actually, the `apacite` module is brand new, developed at the end of 2009, and there are still some issues and bugs to fix. The researcher struggled to present a layout perfectly following APA guidelines, but sometime exceptions could not be avoided. It is hoped that errors would not be too evident and may be forgiven for the not direct responsibility of the writer. What is more, some packages do conflict one with each other for obscure bugs that the researcher cannot explain nor solve. One of them is the fact that the package for *hyphenation* sometimes do not hyphen properly some words. Unfortunately, the researcher could not fix this kind of errors.

⁹<http://scholar.google.com>

¹⁰<http://citeseerx.ist.psu.edu/>

The researcher chose to utilize this particular program for several reasons: even though it is complex and not well-suited for neophyte, LaTeX is an extremely powerful program. What is more, the researcher felt as really important to utilize a software used mostly in scientific studies.

In fact, as the thesis is focused on collaboration in Humanities, it is though important to stress and emphasize a more interdisciplinary approach for both Humanities and hard sciences. Collaboration is need not only within disciplinary groups, but also and above all *between* disciplinary groups. It is a strong belief of the researcher that interdisciplinarity is the main road for all sciences: therefore, an useful tool as LaTeX should be known and used also outside the boundaries of hard sciences.

3.8.2 Wiki

A wiki software has been used as a research instrument for several reasons. The researcher had it installed months before the study began, therefore it did not constitute a further complication for the research process. The wiki had been as a flexible framework, acting as a

- case database
- data analysis framework
- word-processing software

MediaWiki¹¹ is the wiki software engine utilized for the present research. It is a web-based wiki engine used by Wikipedia and all the Wikimedia projects, and it is highly stable, scalable and flexible, supported by a very active community of developers (Koblas, 2006). Being *free software*¹², it supports hundreds of extensions and plug-ins written by his community of users. MediaWiki is written in the PHP programming language, and the version utilized for the study applied MySQL as relational database management system.

The software provides a special mark-up language (a very common feature in wikis) that simplifies HTML language: it has been used mainly for

¹¹<http://www.mediawiki.org>

¹²Distributed under the terms of the GNU General Public License.

formatting text and creating links to other wiki pages. Also, MediaWiki supports rich content generated through specialized syntax: for example, the software comes with support for rendering mathematical formulas using \LaTeX , and extensions that allows exporting of normal wiki pages in \LaTeX or directly printed in PDF.

Templates as editable tools

An crucial feature of MediaWiki is the possibility of creating *templates*.

Templates are text blocks that can be dynamically loaded inside another page whenever that page is requested. The template is a special link in double curly brackets (for example “Disputed—date=October 2008”), which calls the template [...] to load in place of the template. Templates support parameters, so that parts of the text can be substituted for each specific use case. (Wikipedia, The Free Encyclopedia, 2010)

Templates are used for every kind of activity, from creating boxes to standardize information to automatically categorize pages. They can serve many different purposes at once, just depending upon the complexity of the code and the ability of the coder. Furthermore, wikis are often free licensed and open, allowing users and readers to access the code of every page: templates are therefore open to modification and open to be studied from their source code.

The wiki served different purposes and was utilized in different manners during the stages of the research process.

Literature review

Most part of the literature reviewd was digital, so several folders were created to organize the articles to review. The Okular software¹³ was used to read and annotate PDFs, meanwhile one page for each article reviewed was created in the wiki by the researcher. In each of those pages important

¹³<http://okular.kde.org/>

quotations from articles were copied and pasted and personal notes were written.

Furthermore, specific *templates* for *tagging* articles were created: every time a sensitive topic or a keyword was mentioned, a related tag was created by the researcher. In this way, a *controlled vocabulary* of important terms and keywords *emerged* from literature; each of those keywords created a *category* in the wiki that included all the articles (and interviews) that mentioned the particular keyword.

For example, if the word `collaboratories` was mentioned within an article, the researcher marked up the word with the code

```
{{Tag|WORD}}
```

in this way:

```
{{Tag|Collaboratories}}
```

This automatically created the category in the wiki called

```
[[Category:Collaboratories]]
```

All the articles marked up with the template `{{Tag|Collaboratories}}` were stored automatically in the category `[[Category:Collaboratories]]`. This happened for all the keywords the researchers decided to mark up.

Therefore, the emerging list of categories was a sort of *controlled vocabulary*: each important topic in the literature was chosen to be a TAG (thus, a category), and sometimes synonyms were conducted to the same category. Although the system was far from being a proper controlled vocabulary, compiling the list during the literature review helped the researcher finding the core of fundamental keywords.

Moreover, the list revealed to be *weighted*: as the categories listed showed automatically how many articles were contained inside, a quick glimpse of the category/keyword list could offer an ostensible but helpful estimation of the knowledge base.

In fact, the vocabulary was used to develop the *theoretical and conceptual framework* (Pickard, 2007), that has been represented by a *concept map* (fig.1). Eventually, both the concept map and the vocabulary were used

as a framework for the *critical review*. The list of 74 categories is available in **Appendix 2**.

The list of TAGs is far from being complete and coherent, because neither all articles were submitted to the same procedure nor the researcher had been totally coherent in assigning TAGs. Nevertheless, the procedure proved itself to be engaging and helpful, forcing the reviewer to develop a comprehensive list of topics and keywords to define the domain of the study directly taking topics from literature. Far from being a quantitative and exact approach, this procedure could still be considered partly *bottom up*, (at least in the phase of pulling out meaningful words from articles), and could be developed further with proper resources of time and expertise in research methods and programming.

In fact, the procedure described above can be accepted as experimental, as an attempt of the researcher to both show flexibility of wikis and engage in a negotiation with the software and methods for data analysis.

Interviews

Interviews were semi-structured and recorded via Skype, with no webcam used. They can be so regarded as telephonic interviews, with the additional feature that, being online, both the interviewer and the interviewees could browse the Web and interact, suggest websites, visit mentioned projects, check information. The interviews were transcribed directly, and as soon as possible, in the wiki and consequently annotated and marked up for data analysis.

A procedure and the necessary technical tools were developed for *transcluding* important claims: when in the transcriptions were found *facilitators* and *barriers* regarding the research question, they were highlighted with different colors and marked up with a particular code. The code, a dedicated template, *transcluded* automatically the claims into other specific pages, and directly collected them in categories.

3.9 Data analysis

For data analysis, “constant comparative analysis” has been applied. Strauss (1987) originally developed it for use in the grounded theory methodology of Glaser & Strauss (1967). The strategy involves examining data while comparing the with all the similar or different data gathered during the fieldwork. This strategy is very common and it must come out from (and be grounded in) raw data, in a *bottom up* approach: categories have to be inductive and emerge directly from the data, and not be established a priori, although it is inevitable that prior research will influence choice of important issues. Melia (1997, p. 31) states:

The original version of grounded theory stressed the idea that theory emerged from, and was grounded in, data. Careful analysis of data items using the constant comparative method would lead to the emergence of conceptual categories that would describe and explain the phenomenon under study.

Nevertheless, Strauss did revise his theory to still drive the coding from data but not also have a more structured approach. Strauss and Corbin (1998) suggest to divide the coding in three series of activities:

- open coding
- axial coding
- selective coding

Using a wiki as a research instrument revealed extremely helpful for memo writing and interacting with data. The wiki acted a framework where the researcher could construct and develop proper tools which helped him in the analysis, as the before mentioned TAGs and template for highlight and transclusion of meaningful text.

3.9.1 Open coding

Open coding is defined as “the analytic process through which concepts are identified and their properties and dimensions are discovered in data

”Strauss and Corbin (1998, 101). It is an initial phase where the researcher examine the data searching for difference and similarities of single parts of the data themselves. Researcher had to identify discrete concepts, which are the “basic units of analysis of the emerging theory” (Strauss & Corbin, 1998). Pickard (2007) suggests to create a category for each concept discovered: the category can then be described with a set of properties and dimension to add clarity. Developing the list of categories will lead to complete the basis of the emerging theory.

Using interviews as data collection technique, the data provided were transcriptions of in-depth conversations. The researcher executed this phase of data analysis with the help of the wiki system. Using a similar approach to the process of tagging content described above, the researcher created ad hoc templates in the wiki to highlight sentences or little paragraphs with different color. Color chosen were blue or red: blue indicated elements positive for the research questions, red indicated negative elements. For example, facilitators to collaboration were colored in blue, while barrier to collaboration were colored in red. All facilitators and barriers to were *transcluded* with the same wiki template in a dedicated page, which listed all facilitators and barriers to collaboration.

Although, due to scarcity of time, the researcher could not dedicate plenty of time to development of tools in the wiki regarding data analysis, it is important to state that open coding is a phase of data analysis that could gain much help from a digital system like the wiki proposed.

In fact, Pickard (2007, p. 243) describe the process of creating and organizing categories as follows:

I ask my students to imagine they have a row of shoe boxes. On the side of each shoebox they write the name and definition (discrete properties and dimensions) of a category; they create each new shoe-box in response to an actual item of data, not from an abstract concept. That means that every shoe-box contains at least one item of data; each time a new item of data is located in the evidence it is cut out and dropped in that box or photocopied if it can fit into more than one box.

Within a proper digital environment, the process describe above could be applied with much less effort (and waste) than cutting documents and shoe-boxes. In the wiki system utilized by the researcher, all data were reduced to a textual form and could immediately be categorized in few clicks. Also single parts of the transcriptions could be organized with *ad hoc* templates (as the ones developed by the researcher), and much more could be done with proper time and computer skills. Simple action like copy, cut and paste could lead to extremely sophisticated organization of data, within a framework complex enough to allow hyperlink, categories and above all flexibility in customization and development of personalized tools. The wiki system proposed by the author is just a single instance, but it would like to suggest that a flexible approach could reveal extermely helpful for data analysis and the whole research process.

3.9.2 Axial coding

Moving from general categories of concepts to related sub-categories is moving from open coding to axial coding. Handling the categories themselves is a more refined task whose aim is the “identification of the conditions that give rise to a particular phenomenon and the context it occurs” (Pickard, 2007, p. 243). This type of coding is conducted in tandem with data collection: in fact, the researcher started to analyze the data meanwhile he was interviewing. Each interview was informed by the previous one, and the author often asked the interviewee for confirmation of hypothesis and relations between concept raised in the analysis. The phase of axial coding is to searching for links between categories, and checking the content of the category itself.

3.9.3 Selective coding

Selective coding is “the process of integrating and refining a theory” (Strauss & Corbin, 1998, p. 143). It is the final stage of data analysis, where theory has reached a saturation and no new connections, properties and relations are emerging from analysis. *Conceptual framework* is the outcome of this phase. In this phase grounded theory is demanded to provide a section

for recommendations and suggestions of further research. The result of the selective coding can be read in the concept map developed to illustrate the conceptual framework, in **Chapter 4**.

References

- Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, *63*(4), 469–486.
- Erlandson, D. (1993). *Doing naturalistic inquiry: A guide to methods*. Sage Publications, Inc.
- Flick, U. (2002). Qualitative research-state of the art. *Social Science Information*, *41*(1), 5.
- Ford, J. (1975). *Paradigms and fairy tales: an introduction to the science of meanings*. London.
- Greene, J. (1990). Three views on the nature and role of knowledge in social science. *The paradigm dialog*, 227–245.
- Heisenberg, W. (1958). *Physics and philosophy*. New York: Harper.
- Koblas, J. (2006). *Wikis: Tools for information work and collaboration*. Cambridge: Chandos Publishing.
- Kumar, R. (1999). *Research methodology: A step-by-step guide for beginners*. London: Sage Publications.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. London: Sage Publications.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. London: Sage Publications.
- Maykut, P., Morehouse, R., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. Routledge.
- Melia, K. (1997). Producing plausible stories: interviewing student nurses. *Context and method in qualitative research*, 26–36.
- Patton, M. (2002). *Qualitative research and evaluation methods*. London.
- Pickard, A. (2007). *Research methods in information*. London: Facet Publishing.
- Stake, R. (1995). *The art of case study research*. London.

- Stenhouse, L. (1984). Library access, library use and user education in academic sixth forms: An autobiographical account. *The research process in educational settings: Ten case studies*, 211.
- Strauss, A. (1987). *Qualitative analysis for social scientists*. Cambridge.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London: Sage Publications.
- Warden, B., & Wong, S. (2007). *Introduction to qualitative analysis*. (Retrieved 15:56, June 24, 2010 from <http://www.eval.org/SummerInstitute07/Handouts/si07.wongF.pdf>)
- Wenger, E. (2006). *Communities of practice: a brief introduction*. (Retrieved 17:03, June 5, 2010 from <http://www.ewenger.com/theory/>)
- Wenger, E., & Snyder, W. (2000). Communities of practice: The organizational frontier. *Harvard business review*, 78(1), 139–146.
- Wikipedia, The Free Encyclopedia. (2010, April 16). *MediaWiki*. (Retrieved 10:33, April 17, 2010 from <http://en.wikipedia.org/w/index.php?title=MediaWiki&oldid=356386203>)
- Yin, R. (2002). *Case study research: Design and methods*. London: Sage Publications.

Chapter 4

Analysis and findings

4.1 Introduction

Data analysis was conducted through “constant comparative analysis” and led to the categories here going to be presented. Interviews revealed themselves to be extremely interesting but challenging, providing in-depth conversations sometimes difficult to analyze.

The wiki used by the researcher revealed itself as a helpful instrument for analysis, because it allowed the researcher to follow exactly constructivist procedures suggested in research methods literature: particularly, memo writing and open coding were helped by the possibility to interact in many ways with transcripts of interviews, parsing texts and extracting data. Wiki templates have been used during the whole research process to highlight in transcripts of interviews meaningful sentences and data. Phrases were then colored in different colors to distinguish, for example, facilitators and barriers to collaboration in digital libraries.

During the whole research process, single terms and keywords of the research questions have been questioned and debated with interviewees. Actually, a considerable amount of time of the interviews has been used for discussion of crucial terms like *wiki*, *collaboratory*, *collaboration*, *digital library*. These conversations happened to be very helpful and led to important insights and results. Key informants helped defining boundaries of the research: Alison Babeu provided a dedicated bibliography, while the long conversation with Umberto Eco enlightened deeply the Humanities context, especially the Italian one.

Participants of the study contributed greatly to the study with their intelligence and experience arguing about concepts and terms used in the research questions and objectives.

Therefore, these results will be presented as they were found: as the outcome of a *conversation*. Every important topic discussed in those conversations led to a result that will constitute a paragraph or a section. The present chapter will also explore in-depth causes and will highlight keywords, issues and results raised by the research.

4.2 Different types of Digital Libraries

The first discussion during interviews regarded the term “digital library”. This actually confirmed a well-known result: the fact that ideas, definitions and practices around digital libraries are far from being homogeneous. Literature suggested, and interviews confirmed, that the term *digital library* is used for indicating projects (or, from a complementary point of view, *functionalities*) very that are different one for each other. They go from the most common tasks of preservation, storage to most advanced tools for philology and textual analysis. What’s new is that data gathered both with literature analysis and interviews suggest a convergence and an increasing attention to collaboration.

Nonetheless, interviewees explicated a vision of digital libraries that have been coded in Table 3. During the conversations, it came out clearly that participants did want to clarify what they intend with the term “digital library”, and clarification lead to assert the existence of two main models of DLs.

Interviewee I4 declared:

Historical projects (regarding DLs) has been of these two kinds:

- *repositories* in which to store digitized texts, in various ambit: some academic, some other amateur (like the Gutenberg Project¹ or Liber Liber²), some other institutional. Often [the repositories were built] without thinking at the aims they could accomplish.
- the *scientific-academic digital library*, which tries to provide users with tools for working on texts. Often [these tools consists] in linguistic analysis tools, which are quantitative, like concordances, frequencies, etc. Sometimes, user can search on lemmatized texts, for example in the Perseus project.

¹<http://www.gutenberg.org>

²<http://www.liberliber.it>

Other interviews upheld the same idea. It is important to notice that the subdivision here provided must be intended as an abstract and incomplete portrait: several projects do not fall exactly in one category, and sometimes things are more complex than suggested. Thus, it is the opinion of the researcher that such discrimination arisen from interviews can be useful, intending the two main kinds of digital libraries as two limits of a continuous range.

Table 4.1 will sum up some features of types of digital library involved, gathered from interviews. Following paragraphs will explain the division.

Kind of DLs excluded from the table

From Table 3, other minor types of DLs suggested by the interviewees have been excluded. The main reason is that there is practically no literature studying these new types of libraries. As very innovative projects, confusion is still high and there is no systematic study about them. Only one interviewee, I4, suggested that there were different directions for future digital libraries, and the researcher preferred to present only models of DLs widely used and somehow stable.

Interviewee I4, who has a broad vision of DLs had competences in different Humanities disciplines, sorted the different kinds of DLs by audience. He highlighted clearly that, aside repositories and research projects, the ideas around DLs are still focused on the traditional library:

Finally, there is the ultimate audience, broad and undifferentiated, which has been less explored. Also big projects like Gallica³ or American Memory⁴, strongly mimic the model and the structure of a traditional library, with catalogs and sometimes thematic routes. That are exactly the same kind of expositions that we often find in our libraries.

Also new projects like Europeana⁵ (that is a sort of meta-aggregator of digital libraries and online museums) and the World Digital Library⁶ still

³<http://gallica.bnf.fr/>

⁴<http://memory.loc.gov/ammem/index.html>

⁵<http://www.europeana.eu>

⁶<http://www.wdl.org/en/>

Table 4.1: Models of DL arisen from interviews.

| | Repositories | VRE |
|----------------------|---------------------|------------|
| User | Generic user | Scholars |
| Context | Amateur/Scientific | Scientific |
| Area | Librarianship | Philology |
| Focus | Collection | Text |
| Text | Static | Dynamic |
| Granularity | Low | High |
| Collaboration | Mixed | Mixed |
| Technology | Simple | Advanced |
| Provider | Data | Services |

mimic the traditional model of paper libraries and do not explore deeply the potentialities of the digital world.

Furthermore, interviewee I4 emphasizes the lack of DLs created explicitly for teachers and didactics:

[The projects] which has been less explored are projects dedicated to a non-specialized audience, or at least an audience who does not use texts for linguistic analysis, but [an audience that] could use them for literature didactics, or creation of supporting tools, etc.

4.2.1 Digital library as a *repository*

This kind of digital libraries is the most common and directly stem from the original model of traditional libraries. As paper libraries do, these digital libraries provide access, storage and preservation of the content. They are often run by librarians, and have often a great attention to metadata.

Context of repositories can be either academic (i.e. in Open Access institutional repositories) or amateur (i.e. Liber Liber, Gutenberg project, Wikisource⁷). Inside these areas, projects can have different complexity.

Often, volunteer projects as Liber Liber or Gutenberg are sustained by a community of volunteers which digitizes and proofreads public domain

⁷<http://wikisource.org>

books. Similar but institutional projects share the same aims and model; differences often arise in scale, scientificity, work-flow, funding, but not in the model of the digital library.

On the other hand, some digital libraries do not have *digitization* as an objective: working with born-digital documents, they are focused on preservation, storage and access of documents (sometimes multimedia content is allowed too). This is the case for example of Open Access institutional repositories.

The digital library repository's focus is more the whole collection of texts than a single one; a text is intended as static objects which need to be accessed, retrieved, and preserved. In these projects granularity is therefore low, because they try to aggregate texts and books and do not aim to work upon them. This kind of digital library often provides no or few services for working with or around texts, and thus can be labeled as mere *data providers*.

Collaboration too is mixed: in amateur projects is often high, because digitization work is done collectively, while in open access institutional repositories is low, because students and scholars just need to upload their articles and fill the related metadata.

Technically speaking, systems for repositories are often not too advanced. This is especially true for volunteer projects: in projects as Gutenberg and Liber Liber, mailing lists are used for communication and databases are used for storage. On the other hand, Wikisource offers an integrated wiki system which allows both access and a framework for digitization and proofreading of texts. Interesting enough, Liber Liber will move to the same wiki engine, MediaWiki⁸.

Indeed, professional digital repository systems as DSpace⁹, Fedora¹⁰, Greenstone¹¹ and Invenio¹² are designed directly for storage, with great attention to metadata and interoperability, allowing harvesting through the OAI-PMH protocol. Creating an useful and efficient infrastructure for shar-

⁸MediaWiki (<http://www.mediawiki.org>) is also the wiki engine used by Wikipedia.

⁹<http://www.dspace.org>

¹⁰<http://www.fedora-commons.org/>

¹¹<http://www.greenstone.org>

¹²<http://cdsware.cern.ch/invenio/index.html>

ing and access of digital objects, they lack in user interaction and social tools, being often too static and rigid. Moreover, these projects rarely explore Semantic Web tools and practices.

4.2.2 Digital library as a *Virtual Research Environment*

This model of DLs is the one mostly used by Digital Humanities. These digital projects are actually more virtual working spaces than simple digital libraries: they are digital environments where scholars and researchers can work with, around and upon texts. Though less traditional, these digital libraries have their roots in the first ideas of Bush (1945) and Licklider (1965), namely the *memex* and the *procog system*. What is more, these new projects add a strong centering in collaboration, aggregation and networking where old visions were more focused on individual knowledge or “intellect augmentation” (Engelbart, 1988).

The focal point of these projects is more on texts themselves rather than collection; therefore, they are more concentrated on providing *services* and functionalities than simple static digital objects. They offer tools for working with texts, that vary from linguistic analysis tools from annotations. These kind of digital libraries often do not call themselves “digital libraries”, but prefer the term *electronic editions*.

Interviewee I5 stated:

One thing is the *digital library*, which belongs to the librarian area; another is the *electronic edition*, which belongs to philology. Obviously there is some overlapping: on one side electronic editions are growing and developing, offering more features. On the other hand digital libraries are becoming more granular.

Therefore, in the scholar context this kind of digital libraries is more related to philology and ecdotic than librarianship.

In these DLs, texts are dynamic, not fixed and static objects that cannot be manipulated. Granularity of texts is often really high, and different technologies are often involved to provide complex tools for study and analysis. For example, mark-up languages as TEI or XML are used to code texts, and therefore to provide services for text and data mining. Statistical tools are

used for quantitative analysis, as well as syntactical treebanks, as recalled by interviewee I1:

[...] Is principally concerned with *treebanks*, databases of syntactic structures, represented by tree structures. The project Alpheios¹³ has created a collaborative platform for these tools: you go on a text and you can *annotate a syntactic structure*. They want to go further and expand [this model] with *multiple annotators*, plus also *different versions of the same text*.

Semantic Web

Moreover, Semantic Web technologies are receiving great attention from developers of VRE. Though sometime a controversial term¹⁴, Semantic Web has been firstly introduced by Berners-Lee, Hendler, and Lassila (2001), and it's now widely used to indicate a cutting-edge field of computer science research and technologies.

It mainly describes methods and technologies to allow machines to understand the meaning of information on the Web, with a particular attention about the the availability of *machine-readable* metadata that would enable automated software agents to access the Web more intelligently. The agents would be able to perform tasks automatically and locate related information on behalf of the user.

Moreover, while the term "Semantic Web" is not very defined, it's often used to describe the model and technologies proposed by the W3C (*World Wide Web Consortium*). These technologies include RDF (*Resource Description Framework*) and OWL (*Web Ontology Language*), which are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain (Wikipedia, 2010a).

Many of the technologies involved have their roots in decades of Artificial Intelligence research, and there are many connection with the disciplines on

¹³<http://alpheios.net>

¹⁴There are many critics around Semantic Web, regarding different aspects: from the practical feasibility to to privacy and censorship issues. For a deeper understanding, see (Marshall & Shipman, 2003).

Librarianship and Information Science (i.e. subject classification, categorization, tagging). Development of authority files, controlled vocabularies, thesauri, and ontologies are a remarkable part of the work of researchers in VREs and sectoral digital libraries. These tools are developed within very different disciplines for very different aims: for example, at CERN Library (in collaboration with the Computer Science Department) they are constructing a controlled vocabulary for a brand-new version of the SPIRES document search engine, which will cover completely the domain of *High Energy Physics* and retrieve documents, metadata and every kind of information. At the same time, various philological projects allow scholars to annotate and develop ontologies upon texts.

Semantic Web technologies are then very careful to standards, formats and interoperability, as well as the possibility of releasing data in raw (and open) form for further reuse from third parts. The whole concept of open data and reuse is a new trend highly supported by Tim Berners-Lee himself, and is gaining much attention by governments and institutions. In fact, if an institution is publicly funded, his data should be public too in an open and reusable format: this is way several institutions, from the UK Government to the CERN Library, are gradually releasing their data.

4.2.3 Focus on VREs

Eventually, digital library is a label that cover very different projects. What is more, some interviewees suggest other kinds of digital libraries (neither traditional nor academic/scientific) for different audiences. For example, there could be digital libraries for general users which provide maps of topics and places in literature, or learning objects for teachers and professors.

Furthermore, it is important to state that data gathered, observation and literature analysis suggests that there is some convergence of these two kinds of digital libraries. For example, cutting-edge institutional repositories are becoming more granular and adding more services (i.e. annotation, comments and social tools in CERN's Invenio digital library system), while electronic edition projects are fostering collaboration between members and moving towards to becoming Virtual Research Environments.

The present study focused mainly on the this latter kind of digital libraries: projects build for scholars as an aid and a framework for research. The next section will enlighten in details the topic of collaboration in these environments.

4.3 Collaboration

Literature and interviews both confirmed the increasing importance of collaboration in digital libraries. As e-science and cyberinfrastructure are growing, Humanities are exploring the same approach to a strongly collaborative work in digital environments.

Interviewee I4 is rather clear, stating:

The topic of collaboration is increasing every day; in electronic editions, it is the *keyword*.

Interview I1 too was asked if collaborative frameworks could help the field of Humanities:

Absolutely. I believe this is the only way for a science which needs to innovate itself.

All other interviewees confirmed the importance of collaboration. Yet, it is crucial to define clearly what, in this study, has been intended for collaboration. In fact, the term “collaboration” has been another important keyword questioned during interviews. Collaboration is a broad term that covers many means. Kouzes, Myers, and Wulf (1996), exploring issues regarding collaboratories, identify 4 broad categories of collaboration among researchers:

- *peer to peer*
- *mentor-student*
- *interdisciplinary*
- *producer-consumer*

For the present research, collaboration has been a term used in the broader scope possible, not to force meanings or prevent dimensions unforeseen by the researcher. It is in fact an opinion of the researcher that, being collaboration a boost for innovation, stricter definitions could only obstacle a deeper comprehension of the phenomenon.

Nevertheless, a discussion about the term collaboration is mandatory, especially to distinguish it between different but similar concepts.

In fact, collaboration between members in digital libraries and VREs is becoming every day more important. According to I5, “there is great interest for this topic and a bloom of proposals of new projects, procedures and standards”.

The literature review offered an overview of the background of e-science and digital humanities: interviews indeed confirmed attention to collaboration in these digital projects. During research process it emerged clearly that collaboration was the focal point of the whole study: though, it was also the most complex and full of issues, regarding especially digital humanities and particularly the Italian environment.

The first result achieved during the research process has been a deeper comprehension of the concept of “collaborative digital library”.

4.4 Collaborative editing

For the scope of the study, the concept of *collaborative editing* has been chosen to represent the ideal and strongest form of collaboration involved in a collaborative project. In fact, in the context of digital libraries the highest form of collaboration and sharing seemed to be the possibility to edit collaboratively either a text, a page or even a single line. “Collaborative editing” is the possibility of freely edit a text¹⁵, and it emerges as the core feature of wiki (and wiki-like) systems.

Therefore, collaborative editing has been chosen as the core feature of the concept of “strong collaboration”. With this term, it was intended the

¹⁵Actually, several types of digital objects could be edited collaboratively (for example, concept maps or pictures), but *texts* are by far the simplest and most explored digital objects edited collaboratively.

highest form of collaboration possible.

Indeed, interviewee I4 upheld the importance of collaborative editing for digital libraries:

Another aspect, more “borderline”, regarding research, is the one which involves the **collaborative editing**. One of the most complicated in the creation of quality archives is to guarantee the quality of both the text and possible mark up of the text itself.

Collaborative editing is then a top expression of collaboration: yet, it is easy to achieve (in wikis and wiki-like systems) but not easy to control. As literature on wiki points out (see **Chapter 2**), it is in fact collaborative editing which deeply affects the whole project and demands high attention and effort from users, who have to constitute a sort of community to make the project up and running. If collaborative editing is open to all users, the community will have to control the quality and the motivations for each edit: this kind of control, called soft security, will concern totally the community of users. Few tools in wikis have been developed to help this quality control (for example, **History** of pages and **Audit trail**, the log of each user’s contributes), and each community is demanded to decide whether or not (and in which grade) use (and develop) practices, procedures, guidelines.

The present study is not a study on online community of users and their practice in highly collaborative projects: but of course is related to these topics. Further research is needed to enlighten issues of strong collaboration, especially from a sociological point of view.

However, data gathered suggested some patterns and results concerning collaborative editing advantages and disadvantages, particularly for the Humanities community in the Italian context.

4.4.1 Consensus and NPOV

Collaborative editing is a method whose aim is always to find a *consensus* between different views and opinions. In Wikipedia, this is explicated in one of the 5 pillars of the encyclopedia: the **Neutral Point of View** (NPOV) is the (utopian) objective of every user and of every single article. The

community speak, discuss and argue to find a convergence point in writing encyclopedic articles. The definition of NPOV states:

Neutral point of view (NPOV) is a fundamental Wikimedia principle and a cornerstone of Wikipedia. All Wikipedia articles and other encyclopedic content must be written from a neutral point of view, representing fairly, proportionately, and as far as possible without bias, all significant views that have been published by reliable sources. This is non-negotiable and expected of all articles and all editors. — (Wikipedia, 2010b)

Thus, generally, collaborative editing is modeled to save incrementally each version of the page, but the version provided is always one, the last one, which is (hopefully) considered to reflect the community consensus. Collaborative editing force all users to integrate and modify a text to find a common solution for everyone, a version that could satisfy (theoretically) all users and integrated (theoretically) all opinions.

4.4.2 Tasks and forms of collaboration

An important outcome is that, in digital libraries and VREs, differences arise depending on whether collaboration is conceived *on* texts or *around* them. Interviewee I3 stated:

I think that right now we [*the interviewer and the interviewee, NdR*] are discussing about *two different forms* of collaboration.

There is the collaboration *in establishing a specific text*, which it does not exist yet: it is particular, there are some specific issues. You need to find a reference edition, you need to find a printed one, etc.

On the other hand, [there is the collaboration] *around* a text, which it does exist, which is improvable; but the text exists and it is trustworthy. Around that text, which collaborative tools could be provided? [We could] gather critic texts, different interpretations, quotations, different uses of that same text in different contexts (maybe academic or didactic).

In this sense, there is a fundamental difference between tasks and the degree of collaboration they allow. For example, a single community could collaborate in establishing philologically a text, basing on a manuscript, but not collaborate in organizing the critic literature around that text, or not willing to share different interpretations. A single task, depending on different factors, could allow a level of collaboration that another task would not allow. Obviously, tasks regarding a major involvement of scholars' interpretation would be the less easy to be accomplished socially.

Moreover, I3 does emphasize that not all tasks can be executed collaboratively, and some services in these DLs and VREs are just on demand for the user:

If you have corpora and stable texts, you can have *on demand* tools upon them. But these tools are for the single user. I don't see collaboration in this.

Again:

I think these kind of tools [statistical analysis tools] are very little collaborative, at least from the user's point of view. Softwares like Word Cruncher¹⁶ execute quantitative analysis of texts, and they can be collaborative only during the software development, but not when it's done and you have it available and running. . .

In this case, I3 makes a point highlighting the fact that each task, feature or service of a collaborative digital library is different, and demands different degree of collaboration or socialization. Each tasks should be studied and discussed to understand the degree of collaboration that could be more suitable for the designated community (and related project).

4.4.3 Lowering costs

Interviewee I4 suggested that collaboration could lower costs of certain tasks:

¹⁶<http://www.wordcruncher.com>

This kind of operations are often really expensive; they require both time and rare competences (for example, in the case of manuscripts). It would be really interesting to study instruments in which build cooperatively critical editions, or at least high quality diplomatic editions.

He then continues:

I rather think that the opportunities given *interoperability* (which can be seen as collaboration among different softwares) and *collaboration* (among different people), are very important. They can lower costs, especially in the Humanities, where critical editions can cost 20 years [of work] and lots of money and even doubts about funding. *Collaborative work among scholars*, even from different disciplines, could *solve* these problems.

Lowering costs is a clear advantage of collaborative projects, and sometimes is a key feature that is unavoidable. It is also a feature deeply correlated to the *openness* of the projects: the more users can join, the more the work can be distributed. The commonest and simplest example is always Wikipedia, a titanic encyclopedia that could not ever be written without the help of millions of users that freely join the project and contribute for free. A commercial attempt of producing Wikipedia is just unthinkable. This phenomenon of “letting users do the hard work”, commonly known as *crowdsourcing*, is a new feature of the so-called Web 2.0 and raises several advantages and disadvantages.

In the context under study, that is the Humanities scholar environment, particularly within the Italian area, it is highly controversial if it would be a good idea to let normal users join projects oriented to scholars and researchers. Interview with Umberto Eco provided an clear and critic opinion of the professor towards open collaborative projects, and also the other interviewees share several doubts about it. They will be discussed later on.

Nevertheless, it would not be mandatory to open a project to indiscriminated audience to lower costs. In fact, creating a common collaborative framework for Humanities scholars could avoid fragmentation of both insti-

tutional funds and competences that could be concentrated in a distributed framework, still in a controlled environment.

4.4.4 No “best edition”

Moreover, collaborative editing somehow questions the concept of an “ultimate version” of a text. I4 stated:

What is more, without having the presumption to aim at the “best edition” (actually, we can just dump this idea). [There could be projects where] philologists of different nationalities cooperate and suggest interpretations. This would be a brand new opportunity to change scientific cooperation in the philology area.

Editing collaboratively (for example, in a wiki system) is an experience that slowly make the user cast doubt on several practices and habits. This is due to the fact that every sentence (namely, every opinion and thought) can be doubted and questioned by other users. Every text (i.e. an encyclopedic article, a didactic book as well as a translation) is the sum of several interpretations, versions, modifications and contributes of different users.

Thus, the common idea of a final, ultimate version of a text is *doubtable* too. Strong collaborative projects as Wikipedia or Wikisource teach users that there is no ultimate version, and that every user can add, edit or question something. Each new person can add a new meaningful interpretation or information, and this can be applied also to critic or diplomatic editions, that would interest philologists and in general Humanities scholars. This is a very important consequence of the wiki-culture, and it is absolutely non trivial the possible transferability of such a concept within the cultural and scientific environment of both hard and soft, especially Humanities.

4.4.5 Neutralization process

Interviewee I1 interestingly compared the process in Wikipedia with a similar process in philology. He noticed that collaborative editing in Wikipedia, struggling for reaching a consensus among users in the name of the NPOV,

is incredibly similar to a similar process happening in philology: the process of establishing, assess and evaluate a text.

Exactly, this is fundamental: this [*the established consensus of a page, NdR*] is *the outcome of a neutralization process*. From the philological point of view, the possibility of tracking this process (through the **History** or other more sophisticated tools) is crucial. Because the Neutral Point of View [one of the pillars of Wikipedia] is the result of a process of neutralization. In philology, to track and follow this process would be absolutely important.

Moreover, this process in wikis (and other collaborative projects) is tracked by the **History** of a page: this tool is thus a log, a stratification of the various version of a page, then it is very helpful from the philologist's point of view.

For interviewee *I1*, the issue of the neutralization process, allowed by the log of page's versions, is directly linked with the issue of *authorship*, allowed by the automatic tracking of each user's edits.

Another key factor is a philologist is *authorship*: to find out who did what, to tracking each user's contributions. You need to individuate who wrote something.

In fact, wikis do already own basic tools for log and track edits, both for users and pages' versions. Though, they are still too simple and plain to serve communities of scholars. Interviewee *I1* continued:

I wanted to emphasize the importance of this [authorship matter]. The centrality of the user (in Web 2.0 there is a passage from centrality of document to centrality of users) takes us to these "bio-bibliographies".

Within philology, "the history of a text is done through the history of the versions of the text". Thus, the text does not exist without being in relationship with the people who have something to do with the text. This *traceability demands to be also theoretically emphasized*, because in crucial, in philology.

4.5 Community

In the context of collaborative digital projects (particularly, digital libraries and knowledge production projects), the community who is behind the project is the most important stakeholder. It is the main character, because a collaborative project without its members is just an empty (virtual) space. This is especially true for wikis, in which the term often indicates the unit of both the community and the infrastructure itself (Koblas, 2006).

Therefore, for the scope of this study it is mandatory to explore the concept of scholar community in digital collaborative projects.

According to Wenger (2006) “communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly”. In his previous works Wenger did also differentiate from *communities of practice* (CoP) and *communities of interest* (CoI), stating that members of the CoP are active practitioners wherever members of CoI are only interested in sharing information and discussing a particular topic that interests them. Within the present study it will be assumed that huge and open online communities (i.e. the Wikipedia community) as well as more institutional communities are both communities of practice, because the sharing of information and discussion is aimed to a better understanding of practices and to an actual production of knowledge.

Nonetheless, even if both assumed as CoPs (which is not immune of issues), these communities vary largely one from each other.

An helpful discrimination between online communities dedicated to knowledge production was suggested by professor Eco, in his interview as key informant. After discussing about Wikipedia and the *wisdom of the crowds*, the researcher asked him about collaboration in scholar communities. For the sake of comprehension, this section of the interview will be entirely reported:

Researcher: What is your opinion about scholar collaboration in humanities?

Eco: This is another topic. Congresses were made by ecdotic scholars to investigate this topic. These are truly auto-controlled communities.

Researcher: Communities of practice.

Eco: Yes, but where we know that a single scholar belongs to a single university, we know where he comes from. In this case, it happens something similar to when people used to collaborate in writing a book, and they needed to take the train once a week to meet and discuss. It is a collaborative team work that is controlled by someone. It is not the wisdom of the crowds. It is simply the scaling and the simplification of a collective research work that once required filthy travels and nowadays it can be done online daily. [...] I'd rather call them *uncontrollable* and *controlled* communities.

Researcher: It is very interesting when these controlled communities (which are granted and comes from determined institutions) do not have a hierarchy or a chief, but they auto-control themselves. [...] In your opinion, is this auto-organization possible also in these scholar communities?

Eco: I recall a conference in Bologna, about ecdotic studies, that was dedicated mainly to digital humanities projects and text research environments and functionalities. Evidently, this was an *headless* community, auto-controlled and headless. But “headless” is a phrase: because in scientific communities which self-legitimate there's always someone who gains more authority: if an important philologist propose an interpretation, the others will follow.

This study will assume hence the discrimination of communities in *controlled* and *uncontrollable*.

Controlled communities are relatively limited and more homogeneous groups of people, who share precise competences or at least a domain, where anonymity is not appreciated, authorship is fundamental and real life boundaries, connections and hierarchies are exported within the digital world. This is the case for example of communities of scholars: they share a precise context, aims and competences. They too wish to receive credit or some form of remuneration (money, reputation, authorship) from participa-

tion in an online community. Often, roles and authority gained in real life force a hierarchy also online, in the digital environment.

This is quite different from totally open and “anarchic” communities as the Wikipedia community¹⁷, where participants are volunteer and do it “for fun”. These communities do have less boundaries and limits, and participants get involved for their choice, without any evident reward. Both these communities and scholar communities can be defined communities of *peers*; on the other hand, open communities are definitely more flat and horizontal.

Therefore, the so-called “wisdom of the crowds ” is very different from scholar collaboration, in terms of scale, openness, hierarchies, freedom.

An in-depth analysis of differences and similarities of these two phenomena is beyond the scope of the present study: nonetheless, further research is needed to gain deeper understanding of two diametrical forms of collaboration that share (or could share) much more than they at the present days.

For the scope of present research, it will suffice to say that scale, authorship and openness seem to be crucial dimensions in which open and scholar communities sit at the opposite extremes. Still, similarities arise to a deeper and non trivial observation. Communities dedicated to knowledge production try to auto-control themselves, and policies and procedures are negotiated directly within the community. Both type of communities then are, theoretically, community of *peers*, in which agreement and consensus should be reached all together. Further research is needed to verify these outcomes.

4.5.1 Control, review and quality

The issue of control is probably one of the most problematic in online collaborative projects, especially if open and wiki-style. Projects like Wikipedia commit the review and the quality of the content directly to the same community why is producing that content. It is a sort of peer review, though not formalized and relegated to the good willingness of users. The most con-

¹⁷Actually, it would be more precise to state that the Wikipedia community is composed by several communities of practice, because of many users collaborate only with a small subgroup of fellow participants who share the same interest about a topic.

troversial aspect of wiki projects is in fact this: how it is possible to trust information if you are not sure about the author and the reviewer? This question would not be answered here, going far beyond the scope of the present study. Nevertheless, it will suffice to say that the auto-organization of the Wikipedia community seems to work in assessing quality of the content, on average. Several quantitative studies, as Wilkinson and Huberman (2008), suggest that the more users contributes and edits a page, the more the page, on average, is accurate and provides trustworthy information. Further and more qualitative research is needed to evaluate and comprehend these outcomes.

Regarding Internet, Wikipedia and other peer-generated information, Eco thinks is always a problem of *filtering*:

So Wikipedia, as the whole Internet, has the problem of *filtering the news*. *It keeps both false and real news*; but the rich knows filtering techniques at least for the areas they know how to check. [...] Collective filtering is useless, since it could yield to fluctuations. I noticed that in a certain period of Berlusconi's triumph people went looking for information about me on the right-winged books and placed them in Wikipedia: as correctness prevents me from changing them directly, I left them on. But obviously it was an entry made by the winners of the moment.

The collective control is therefore useful up to a certain point: it is conceivable that if one gives a false length of the equator, sooner or later someone comes and corrects it, but correction of more subtle and difficult issues is more complicated.

In scholarly publishing, peer review is much more formalized and demands competences and economic resources. In collaboratory digital libraries, it is not yet understood (or foreseen) how the process of assessing quality would work. Data suggest it is a matter of both authority and community.

Eco strongly believes in the difference of open, *uncontrollable* and anarchic communities of users and *controlled* community of scholars.

Take for example the journal *Nature*. In the scientific world, if an article appeared on *Nature*, where there is peer review and large control, it is taken seriously. It is true, in any case, that *Nature* could make an error, and exclude a brilliant article: nevertheless, *Nature* is believed a landmark of reliability, with fringe boundaries. There is always the possibility of an error, or event a little academic revenge. . .

He believes in the final convergence on a consensus between peers, but only in a community with defined boundaries, with defined names and curricula, which guarantee authority of information and possibility of control.

I am a disciple of Peirce, who states that the scientific truths get assessed and approved by the community. He intended the *scientific community*, at his time much more divided by the normal crowd. The slow work of the community, though errors and revisions, carries on the “torch of the truth”, as he said in the XIXth century.

And again:

Things eventually *get fixed*: these are the controlled communities, which are not anarchic, but with a fringe authority. This doesn't regard Wikipedia, in which the anarchy is much bigger.

Thus, for Eco it is not possible to compare communities of scholars and communities of general active users as in Wikipedia. Also other interviewees shared a similar (yet less incisive) thought: some of them sincerely confessed that opening a scholar collaborative project to general users would be often counterproductive, because people could vandalize or worse insert false information and argue with no competence about very specialized topics.

Though, the researcher somehow feels that Eco did attribute some flaws to collective processes that he did not see (or appear to see) in scientific communities. He trusted the scientific community to reach, eventually, a convergence that he's not willing to attribute at the communities of active users of Wikipedia.

Although the researcher is absolutely not competent of *sociology of science*, yet, it is his humble opinion that a statement like this may be disputed in the future. Sociology of science is a quite sectorial and young field of study, but provides interesting insights about behaviour of scientific communities (and single scholars), illustrating a much more (and disturbing) picture of science, research and scholarship. Mechanisms of knowledge production and science discovery are not immune of the influence of external factors as grants, scientific fads, trends, mainstream science, competition among scholars, etc. Even the existence of a core, mainstream science is a political fact that prevents a free and totally open science and research, especially in developing countries (Guedn, 2009). Further and multidisciplinary research is then needed to understand similarities and difference of scholar and amateur communities, that are maybe more related than expected.

4.5.2 Fear

This explains (and it is caused by) the diffidence of scholars regarding collaborative projects, especially open ones, as Wikipedia. Wikipedia is open to *everyone*, and this is a flaw that few scholars tolerate.

In fact, open communities are still high controversial and very difficult to comprehend in all their facets, also for inside participants. Many factors are involved: social, economic, psychological ones. Moreover, interviewee *I5* admitted that scholars are often scared to a position, and an authority, which has been gained through long years of study, and that nowadays is sometimes underestimated.

The word “collaborative” really scares the cultural establishment. Because it takes away *power*, and the *goal*. Yes, a philologist could be a reviewer, also with the so-called “educated non professionals” who could execute bulk jobs; he could control and assess quality. But still, this really scares academic world.

In the conversatoin with Eco, the topic came up:

from mighty to influential

Interestingly enough, I5 indicated with the label “educated non professionals” what professor Eco called “motivated crowd”, both emphasizing competence and motivation of non scholar users in open communities. Further research is needed to enlighten the confusion regarding various types of users in open communities, and also motivation, aims and jobs of these subgroups.

4.5.3 Boundaries

Especially in the Italian environment, it is still strong the idea that *control* means *boundaries*, limits. Eco itself confirmed this impression, providing several examples of categorization, discrimination between different kind of people, during the whole interview. At the beginning of the conversation, he made a distinction between two kind of information *users*. He stated:

I once made a distinction between things good for the *poor* and things good for the *rich*, where rich and poor have no immediate connotation in terms of money, but in terms, say, of *cultural evolution...* *A graduate is a rich, an illiterate is poor.* There can obviously be a big entrepreneur who is poor and a little clerk who is rich.

Television thus is good for the poor and bad for the rich: it taught the poor to speak Italian, it is good for old women who sit alone in the house. And it harms the rich because it prevents him to go out and see things more beautiful at the cinema, it narrows his ideas.

The computer in general, and the *Internet* in particular, *is good for the rich and bad for the poor.* That is, Wikipedia is good for me, because I am able to find the information I need, I do not trust it, because everyone knows that as Wikipedia grows, the errors also grow. I found steep follies written about me, and if no-one pointed me at them, they would stay there still.

The rich are grown people, they can compare the information. I look at Wikipedia in Italian, I’m not sure that the news is correct, then I go to check the English version, then yet another

source, and if all three tell me that this gentleman died in 371 AD I begin to believe it.

To explain the success of open projects, he suggested an interesting definition. He divided the normal, “low” crowd, which often supports every kind of falsity (“If you ask to 6 billion of inhabitants of the world, the majority will tell you that the Sun is going round the Earth. There’s nothing to do about it”) to what he called the *motivated crowd*, which can actually create trustworthy information, as (sometimes) happens in Wikipedia.

We must therefore find another criterion, which I think is the *motivated* crowds. People who work on Wikipedia are not just an aristocracy, just professors, but they are not the indiscriminate crowd either: they are the part of the crowd who feels motivated to work with Wikipedia.

Thus, he continues:

Here it is: I’d replace the theory of “wisdom of crowds” with the theory of the “wisdom of the *motivated* crowds.” The general crowd says that we should not pay taxes, the motivated crowd says that it’s fair to pay them. In fact, it is not a digger or an illiterate who contributes on Wikipedia, but someone who already belongs to a cultural crowd for the very fact of using a computer.

4.6 Authorship

Collaboration between scholars, especially in the field of humanities, is intended to be the most important topic of the study. Data gathered suggest a consensus in noticing the same barriers to collaboration. The main barriers found are related to *authorship*. This outcomes primarily refer to the Italian community of humanists, but some of the results can be extended to other national communities.

The term “authorship” was chosen to indicate different but deeply correlated issues. Data gathered assessed the idea of *authorship* as a main issue

for collaboration, which tends to presents in different facets summarized as follows:

- *myth of the lonely scholar*
- interpretation
- attribution
- intellectual property

These minor issues will be analyzed as aspect of the main authorship problem.

4.6.1 The *myth of the lonely scholar*

It is meaningful that all interviewees agreed in confirming the fact that humanists are used to work independently, aside any discipline or nationality, perpetrating the “*myth of the lonely scholar*”. This concept is part of the core culture of all Humanities and it is accompanying the entire community of humanists from the roots of culture itself. Interviewee I5 stated:

We need to state again that, so far, in humanities, work has been primarily individual. There is the *myth of the lonely scholar*, as they say in the Anglo-Saxon area. At the same time, there is the *myth of the lonely author*. It is a romantic idea, where the author, the scholar, is always working alone: this is still true.

He then continues claiming this behavior to be an obstacle for collaboration:

When you make a research in hard science, it is often a collaborative team work. In humanities you use need to go in a library, in an archive, in your *scriptorium* and write until you have finished you work. How much Digital Humanities have changed this behavior and culture needs still to be assessed.

Literature and other interviews confirmed the same views.

4.6.2 Interpretation

If the scientific method is founded on experiments and measurability of data, in Humanities there is no such a thing. Humanities found themselves on authors' and scholars' *interpretation*. Each individual interpretation is legit.

Eco explained this difference between hard and soft science as follows:

Science is *cumulative-destructive*, it stores what it needs and throw away what it doesn't require. Humanities are totally *cumulative*, they don't throw away anything: in fact, there is always a return to the past.

On the other hand, they are totally destructive in the way, as Maritain stated regarding to Descartes, "a philosopher is a *novice in the Absolute*". For Descartes, everything that philosophy stated before him was false. If a mathematician did that, it would be the end of mathematics.

Hence, we have one crucial difference between hard sciences and humanities: if in hard sciences the entire community of scholars and scientists pursue the the ideal of universal agreement, humanities are less tied to the concept of *agreement*. *Measurability* and *verifiability* are pillars of philosophy of science, and they are not in the humanities. Differences between these two fields of human thought has been investigated for centuries: for the scope of the present study it will suffice to state that measurement in hard sciences provides a framework where to converge that is not available in Humanities.

While hard sciences struggle to converge to a consensus and to theories tested and verified by scientists all around the world, Humanities do not seek a convergence, but they maintain dignity to the single view of the individual.

Of course, each interpretation needs to be correctly motivated: there is still the aim of being "scientific". Nevertheless, there is surely less struggle to achieve an agreed, impersonal and *objective* opinion: often, there's no objectivity¹⁸ at all.

¹⁸The researcher knows that the term *objectivity* can be misread in the context of soft

In philology, different scholars can argue about how to read a line or a word, and agreement is not mandatory, sometimes not usual at all. This is directly an obstacle for collaboration, because collaboration itself premises a consensus between parts.

In this way, authorship cannot be forgotten. The authority of the scholar directly affects the importance of his interpretation, and viceversa. Interviewee I5 claimed that authority is still crucial in philological sciences:

In philological sciences, in particular, there is one thing not to be underestimated: We read an edition of a particular curator or publisher because we know that he is an international expert. I can read the Petrocchi's critical edition of Dante over another because I prefer, I respect, I trust more that curator or publisher. In Humanities and philological sciences the author, *the scholar is king*.

4.6.3 Attribution

Interpretation also brings with itself all problems related with attribution and authorship: in humanities, an interpretation is owned by the author, who demands to be recognized.

Eco stated:

For what are soft sciences, *there is absolutely less impulse to collaboration*. There is much more interest to be the main character of an idea, than being just a “water carrier”.

That's for sure. A scientist in these cases is used to not being mentioned and to know that however is carrying forward a fundamental research. In soft sciences, this happens only to the exploited student who is sent to gather data that the professor will sign and profit by.

That's an old story, there's no escape from that. . .

Far from being only a intellectual property problem, it is more a cultural issue. Humanists are sometimes really jealous of the texts they are studying sciences: however, for the sake of the argument, here it is used to highlight the concept of convergence and scientificity

and working on. This is obviously an obstacle to sharing and collaboration, because if in an environment competition is hard collaboration it is not likely to happen. According to interviewee I4:

Usually, when a philologist or in general an humanist is working on a text, he tends to consider it *his property*... He surely doesn't like other people to study the same text.

He continued:

Working in different disciplines within Humanities, I felt and shared the feeling that computer science is important to renew your research and that collaboration is important: nonetheless, at the end of the day everyone has his own personal project, his own texts which is working on and he also gets mad if someone else copy him or just work on the same text.

Moreover, he correlates this approach with hard science different approach:

In the Humanities, a scholar writes and that is his intellectual product, that is his accomplishment. In STM, there are patents, technologies, tolls to be developed. Furthermore, aside some rare exceptions, no theory is developed on a single researcher, but on previous results. [...] For humanists, this is more complicated, because interpretation has value on its own.

4.6.4 Intellectual property

Moreover, intellectual property is a problem on its own. Because each author's interpretation is so significant, intellectual theft are very feared. Interviewee I3 confessed:

This is [a complicated issue] complicated, I admit that I'm a bit more conservative about this. In Humanities a scholar writes and his byproducts what is its intellectual products, research, is that it produces.

Within science, there are patents, technologies, tools are developed. Moreover, apart from rare exceptions, no theory is constructed on the singular individual, but more on past research. [Technological and scientific] innovations innovate only 5 percent of earlier theories. Contributions in these theories are multiple and minimal.

For us [in Humanities] it is a bit more complicated, because the *interpretation* has a value in itself. So I understand the issue of protecting even the letter of what I write, so no one appropriates it to write his text. There is an authorship problem, which is important and must be addressed. This also does not cover the speech open license and permit closed, but the recognition of intellectual authorship.

Internet in fact undermines the effective protection of intellectual property, and people copy. [...] The issue of *economic* rights is almost *insignificant*, there are virtually no revenue for humanistic work. I am convinced that the protection of open content, along with the protection for those who first expressed a particular idea in a particular form is a goal to pursue.

Even if it's not an economic problem, stealing of intellectual property is very feared in Humanities, due to the high value of each singular interpretation. Collaborative projects don't often pay enough attention to each author's work, with no sufficient effort in developing tracking systems for author's contributes. Thus, collaboratories digital linbraries for Humanities should pay extra attention to technical and management tools for tracking of edits and attribution of authorship.

4.7 Scientific social networks

It is beyond the scope to investigate deeply differences between collaboration and communication, but as they were topics involved in the study it is mandatory to provide at least a common definition.

During the designing of the present study, the researcher assumed collaboration and communication as linked but different concepts. Although

they are obviously bounded (no meaningful collaboration is possible without communication, communication itself is a form of collaboration), they are not the same thing.

The term “Web 2.0”, although is useful for suggest a cyberspace more oriented towards social interactions and users rather than services and documents, is also an evergreen label used largely for marketing. What is more, it often confuses different concepts as communication or collaboration, using them both to express a generic fostering of social interaction.

In the present research, the terms will be assumed as related but not equal. A collaboratory digital library is a space for collaboration in study and research, and therefore needs social features to allow communication between members. Therefore, here it will be assumed that a system for collaboration is also (but not only) a system for communication.

The contrary is not always true. According to Boyd and Ellison (2007), social networks sites are

as web-based services that allow individuals to

1. construct a public or semi-public profile within a bounded system,
2. articulate a list of other users with whom they share a connection, and
3. view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site.

Generic social networks (i.e. Facebook¹⁹, MySpace,²⁰) or even task-oriented social networks (LibraryThing²¹ and aNobii²² for books, Last.fm²³ for music, etc.) are focused on social interaction and communication, and they do not allow the *collaborative editing* of a text or a digital object, that is what in this study it has been assumed as the core of strong collaboration.

¹⁹<http://www.facebook.com>

²⁰<http://www.myspace.com>

²¹<http://www.librarything.com>

²²<http://www.anobii.com>

²³<http://www.last.fm.com>

Therefore, in the scope of this study social networks will be not treated as collaborative frameworks, VREs or laboratories, due to the fact that the former are concentrated in communication and social interaction and the latter in collaboration and production. A social network therefore is not necessarily a collaborative network, if it does not provide tools for collaborative *production*. Viceversa, a collaborative network is also a social network (and this is a reason why Wikipedia is often regarded as a social network).

The difference here explained is a personal opinion of the researcher (motivated by literature, personal experience and different conversations with net experts), that thus was discussed about this topic with the interviewees. Here is an excerpt of the conversation with interviewee I5:

I5: Where do you make the *social* part finish and start the *wiki* one?

Researcher: Obviously, everything is blurred together, but I make this difference: in wikis, communication, the social part, is *upon* something, is aimed to knowledge production. The social component is very important, also in wikis. But there is a difference between, for example, a generic wiki project and Facebook. In the first, you produce and communicate, in the latter you just communicate. In the first, communication is connected to a task, collaborative production of “something”. We are so a community of practice, because we *make* something (in Wikisource, for example, linked to books and texts).

I5: In the electronic environment, everything is much more complicated, and distinctions as well, but you’re right on this. On Facebook, few days ago I wrote a post on a congress’ keynote; I tagged all my contacts related to Digital Humanities, and some of them commented. But you’re right.

I produced something, but alone, an other commented in a social way. We communicated, but the text was mine.

Thus, a little but clear difference will be maintained among the adjective *collaborative* and *social*: the first suggest a common workplace, the latter a place where to chat, communicate, even share information and opinion, but

not produce digital objects in cooperation with others.

It is not meant to attribute a different value of collaboration instead of communication (or viceversa): just to provide a distinction whose hope is to be a clarification, not a “hierarchy of values”.

Nonetheless, interviewee I3 did actually highlight the importance of scientific social network as a framework for collaboration:

More than a direct collaboration within a collaborative digital library, I think is important the *social networking aimed to research*. What I think is really missing a lot in Italy is just that.

For example, a project as Academia²⁴, is trying to create social networks, similar Facebook, but research-oriented. There are only two or three examples of this kind, and yet not well developed, but I think it would be extremely helpful to work collaboratively.

In a sense, all this could be carried out directly by Facebook, if it allowed *typified* content. In Facebook it’s all one mashed pot of content: you cannot filter by content, and that’s a great pity. If you had a layer on Facebook that gave the “contact” and “contents of those contacts that are focused on research”, that would be great.

Moreover, I4 not only thought this would be important, but also I argued that in the Italian context this would more important than a real collaborative digital library, because is a previous step.

This is what you need, I think, *even before* thinking to collaborate on digital libraries. You need to *establish relationships*, they come first.

Within the DL environment, ideas are very beautiful, but perhaps they require *prerequisites that we have not yet*. First, *in Italy we do not have the digital library*, which is essential [*laughs, ndR*]. Second, we don’t have a research community used to build a research social networks.

²⁴<http://academia.edu>

I believe that these two things are essential to work. Wiki systems, which are best systems available nowadays to do this kind of things, are still used in a very amateur way. Including me; it is not a critic to users or to the instrument itself, but a critic to the *approach*.

Thus, research social networking has been considered as an important prerequisite for real and efficient collaboration within Humanities in the Italian context.

4.8 Laboratory

During the whole research process, a concept emerged from literature, interviews and above all from the same experience of the researcher, who was using the computer (wiki, concept map software, L^AT_EX) as a research instrument: the idea that several tools combined can create a framework, a *laboratory*. This is a simple yet powerful concept that believes in the emergence of a new structure out of single, discrete instruments. The best example is the laboratory of a smith or a carpenter: it is not a disorganized collection of individual tools, but a framework of correlated instruments that together are much more the simple sum of them all. A framework where the smith or the carpenter can work, product and maybe *create other tools*.

This well-known idea is very explicit especially in the technological evolution, where rates of growth strictly follows exponential graphs (Kurzweil, 2005): this is the natural consequence of building tools with previous tools, as well as frameworks with previous frameworks. It is a *multiplicative paradigm*, in which every factor is multiplied by, and not added to, the following factor.

The history of technology teaches us that some mechanisms of the world are exponential, some other no. Technology itself, and scientific evolution too, follow this type of growth: this is the reason of the existence of empirical laws like the Moore's Law (Wikipedia, 2010a), which state that every 18 months a microprocessor would be twice powerful and cost half.

In a minor way, also Eco suggested that few tools combined could simplify greatly the accomplishment of some tasks:

It is a matter of time. When I write, I go on Wikipedia 30-40 times a day, because it is really helpful. When I write, I don't remember if someone was born in the VI century or the VII; or maybe how many n are in "Goldmann"... Just few years ago, for this kind of things you could waste do much time. Nowadays, with Wikipedia and Babylon, which check the spelling, you can save it a lot.

Finally, the researcher strongly believes in the importance of building open frameworks where users can exploit their creativity and competences not only to work on their object of study, but rather to create, build and develop other tools and instruments they could use on texts. This has been true in the little experience of the researcher himself, who developed during the whole research process an incomplete yet helpful framework for accomplishing the research tasks. In fact, the researcher gradually developed a framework that helped him in storing and retrieving information, as well as storing data and analyze them. The framework has been previously overviewed in **Chapter 3**.

When asked, I5 corroborated the concept:

This is exactly the theme of VREs. If I have to distinct tools from two different developers, and I want to have them both available on my VRE, I need standard and interoperability. It is a very complex topic.

Interviewee I3, too, was asked if he agreed with the fact of innovation to be generated by frameworks, that were themselves generated from previous tools:

Of course. The knowledge has always worked like this, with people which discussed and changed opinion and mutually shared information. Collaborative frameworks allow at ideas to become convergent, and to do it without the boundaries of space and time. This is fundamental.

4.9 Amateur and scholar digital libraries

During the research process, a topic emerged clearly as prominent: the comparison between scholar and amateur communities working with digital libraries. Due to different experiences and aims, these communities both are similar and very different. These communities often share the same object of study (i.e. public domain texts), but do differ in work-flow, practice and expectations. A result of the research is that there is a significant difference in approach towards collaborative projects, also projects constructed upon wikis. In fact, scholar communities seem to be not interested in wikis for collaborative digital libraries, because they prefer more specific and complex tools. On the other hand, amateur communities do not have same expectations as scholar ones, and thus find wikis are far more useful and helpful.

Interviewee I4, having a good experience of both the communities, highlighted some differences:

I know both the two different approaches: the first is the amateur approach of people who digitize texts for interest or passion, without having any scientific purpose, and even then the pressures that this entails.

The other approach is the one of scholars or students, generally experts, in which defined collections are created, and where there is a plan work. The job is carried out in substantially different ways. Often texts are difficult to find, difficult to code, etc..

Thus, it is patent the fact that communities of amateurs do not have the same boundaries (especially legal and correlated to funding and scientific methods) of scholar ones: there is much more freedom (others would say “anarchy”). Being volunteer-drive, amateur DLs do have little or no money involved: this is both a disadvantage (it limits the scope and the possibilities of the project) and an advantage (the community does work *for fun*, no money means often no pressure). Wikis do work very well in large communities because they eliminate bottlenecks and allow all users to work without hierarchy restrictions. Other data suggest that these amateurs communities do have less limits and expectations.

A proper comparison between amateur and scholar communities of practice working with collaborative digital libraries could not be accomplished within the present study. The researcher in fact contacted the Liber Liber community to run an online focus group about the imminent shift of the project to a wiki framework. Unfortunately, due to low response of the community and technical problems, the focus group could not be executed.

Nevertheless, interviews conducted offered meaningful data that suggest some clues: further research is needed to prove the conclusions and gain deeper understanding of the phenomenon.

Amateur communities of practice did find a great increase in the last years, thanks to Internet. Amateurs gather in every angle of the Net for discussing about topics and sometimes to produce content and information. This is a prominent part of what has been called “Web 2.0”: the gathering of people in communities of *prosumers*, that are both information producer and consumers.

Regarding digital libraries, the first and most important amateur project is the Gutenberg Project, which was born in 1971 and has been digitizing over 30.000 documents (Wikipedia, 2010c). In all these years, the Gutenberg project has inspired several projects worldwide: one of them is the Italian Liber Liber, that was born in 1993, in Rome, to compensate for the lack of similar projects in Italy.

4.10 The Italian gap

Several interviews provided insights about different issues in Italy, and a draft but meaningful picture of the Italian context.

Moreover, the study had the chance to gather data about the Italian cultural environment to one of his most prominent members. In fact, the best achievement of the present research has been the interview of Umberto Eco as an important insider of Humanities in the Italian context. Professor Eco is a worldwide known professor and author of best-sellers as “The Name of the Rose” and “Foucault’s Pendulum”. He is also a medievalist, semiotician, philosopher, literary critic and novelist: he has written from academic texts

to children's books, as well as many essays regarding very different topics, from politics to literature.

Professor Eco was interviewed by the researcher for a project called "Wiki@Home", supported by the Italian chapter of the Wikimedia Foundation²⁵, Wikimedia Italia²⁶. Wikimedia Italia (WMI) is an association for open culture and open knowledge, and it supports Wikimedia projects as Wikipedia and Wikisource. Wiki@Home is a subproject of WMI which is aimed to interview important members of the cultural and entertainment world.

Within this context, the researcher contacted and gained an appointment with professor Eco to interview him about Internet, collaboration and Wikipedia. The researcher thus exploited this unique occasion to utilize professor Eco as a *key informant* for the study, investigating topics as collaboration between scholars, especially within Humanities in the Italian context.

Interview revealed to be extremely rich and helpful to gain in-depth information about Humanities in the Italian area and their approach to Internet and collaboration. Eco's and others' interviewees opinions will be analyzed in the next paragraphs.

4.10.1 Political, cultural, institutional issues

Interviewee I3 highlighted and summarized several issues, at different levels, of the community of Humanities. He firstly emphasized cultural issues:

There is a mentality and cultural problem, regarding Humanities. In my job I had the lucky chance of knowing different areas of expertise. I always found that people feel computers as an important tool to innovate their research and studies, and also that collaboration is important. But at the end of the day everyone has always his own project, his own studied text; people get mad if they discover that someone else is copying or even studying the same text.

²⁵<http://wikimediafoundation.org/wiki/Home>

²⁶http://www.wikimedia.it/index.php/Who_we_are

But, eventually, many serious issues are *political* and *institutional*. One of the main problem (which unfortunately is present in all levels of the Italian political and enterprise establishment), that is the old age of rectors, professors, deans:

The problem is that my generation [which is composed by] young professors does not have decisional power within universities, so they can't support and endorse this kind of projects.

This issue is a well-known consequence of the Italian tradition of maintaining a position until elder age. Unfortunately, young people do not have the chance of getting authority before they are no more young, and so the whole establishment (political, institutional, economical) tends to be held by elder people, with a consequent lack of attention and interest to innovation.

In the last years, Universities saw their funds being reduced year by year, and the situation is not improving. Obviously, in such a grim situation, there is little space for really innovative projects, and it is very hard to find institutions funding big, cutting-edge projects. Interviewee I3 stated:

Moreover, there is a huge issue of competence and background, and on the other hand a lack of adequate infrastructure. In UK e Germania si investe molto di E-science in ambito umanistico, di Digital humanities. Non solo non si fa, non se ne parla proprio, ed un limite grosso. Anche la stessa comunit di informatica umanistica si fermata su questo punto. Bisognerebbe agire su vari livelli, partire da una piattaforma e farlo vedere, farlo valutare in sede concorsuale.

Moreover:

The issue is providing a institutional framework which will support [these projects]: if there was a national institution for e-science or e-literature, for example, even with limited fundings, that would be the best location for experiments and test.

Academic publishing and career issues

Finally, he emphasized very serious career and tenure issues. In the US academic environment, it is popular the phrase “publish or perish”, meaning the “pressure to publish work constantly to further or sustain a career in academia. The competition for tenure-track faculty positions in academia puts increasing pressure on scholars to publish new work frequently.” (Wikipedia, 2010)

Related to this, as both a consequence and a cause, there is the fact that publication on academic journals is one of the few possibilities for scholars to improve their visibility, and maintain or obtain the position in universities.

Now, there are issues related with tenure and *career*, somehow complicated. In tenure competitions, digital editions and digital publishing do not count. *They are not evaluated.*

Academic publishing is thus the main way to further a career in academia: unfortunately, this runs against the possibility of creating and sustaining online collaborative VREs, because often working in such projects is not seen as a proper task for researcher, and it is not relevant for evaluation of a scholar. Moreover, this is a big obstacle for collaboration and collaborative projects:

Think about what could happen in a collaborative work, where it is not possible to redeem attribution of every edit. This is a serious problem, that will be maybe solved by younger generations of scholars, who are more aware of importance of sharing.

4.10.2 Italian digital libraries

Interviews did also confirm the backwardness of Italy regarding digital libraries and their conservative approach. Only few institutional libraries are actually online, and unfortunately amateur projects still are very important for their digitizing work.

Indeed, all interviewees related to the Liber Liber community expressed the same thought: the fact that Liber Liber was born to be an experiment

(and a model), and it would have been substituted by more institutional and authoritative projects. Unfortunately, this did not happen. Italy still lacks institutional digital libraries comparable with great national projects as the french Gallica²⁷ or the American Memory²⁸, and even in European collective projects as Europeana²⁹ the presence of Italian documents is minimal (and sometimes provided greatly by non-Italian institutions as Gallica). What is more, even though Italy can boast centuries of culture (in almost every field of human creation), still little has been done to digitize this content and make it available online. Several nations, with far more humble and smaller heritages, spend greater effort in digitizing cultural heritage material.

Interview I2, one of the founders of Liber Liber, stated:

At the beginning, we filled a gap, *in Italy there were no e-books at all*, at least in Italian. There was the Gutenberg project, it was in English, and we wanted to have at least an Italian project.

Passing the years, even with a great response from users, *we noticed to be still alone*: until recent years, to the birth of Wikisource and Distributed Proofreaders, there were no other projects that did what we were doing.

Also nations (in Italy very little, sincerely) are moving towards digitization.

Other interviewees share the same thoughts. Interviewee I3 recalled the experience of *Biblioteca Digitale Italiana*, the first and most famous (and, in a certain sense, the *only*) institutional digital library for the general public. He also express the feeling that the real issue is not the lack of resources, but mainly the lack of culture and mentality:

Then, I also think that Italy lacks institutional digital libraries. Apart from the first experiments of *Biblioteca Digitale Italiana*, (which was very professional, TEI marked-up, but actually really small).

²⁷<http://gallica.bnf.fr/>

²⁸<http://memory.loc.gov/ammem/index.html>

²⁹<http://www.europeana.eu/portal/>

Aside from that in Italy there is *nothing* [...].

There's lack of an adequate culture: when they started digitizing in Italy they started from historic bibliographic catalogs. There were beautiful bibliographic cards, of course, handwritten and from the XVII-XVIII century; but definitely not very attractive to the general audience. Other nations are much smaller and culturally poorer [than Italy], such as Norway, are far ahead of us. It is a matter of approach, not just resources.

Interviewee I4 also shared the surprise (and probably delusion) that an amateur project such as Liber Liber is nowadays still very important, because no other institutional projects followed:

To be honest, we all thought that the work of Liber Liber was necessary in the first phase, hoping that the digitization would have been further made by large national projects, universities, institutions, in a professional manner, and not voluntary as we did.

Unfortunately it was not so, at least not in Italy.

Liber Liber is thus still a reference point for Italian digital libraries. What is more, it is not a proper digital library: it does not guarantee certain levels of science, despite the good will and the competence of volunteers. The same framework would be more complex.

Situation of Italy regarding is thus quite clear. There is little attention to innovation, and even the participation of Italy in big European projects still is insufficient, given the richness of Italian cultural heritage. Moreover, regarding digital libraries Italy is focused on preservation, and still maintains a very conservative approach towards innovative and advanced projects. Data from literature suggest also that Italy has so far lost several opportunities to innovate his approach to cultural preservation and access, and that much is to be done to reach the standards in these ambit of other European countries as France or Germany. Recent news tell that the Italian Government has already subscribed a partnership with Google for the digitization of over a million books and documents in the well-known (and controversial) **Google**

Books project, but it is still too early to evaluate outcomes and results of the mentioned settlement,

4.10.3 Amateur digital libraries go wiki

In 2001 Wikipedia rise up and community of volunteers in Internet would never be the same again (Lih, 2009). This project gathers millions of users around the world and in less than ten years has been producing over 15 million articles (over 3 millions just in English)(*List of Wikipedias*, 2010). In 2003, a sister-project is born, called *Wikisource*, a wiki digital library which provides texts free of copyright (Wikipedia, 2010b). The Italian Wikisource has been working since 2005: it started taking content already provided by Liber Liber and then began to digitize documents also providing the images of scanned books, which has been a step of innovation towards other amateur digital libraries. Liber Liber itself got interested in the wiki framework, deciding to adopt one. At the moment of the writing of the study, the moving has not been completed yet.

Interviewee I2, was asked to explain why the Liber Liber project chose to move on a wiki³⁰:

Firstly, to provide [the community] a more efficient instrument for collaborative work than before: [the previous system was in fact] based on a mailing list. The editorial staff used to create the content and subsequently send it to an HTML expert, who had to transfer it on the Web. Though, this created bottlenecks that were unsustainable, given the number of volunteers....

The wiki system thus seems to meet some needs of the community, who demands to smooth the work-flow and eliminate bottlenecks. Wiki systems, allowing collaborative editing, greatly improve quickness of these collaborative projects, due to the fact that almost all tasks can be executed in parallel and few bottlenecks are allowed³¹.

Then, among the various tools that enable online collaboration, some of which are also best suited to manage online libraries

³⁰The researcher knew this information months before starting the study.

³¹This explains why wiki are called like this: *wiki* means “quick” in Hawaiian.

(such as DSpace), producing a data stream OAI-MPH, however, we chose a wiki [the technical name of the software is MediaWiki, ed] because we are a volunteer-based initiative, then we are interested first of all spark enthusiasm. Then, the node to adjust MediaWiki, so that can produce a data stream OAI-MPH can be done later. The opposite, with a specific tool, but little exciting in terms collaborative was much more difficult. I'd rather have a tool that works in terms of social and technical improvements that have a specific platform and already interoperable OAI does not work in terms of collaborative, rather it is designed for university employees who enter data and who are paid to do so.

4.10.4 Wiki is not suitable for scholars

It is very interesting to note that, on the contrary, all scholars interviewed were contrary to the idea of using wikis as collaborative digital libraries. If for I2, it was very important to “spark enthusiasm” and foster collaboration among other things, for scholars collaboration became secondary, instead of other important features of the system.

Interviewee I5 was very clear:

G. and I were astonished when Marco told us he would like to use MediaWiki for Liber Liber. We come from *a world which demands certain scientific features*. MediaWiki is absolutely fantastic for collaboration, but not for example for metadata.

I3 stated:

I was quite contrary with the idea of moving Liber Liber to a wiki, because it's true that is very useful for collaboration, but it is not so good for managing a collection, for being a repository. I would have liked more to have a wiki and another repository program, as DSpace, because it's standard for metadata management.

Instead, interviewee I3 did not focus on technologic issues, stating that every technical aspects was somehow secondary: more issues need to be

solved before.

I believe the technology platform is a secondary problem, in a certain sense. It is not a matter of choosing wiki technology rather than DSpace or Fedora. Obviously flexibility of wiki software among others is clear.

He insisted on the procedures and the model:

For me it's a question of *procedures* and *work-flow*. Basically, when a philologist or generally a humanist starts working on a text, he tends to consider it *private property*, with no doubt he will be disappointed if other people studied the same text. . .

4.10.5 Collaborative editing and variations

It has been previously mentioned that interpretation is an irreducible feature of Humanities culture and scholar method. Collaborative editing is somehow opposite: the aim is always to find a consensus between different views and opinions. It works on the consensus of a particular community of users: even the version can be changed and edited by users, it is always one. There are no different variations or interpretations of the page.

Actually, this is a issue for Humanities, in which the concept of *interpretation* plays a great role. Each scholar (actually, each reader) can provide an interpretation of a text, and they are all theoretically plausible. Regarding this, Eco stated:

Science is *cumulative-destructive*, it stores what it needs and throw away what it doesn't require. Humanities are totally *cumulative*, they don't throw away anything: in fact, there is always a return to the past.

On the other hand, they are totally destructive in the way, as Maritain stated regarding to Descartes, "a philosopher is a *novice in the Absolute*". For Descartes, everything that philosophy stated before him was false.

Each person, in Humanities, is therefore allowed to provide an interpretation: although of course these interpretations need to be motivated and

there is always a principle of authority, in some fields (as philology) a collaborative system would absolutely provide a system for allowing different variations.

This is not a feature of contemporary wiki (or wiki-like) system. This is an important reason why scholars are skeptic about wikis: these softwares are sometimes considered impersonal and too aimed to reach an neutral consensus.

Yet, from the point of view of the researcher, this issue is merely technical: it should be possible to still have both features, collaborative editing and possibility of variations, in the same system. It does not seem an insurmountable technical problem, and the idea of allowing users to use collaborative editing where possible and still let the possibility of proposing different interpretations and variations should be methodologically and theoretically acceptable by Humanities scholars. In this way, the system should foster collaboration and leave intact the cultural prerogative of the CoP.

So it seems to me absolutely positive that through wiki you get used to the need to verify the source, and the fact that there are some in the academic resistance is not a good sign, but the academic community. People should be more stringent with regard to information, but instead attached to a show how old information and wrong, and before you get used to that Wikipedia is a source as much as the "sacred manual" of their discipline, the better. Because even there, for sure, there will be mistakes, however small.

4.11 A Collaboratory Digital Library model

During the conversation, interviewee I3 said an enlightning sentence:

The first image I would suggest of a collaborative digital library is an interface with several tabs, with the text, the quotations, the critical literature, the references, the didactic uses . . . I guess these are the layers where it could be interesting working on.

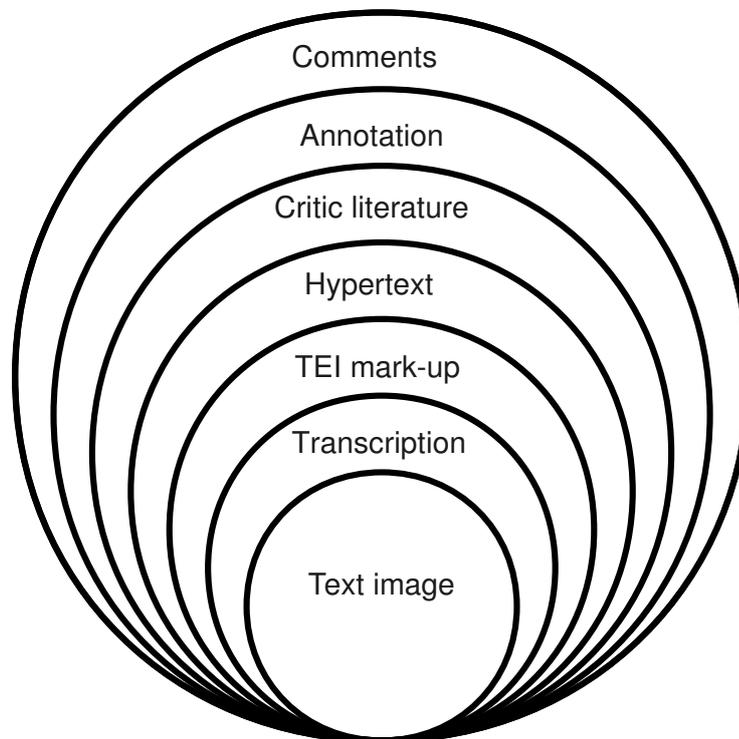


Figure 4.1: Collaboratory Digital Library model

This statement suggests, in a nutshell, the most meaningful outcome of the present study: the idea that a collaboratory digital library is possible, but it needs to be structured in typified *layers*, each one dedicated for a defined task with its own level of collaboration.

This type of design tries to solve several issues. Indeed, it has been explained that authorship and collaboration do have several issues. Although some data suggests that collaborative editing is suitable for certain tasks, several other clues indicate that a strong collaboration system would not easily accepted by the community. Not all tasks and tools are meant to be social, or at least not all of them can be based on collaborative editing. Therefore, tasks demands to be discriminate by their aim and the possible degree of collaboration.

4.11.1 Layers

Every layer of the model represents a different service or feature, with its own level of collaboration. Creating a hierarchy of feature discriminating them by the level of sharing and collaboration is the main result of the present study. The researcher discussed largely with interviewees about collaboration and specifically collaboration in Humanities, concluding that only discriminating different actions and activities could be the key of engaging humanists in collaborative projects. Each layer should be *optional*, *repeatable* and *independent* from the other, to let each community decide whether or not use, add, or eliminate tasks. In fact, each community should auto-organize and adapt the models to his needs and aims.

In the draft concept here proposed, the Collaboratory Digital Library Model (CDLM) is structured like an *onion*. The first layer, level 0, could be an image of the document in which the community is interested. If the document need to be transcribed, another layer could be added, providing a wiki-like system for transcription of the text. This level should allow both collaborative editing and different variations to show different interpretations of the text.

Then a TEI mark-up would be added, utilizing stand-off mark-up to leave the transcription as plain and static text. Moreover, an hypertext could be another layer, as well as critic literature, comments, interpretations, public and private annotation.

4.11.2 From static to dynamic

This passage from the interview of *I5* will enlighten some clues of the model:

Dino Buzzetti theorized that a text had both a dynamical and static aspect: one fixed and one mobile. Either you have textual variations and the text is mobile, or you have a fixed text (maybe critically established), and upon this you apply the dynamism. He viewed the text as the fixed part, upon which you can apply the dynamic part of interpretation.

To establish a the fixed part of the text (i.e. transcription of a printed edition) could be then a task easily done in collaboration (transcribing a manuscript is much harder, demands different interpretations). Interviewees on average agreed on this. Though I4 argued that establishing a text it is not the most interesting part to do collaboratively:

[I imagine] an “onion” with typified layers. The text is the level 0 of the onion, but is also the one with less issues. If we need to collaborate to establish a text, well, when it’s done, you just need to fix typos. Thus, with a single edition the job is done soon. For example, with Liber Liber we digitized the Petrocchi’s critical edition of the *Divine Comedy*: that is now an established text, probably there are no more typos. We are reasonably sure that the paper edition and the digital edition coincide.

Nonetheless, that was one particular edition, there are several others. What is not finished, and will never be completed, the whole environment of the references, quotations, critical literature, uses. This is what happened with the American Dante project: they collected the different editions, then they moved to the critics and comments. In my opinion, this the true field where to work collaboratively.

Furthermore, I5 confirmed the never-ending approach of philology:

Approach to text is not one-sided, there are many as you wish. It depends on the text, different disciplines, and from what you are interested into. In a philology conference, it is very hard to find two philologists agreed. *Text is liquid, dynamic*. This is the reason it is needed to start with images, that are more objective, and then going further to the top, creating different interpretations, different ontologies, ecc. In an electronic archive, you can collect several different editions of the same text, as well as critics, comments [...]. You can make more people agree.

4.11.3 Hypertextuality

For interviewee *I4*, hypertextuality was a bit of a issue. The researcher provided him the example of wikis, where in the normal text it is possible to easily create hyperlinks (i.e. blue words in Wikipedia). In a wiki digital library as Wikisource, this system is exploited to link quotations and cited author and texts to the original quotations, author and texts.³² :

I believe that the approach of having an *hypertext* directly on the textual layer is naive. It is related to a common thought, which is in my opinion wrong, which believes that *hypertext would need to make intertextuality shallow*, evident. It believes that a text itself *is rich of implicit links, and the hypertext makes them explicit*.

He continues arguing:

The problem is that intertextuality is potentially never-ending, also in the very same hypertexts. If people is given the opportunity of insert link in a plain text, these links are potentially infinite, and therefore it would be *technically impossible*, because if links are not *typified*, the very same word could generate different links, different paths. . .

The only way I see is to use different layers and *typify* those layers: one layer dedicated to direct citations, another for back-links, etc.

4.11.4 TEI mark-up

The TEI, meaning “Text Encoding Initiative”, is an international organization founded in 1987 to develop guidelines for encoding machine-readable texts in the humanities and social sciences (Text Encoding Initiative, 2007a).

³²The issue of hypertext is a broad and complex topic that could not be developed by the present study. However, it will suffice to say that hypertext is one of the common and best metaphor of the net itself, and the whole world of digital libraries has taken his roots from the first experiments and conceptualization of hypertext. For further research, see Ridi (2007).

The TEI Guidelines “define and document a markup language for representing the structural, renditional, and conceptual features of texts. They focus on the encoding of documents in the humanities and social sciences, and in particular on the representation of primary source materials for research and analysis” (Text Encoding Initiative, 2007b). TEI guidelines are thus an useful instrument for researchers, who can develop customized languages for their own texts and digital libraries.

Some interviewees were interested in the possibility of collaborative TEI mark-up: collaborative editing is very suitable for textual tasks as marking up a document, and several project, from Wikipedia to PlanetMath.org³³, do involve wikis and wiki-like system for text mark-up.

Interviewee I4 saw the TEI as a very useful layer of the collaborative digital library:

Marking-up a text in TEI could be useful, it could be a very functional and helpful layer, and too it could be done collaboratively. On this mark-up, you could have other services, as data mining tools, etc.

4.11.5 Other layers

Beyond tools that focus tools aimed to work on texts, there several other tasks that can be made *around* them. Interviewee I4 stated:

There is the collaboration *in establishing a specific text*, which it does not exist yet: it is particular, there are some specific issues. [...] On the other hand, [there is the collaboration] *around* a text, which it does exist [...] Around that text, which collaborative tools could be provided? [We could] gather critic texts, different interpretations, quotations, different uses of that same text in different contexts (maybe academic or didactic).

He continues:

What is not finished, and will never be completed, the whole environment of the references, quotations, critical literature, uses.

³³<http://planetmath.org>

This is what happened with the American Dante project: they collected the different editions, then they moved to the critics and comments. In my opinion, this the true field where to work collaboratively.

Therefore, the next level of collaboration is around texts, maybe providing a critic literature framework for researchers, or collecting shared annotations and comments. In the CDLM, these tasks has been represented as higher levels. Moreover, these layers are also the less “objective” and more influenced by personal interpretations and opinions. Which means that are tasks less suited for hard collaboration, or collaborative editing. In this case, the collaboration shift on communication, with the difference previously mentioned: from *collaboration* to *social*.

4.12 Open

There is a deep connection between collaboration and sharing. It is beyond the scope of the present research to investigate these connections: however, it must be emphasized that collaboration is a form of sharing, and openness is a crucial factors for both of them. Highly collaborative projects demand often an high degree of openness, and the other way around. This little result, which needs further studies to be validated, was found during data collection, investigating issues related to strong collaboration. Interviewees were asked to express an opinion upon relationship between openness and collaboration.

Interviewee I3 enlightened the issue of openness in Humanities, directly related with the issue of authorship and intellectual property:

This is complicated, I admit that I’m a bit more conservative about this. In Humanities a scholar writes and his byproducts what is its intellectual products, research, is that it produces.

Within science, there are patents, technologies, tools are developed. Moreover, apart from rare exceptions, no theory is constructed on the singular individual, but more on past research. [Technological and scientific] innovations innovate only 5 percent

of earlier theories. Contributions in these theories are multiple and minimal.

For us [in Humanities] it is a bit more complicated, because the *interpretation* has a value in itself. So I understand the issue of protecting even the letter of what I write, so no one appropriates it to write his text. There is an authorship problem, which is important and must be addressed. This also does not cover the speech open license and permit closed, but the recognition of intellectual authorship.

Internet in fact undermines the effective protection of intellectual property, and people copy. [...] The issue of *economic* rights is almost *insignificant*, there are virtually no revenue for humanistic work. I am convinced that the protection of open content, along with the protection for those who first expressed a particular idea in a particular form is a goal to pursuit.

According to him, there is also the obstacle of the “proprietary” mentality of institution and singular scholars:

There is then the problem of institution, which tend to consider everything as *private property*. The issue of Open Access and Open archives in Humanities is desolating: in Humanities, no ones put there anything of their own.

References

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43.
- Boyd, D., & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer Mediated Communication*, 13(1), 210.
- Bush, V. (1945). As we may think. *The Atlantic Monthly*, 176(1), 101–108.
- Engelbart, D. (1988). A conceptual framework for the augmentation of man’s intellect. *Computer-supported cooperative work: a book of readings*, 36–65.

- Guedn, J. (2009). *Open Access*. National Science Foundation.
- Koblas, J. (2006). *Wikis: tools for information work and collaboration*. Oxford: Chandos.
- Kouzes, R., Myers, J., & Wulf, W. (1996). Collaboratories: Doing science on the Internet. *Computer*, 29(8), 40–46.
- Kurzweil, R. (2005). *The singularity is near: when humans transcend biology*. New York: Viking Adult.
- Licklider, J. (1965). *Libraries of the future*. Cambridge: The MIT Press.
- Lih, A. (2009). *The Wikipedia revolution: how a bunch of nobodies created the world's greatest encyclopedia*. New York: Hyperion.
- List of wikipedias*. (2010, May 12). Meta. (Retrieved 19:04, June 6, 2010 from http://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=1969133)
- Marshall, C. C., & Shipman, F. M. (2003). Which semantic web? In *Hypertext '03: Proceedings of the fourteenth acm conference on hypertext and hypermedia* (pp. 57–66). New York, NY, USA: ACM.
- Ridi, R. (2007). *La biblioteca come ipertesto*. Milano: Editrice Bibliografica.
- Text Encoding Initiative. (2007a). *TEI: Frequently Asked Questions*. (Retrieved 19:49, June 5, 2010 from http://www.tei-c.org/About/faq.xml#body.1_div.1_div.1)
- Text Encoding Initiative. (2007b). *TEI: Guidelines*. (Retrieved 17:49, June 10, 2010 from <http://www.tei-c.org/Guidelines/>)
- Wenger, E. (2006). *Communities of practice: a brief introduction*. (Retrieved 17:03, June 5, 2010 from <http://www.ewenger.com/theory/>)
- Wikipedia. (2010a, June 5). *Moore's law*. (Retrieved 17:03, June 5, 2010 from http://en.wikipedia.org/w/index.php?title=Moore%27s_law&oldid=366166522.)
- Wikipedia. (2010b, June 16). *Neutral point of view*. (Retrieved 18:35, June 15, 2010 from http://http://en.wikipedia.org/w/index.php?title=Wikipedia:Neutral_point_of_view&oldid=368361231)
- Wikipedia. (2010c, May 26). *Project Gutenberg*. (Retrieved 17:34, June 5, 2010 from http://en.wikipedia.org/w/index.php?title=Project_Gutenberg&oldid=364370296)
- Wikipedia. (2010, May 16). *Publish or perish*. (Retrieved 16:57, June 9, 2010

- from http://en.wikipedia.org/w/index.php?title=Publish_or_perish&oldid=362454397)
- Wikipedia. (2010a, June 14). *Semantic web*. (Retrieved 19:35, June 15, 2010 from http://en.wikipedia.org/w/index.php?title=Semantic_Web&oldid=367996192)
- Wikipedia. (2010b, May 12). *Wikisource*. (Retrieved 18:53, June 6, 2010 from <http://en.wikipedia.org/w/index.php?title=Wikisource&oldid=361691031>)
- Wilkinson, D., & Huberman, B. (2008). Assessing the value of cooperation in Wikipedia. *Arxiv preprint cs/0702140*.

Chapter 5

Conclusions and recommendations

The global aim of the present study was to understand if a collaboratory digital library could be helpful for the Humanities community in the Italian context. Research questions were declared in a sequence to go deeper in details regarding functionalities and structure of such digital libraries. They were:

H1: Can a digital library for the Humanities community within the Italian context be *collaborative*?

H2: Can a digital library for the Humanities community within the Italian context be a *collaboratory*?

H3: Can a digital library for the Humanities community within the Italian context be built upon a *wiki*?

and objectives were:

- to analyze perception of Humanities scholars (particularly of Digital Humanities) towards collaboration and digital collaborative projects
- to investigate wikis to be a possible framework for a collaboratory digital library

The first two questions, as previously explained, are very similar: they did center the main topic of the study, that is collaboration in a scholar online community of practice. Both focus on exploring willingness of the community of scholars to join (or just imagine) a collaborative project aimed to research and study. First question is more general, second directly refer to the e-science and cyberinfrastructure trend that is finally spreading in the academic world. Answer to these two questions, as expected, is *positive*.

Interviewees did in fact support the core hypothesis of the study, that is the importance of collaborative projects in scholar communities, even in the Humanities (and Italian) context. Though, they did raise many issues and barriers unforeseen by the researcher.

5.0.1 Wiki as a digital library

Another objective of the research was to discover if a collaborative digital library could be built upon a *wiki*. Wiki has been proposed as a well-known and widely used framework for collaboration, being utilized by online volunteer communities as well as business organizations. The researcher himself is an active user in a wiki digital library (Wikisource), and decided to test the willingness of scholars to utilize such an instrument.

At the beginning of the study, the main idea was to propose the wiki to the scholar community as a collaborative framework.

It must be admitted that the perspective of the research was very technology-driven. The idea was that, if wiki systems had proven to be a valid framework for huge and complex collaborative projects, thus they should have been suitable also for Humanities community, particularly in the Italian context.

Eventually, the perspective has been completely *reversed*: during the research process, the researcher gradually changed his mind and understood that his approach was wrong.

Applying research methods and following the procedures did teach the researcher that, instead of focusing on the technical aspects and the software interface, the right way to reach meaningful results was to *directly ask the community*. There has been quite discussions with the advisor on these topics, and it must be admitted that the initial approach of the researcher, focusing mainly on technical aspects and on the proposal of the wiki, would have forced analysis and results.

Thus, contrarily to what expected at the beginning, data suggests that answer to this third research question is **negative**. Or, at least, the outcome is negative intending *wikis as softwares*: scholars and researchers demand more complex and modular systems able to support advanced Semantic Web technologies and with a strong attention to metadata and protocols as OAI-PMH or OAI-ORE. Nowadays this is not available for most wiki engines, that are more basic and that focus more on other features.

On the other hand, the answer can be evaluated as **positive** intending wiki as a synonym of *collaborative editing*, a strong form of collaboration

that could really enhance research in Humanities. In this sense, intending more broadly the term wiki, as an analog of “wiki-like”), answer can be regarded as **positive**. Moreover, the growth of Wikisource and the fact that a project like Liber Liber is moving on MediaWiki, suggest that, if not for scholars projects, wikis are suitable for amateur projects. Nevertheless, this outcome need to be validated by further and more focused research.

The following paragraphs will illustrate conclusions and recommendations suggested by the researcher.

5.1 Different types of digital libraries

Data gathered from interviews and literature suggested a distinction between two main types of digital libraries perceived by scholars. In fact participants shared a vision where digital libraries are intended either as *repositories* or *Virtual Research Environments*.

Interviewee I4 declared:

Historical projects (regarding DLs) has been of these two kinds:

- *repositories* in which to store digitized texts, in various ambit: some academic, some other amateur (like the Gutenberg Project¹ or Liber Liber²), some other institutional. Often [the repositories were built] without thinking at the aims they could accomplish.
- the *scientific-academic digital library*, which tries to provide users with tools for working on texts. Often [these tools consists] in linguistic analysis tools, which are quantitative, like concordances, frequencies, etc. Sometimes, user can search on lemmatized texts, for example in the Perseus project.

Digital libraries are often intended as repositories where to store, preserve and provide access to entire collections of documents, either born-digital or

¹<http://www.gutenberg.org>

²<http://www.liberliber.it>

digitized. These projects do follow a traditional vision of digital libraries that is not far from paper libraries: they focus on entire collections, and treat texts as static objects. On the other hand, VRE are much more innovative projects created for scholars and researchers, especially in the field of Digital Humanities. These digital libraries provide many tools (often, Semantic Web technologies) for research and treat texts as dynamical objects to research and investigate. Interestingly enough, this vision of DLs is much more close to the early pioneers (visionary) work of Bush, Licklider, Engelbart, Busa.

This perception was confirmed by other interviews, and **table 4.1** was completed collecting features from interviews.

5.2 Importance of collaboration

Literature and interviews both confirmed the increasing importance of collaboration in digital libraries. As e-science and cyberinfrastructure are growing, Humanities are exploring the same approach to a strongly collaborative work in digital environments. All interviewees, while having different opinions, confirmed the fact that collaboration is evaluated as a positive factor for research and innovation, and so online collaboration in apposite frameworks. Interviewee I4 affirmed:

The topic of collaboration is increasing every day; in electronic editions, it is the *keyword*.

while interviewee I1 too was asked if collaborative frameworks could help the field of Humanities:

Absolutely. I believe this is the only way for a science which needs to innovate itself.

For the scope of the study the concept of *collaborative editing* has been chosen to represent the ideal and strongest form of collaboration involved in a collaborative project. In fact, in the context of digital libraries the highest form of collaboration and sharing seemed to be the possibility to edit collaboratively either a text, a page or even a single line.

Thus, collaboration in collaborative project is in fact intended as collective editing.

Interviewee I3 upheld the importance of collaborative editing for digital libraries:

Another aspect, more “borderline”, regarding research, is the one which involves the **collaborative editing**. One of the most complicated in the creation of quality archives is to guarantee the quality of both the text and possible mark up of the text itself.

More specifically, the creation of collaborative digital environments could have different advantages:

- lower costs (i.e. critic and diplomatic editions)
- introduce the concept of “no ultimate version” within Humanities
- foster the neutralization process within philology and literary critics
- enhance collaboration at many levels (i.e between scholars, subdisciplines even scientific fields)
- innovate a traditional field as Humanities in the Italian context

Interviews confirmed that collaborative editing could lower costs of digitization and DLs, and also stimulate discussion and collaboration among various disciplines. Interviewee I3 stated:

I rather think that the opportunities given interoperability (which can be seen as collaboration among different softwares) and collaboration (among different people), are very important. They can lower costs, especially in the Humanities, where critical editions can cost 20 years [of work] and lots of money and even doubts about funding. Collaborative work among scholars, even from different disciplines, could solve these problems.

Moreover, collaborative editing somehow questions the concept of an “ultimate version” of a text. I4 stated:

What is more, without having the presumption to aim at the “best edition” (actually, we can just dump this idea). [There

could be projects where] philologists of different nationalities cooperate and suggest interpretations. This would be a brand new opportunity to change scientific cooperation in the philology area.

This is due to the fact that every sentence (namely, every opinion and thought) can be doubted and questioned by other users. Every text (i.e. an encyclopedic article, a didactic book as well as a translation) is the sum of several interpretations, versions, modifications and contributes of different users.

Interviewee I1 noticed that collaborative editing in Wikipedia, struggling for reaching a consensus among users in the name of the NPOV, is incredibly similar to a similar process happening in philology: the process of establishing, assess and evaluate a text.

This is fundamental: this [*the established consensus of a page, NdR*] is *the outcome of a neutralization process*. From the philological point of view, the possibility of tracking this process (through the **H**istory or other more sophisticated tools) is crucial. Because the Neutral Point of View [one of the pillars of Wikipedia] is the result of a process of neutralization. In philology, to track and follow this process would be absolutely important.

This process in wikis (and other collaborative projects) is tracked by the **H**istory of a page: this tool is thus a log, a stratification of the various version of a page, then it is very helpful from the philologist's point of view.

5.3 Community

The present research intended the phrase *community of practice* as defined by Wenger (2006):

Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly.

Although observation, literature and interviews suggest that collaboration is becoming more important, they also suggest that there is a distance between scholar communities of practices, and the general, undiscriminated audience that compose the crowd behind huge collaborative projects like Wikipedia. Scholar communities do in fact agree in considering very importance collaboration among peers, intending people who share the same competences, culture and expertise. There is absolutely no consensus upon *opening* the community or scholar collaborative projects to generic users, wiki-style.

Professor Eco called these two different communities *controlled* and *uncontrollable*; he suggested that a closed community of peers could reach goals and collaborate in a much more helpful and efficient way. Especially, he insisted on the trustworthiness of the collaborative work of such a community, which has inside procedure, policies, cultural aspects that prevents errors and guarantee a control and a (peer) review of the information.

According to him, generalized communities of people do not reach the same level of reliability, though he admits that the collective filter of a “motivated crowd” does work on average.

5.3.1 Quality and reliability

It is thus still very present that quality and authority of peer produced information is depending by the community who produced it. A lot of research has been done regarding the problem of authority and reliability in Wikipedia and other open projects, and more research has to be done to understand and comprehend the phenomenon in its broadness. It is the opinion of the researcher that only the surface of these issues has been scratched. Wikipedia opened a window that we did not expect to exist: it shows us a world where expert are important, but not so much. Crowds of active and good-willing users can flock and creates coherent and complete enciclopedic articles, almost about everything.

This is a complex issue that is very feared, for many reasons, by scholars and academy in general. Their comments and critics are often grounded and very agreeable, but it is still valid the opinion that this whole phenomenon

of crowdsourcing, “wisdom of the crowds” and value of expertise (as well as perverse mechanisms in academy and scientific communities) are yet to be understood.

This is also a reason why the term wiki is so feared: aside technical limits, it reminds to Wikipedia and this new kind of open, indiscriminated communities of practice. Further and specific research is needed to assess these outcomes and investigate in details the conflict between scholar and amateur communities. Other research is then needed to really understand inner mechanisms of both the communities to comprehend differences and similarities, particularly in knowledge production aimed communities. It is the humble opinion of the researcher that these communities are more alike than different, and that amateur communities still have something to offer (and teach) to scholar ones.

5.4 The Italian context

Interviews collected several meaningful data about issues in Humanities within the Italian context. Many of them are related to mentality and culture, as the elder age of the cultural and academic establishment, with consequential lack of interest to innovation. This is a well known problem of Italian politics and institutions: the ruling class is aged, younger generations have extreme difficulties in gaining leading positions and therefore innovation and progress-related issues are not the primary agenda. Academia makes no difference: young professors are rarity and thus they have no power to decide and direct funds on innovative projects, as it could be a collaborative digital library.

Moreover, regarding collaboration and collaborative projects, there are serious issues related to career and tenure, as the fact that digital publishing is not evaluated in tenure competitions. Interviewee I3 states:

Now, there are issues related with tenure and *career*, somehow complicated. In tenure competitions, digital editions and digital publishing do not count. *They are not evaluated.*

This means that all publications needs to be on paper and thus there

is little interest in publishing (and researching) in the digital environment. Obviously, also people interested in these topics are discouraged by the lack of dedicated infrastructure and procedures. This situation also acts as an obstacle to collaborative projects, because there would be much more difficulties in giving right authorship of contributions to every user.

Think about what could happen in a collaborative work, where it is not possible to redeem attribution of every edit. This is a serious problem, that will be maybe solved by younger generations of scholars, who are more aware of importance of sharing.

The solution is then offered by the same researcher: a “rejuvenation” of the ruling class would probably have the outcome of making it more sensitive to innovation and progress-related issues. A practical step could be the evaluation of digital publishing for tenure and scholarship competition. This would foster innovation in research and could help the birth of collaborative projects as a collaboratory digital library.

5.4.1 Italian digital libraries

Towards digital libraries, and the digital environment in general, Italy maintains a conservative approach and quite reluctant to innovation. Few institutional digital libraries have been built in the last decades, and few investments are being made for digitization of cultural heritage. The limited amount of Italian material in projects like Europeana make Italy one of the nation with greater potential and less effort in preserving (and providing access) digitally cultural heritage.

Many interviewees confirmed the fact that amateur projects like Liber Liber and Wikisource are still very important because no other institutional projects actually followed. In fact, Liber Liber was born in 1993 to fill a gap, as I3 recalls:

To be honest, we all thought that the work of Liber Liber was necessary in the first phase, hoping that the digitization would have been further made by large national projects, universities, institutions, in a professional manner, and not voluntary as we did.

Unfortunately it was not so, at least not in Italy.

Obviously, some work has been done, as Biblioteca Digital Italiana, probably the most famous digital library in Italy. Yet, I4 argued:

Then, I also think that Italy lacks institutional digital libraries. Apart from the first experiments of *Biblioteca Digitale Italiana*, (which was very professional, TEI marked-up, but actually really small).

Aside from that in Italy there is *nothing* [...].

There's lack of an adequate culture: when they started digitizing in Italy they started from historic bibliographic catalogs. There were beautiful bibliographic cards, of course, handwritten and from the XVII-XVIII century; but definitely not very attractive to the general audience. Other nations are much smaller and culturally poorer [than Italy], such as Norway, are far ahead of us. It is a matter of approach, not just resources.

Recent news tell that the Italian Government has already subscribed a partnership with Google for the digitization of over a million books and documents in the well-known (and controversial) **Google Books** project. It is the opinion of the researcher that, although it is still too early to evaluate outcomes and results of the mentioned settlement, Italy has so far lost several opportunities to innovate his approach to cultural preservation and access, and that much is to be done to reach the standards in these area of other European countries as France or Germany. Unfortunately, there is little hope that this will be done in the following months, given no sign at all of a change of mentality in both the academy and the government.

5.5 Wikis is for amateurs, not scholars

The third research question, asking if a wiki could be suitable as the framework for collaborative digital libraries, had to unexpected outcomes. In fact, the question had to different answers by two defined communities. Data gathered from scholars clearly suggest that wikis are to basic systems for scholar projects, and even if flexibility and focus on collaboration

are considered as important features, lack of other more technical features (i.e. metadata standard, protocols, etc.) make them unavailable for this advanced community of practice.

On the other hand, although proper data collection could not be applied for amateurs, some data suggest that wikis are actually suitable for amateurs community of practice, especially regarding digital libraries. Wikipedia is a self-explaining example, when Wikisource is a dedicated digital library wiki which is increasing his community day by day, reaching an ever increasing visibility worldwide. In Italy, the Liber Liber project decided an important shift of infrastructure, moving on a wiki system. Interviewee I2, was asked to explain such a decision:

Firstly, to provide [the community] a more efficient instrument for collaborative work than before: [the previous system was in fact] based on a mailing list. The editorial staff used to create the content and subsequently send it to an HTML expert, who had to transfer it on the Web. Though, this created bottlenecks that were unsustainable, given the number of volunteers....

Thus, wiki do serve communities because their enhancing of collaboration and enthusiasm. They are very efficient in smoothing workflow and make it faster (“wiki” in Hawaiian means *quick*). Wikis do work very well in large communities because they eliminate bottlenecks and allow all users to work without hierarchy restrictions. Other data suggest that these amateurs communities do have less limits and expectations. than scholar communities, and wiki perfectly fit their purposes.

5.6 Solving the Authorship problem

Evidently, authorship and collaboration are two faces of the same issue. Solving authorship related problem would foster collaboration, and viceversa.

Chapter 4 extensively provided a picture of issues and sub-issues of collaboration related to authorship, at many different levels (cultural, institutional, political, technical). Thus, many improvements needs to be pursued

at the same levels. Issues were divided in sub-issues as follows:

- *myth of the lonely scholar*
- interpretation
- attribution
- intellectual property

In fact, the great value attributed to the interpretation often make collaboration problematic. Collaborative editing, intended as the strongest form of collaboration in online projects, demands convergence of opinions and consensus. Variations, opinions, different interpretations should be accepted as well. Thus, further study and development is needed to make softwares and collaborative systems adequate for Humanities scholar needs. For example, new technical features are needed in wiki and wiki-like softwares to host different variations, or to allow less strict forms of collaboration. Collaboratories need frameworks which support different complexities and forms of online collaboration and communication.

Regarding institutional and political issues, some advice were given directly by interviewees. I3 suggested the creation of institutions for support and development of e-science and e-research projects.

The issue is providing a institutional framework which will support [these projects]: if there was a national institution for e-science or e-literature, for example, even with limited fundings, that would be the best location for experiments and test.

These institution would act as landmark for communities of practice, who are nowadays alone and disorganized. Coordination would certainly be helpful.

Moreover, political actions could be pursued. Politically, as previously mentioned, an important goal could be to make digital publishing evaluable in tenure and scholarship competitions. This could be a concrete help for scholars and researchers and a boost for those who want to exploit digital environment for research in Humanities.

Technically, importance of logs, history and contributes' traceability was confirmed by several participants. I1 highlighted that, even if these tools

are present right now, they are too basic and need not only to be improved technically, but also a deeper reflection is needed to give the right importance to these issues, truly crucial in Humanities:

I wanted to emphasize the importance of this [authorship matter]. [...] This traceability demands to be also theoretically emphasized, because in crucial, in philology.

5.7 Collaboratory Digital Library model

In Chapter 4 a model for a Collaboratory Digital Library (CDLM) has been suggested. It acts as a sum of the outcomes of the present research, integrating them in a draft, yet constructive proposal. The idea came directly from interviewee I4:

The first image I would suggest of a collaborative digital library is an interface with several tabs, with the text, the quotations, the critical literature, the references, the didactic uses . . . I guess these are the layers where it could be interesting working on.

The CDLM is structured in *layers*, which are divided per tasks, which are also discriminated by their level of collaboration. Task suggested by data gathered were organized as follows:

- Image: the image of a text is intended to be the level 0 of our digital library. All the further philology, ecdotic and librarian work is conducted upon that.
- Transcription: Scholars collaboratively transcribe the image of the text, maybe aided by OCR softwares.
 - Variations: in this layer collaborative editing must admit (methodologically and technologically) variations and different interpretations of the text.
- Mark up: several mark-up languages (i.e. TEI, XML) can be applied on a different layer, utilizing *stand off* mark-up.

- Hyperlink: another possible layer can be the hyperlink layer, in which exploit the other dimensions of the text within the digital environment. It can be used to highlight *intertextuality*.
- Annotation: annotation can be collaborative or not. This layer allows both collaborative editing of notes and private notes.

The first tasks are more objective and focused on consensus: in which thus collaborative editing (with some exceptions) is more useful. They are in fact tasks operated *on* texts, following I4:

There is the collaboration *in establishing a specific text*, which it does not exist yet: it is particular, there are some specific issues. [...] On the other hand, [there is the collaboration] *around* a text, which it does exist [...] Around that text, which collaborative tools could be provided? [We could] gather critic texts, different interpretations, quotations, different uses of that same text in different contexts (maybe academic or didactic).

The next level of collaboration is *around* texts, maybe providing a critic literature framework for researchers, or collecting shared annotations and comments. In the CDLM, these tasks has been represented as higher levels, being less objective and more influenced by personal interpretations and opinions. Which means that are tasks less suited for hard collaboration, or collaborative editing. In this case, the collaboration shift on communication, with the difference previously mentioned: from *collaboration* to *social*.

Such a model should be implemented in open source, allowing people to use, modify, and customize the software and, what is more, to create and modify their own tools for research. This study wanted to emphasize the importance of freedom of reuse and remix of tools: data suggest that such a open and free approach foster collaboration, interoperability and innovation. To be real laboratories the projects have to allow participants a certain degree of freedom, also in modifyng and adapting the environment to the needs of the community. Huge collaborative projects as Wikipedia showed that empowering users help projects adapt and evolute faster, making them

highly flexible and scalable. Much smaller projects, as a collaboratory digital library, would have no problem in applying the same principles within a much more controlled and homogeneous context.

In the end, the present research gathered data from a restricted yet significant sample of Digital Humanities scholars, to investigate their towards collaboration and collaborative e-research projects. The CDLM is here presented as a proposal, summerizing different outcomes of the study: it is hoped that it will be an helpful draft for further discussions and research.

Appendix 1

The following conversation with Umberto Eco happened April 24th 2010, in Milan. Professor Eco welcomed the researcher for an interview on collaboration, Internet and Wikipedia, under the project Wiki@Home, powered by Wikimedia Italy. The Italian transcript of the interview is available online at the URL http://it.wikinews.org/wiki/Intervista_a_Umberto_Eco, while the English translation will be available at the URL http://en.wikinews.org/wiki/Interview_with_Italian_professor_and_writer_Umberto_Eco. The interview is released with a the Creative Commons license CC-BY-SA 3.0³.

The researcher wants to thanks again professor Eco for his availability and kindness.

5.8 Interview

R.: *Thank you very much for the opportunity you gave us. Our community really wanted to interview you, especially since you are one of the very few important exponents of Italian cultural world who approached Wikipedia without bias, describing and criticizing, but nevertheless using it. You wrote several articles about it, the latest in 2009 if I remember well. Could try to re-explain your views about Wikipedia?*

Eco: I am a compulsive user of Wikipedia, also for *arthritic* reasons: the more my back hurts, the more it costs me to get up and go to check the Treccani, so if I may find someone's birthday on Wikipedia it's all the better.

I am a car user and I could not live without it, but this does not prevent me from stating all the defects and the troubles of cars.

I once made a distinction between things good for the *poor* and things good for the *rich*, where rich and poor have no immediate connotation in terms of money, but in terms, say, of cultural evolution... A graduate is a rich, an illiterate is poor. There can obviously be a big entrepreneur who is

³This license allows you to share, remix and use for any purpose the transcription, under the condition of attribute the work to the author and distribute you eventual derivative work with the same license. More information at <http://creativecommons.org/licenses/by-sa/3.0/>.

poor and a little clerk who is rich.

Television thus is good for the poor and bad for the rich: it taught the poor to speak Italian, it is good for old women who sit alone in the house. And it harms the rich because it prevents him to go out and see things more beautiful at the cinema, it narrows his ideas.

The computer in general, and the Internet in particular, is good for the rich and bad for the poor. That is, Wikipedia is good for me, because I am able to find the information I need, I do not trust it, because everyone knows that as Wikipedia grows, *the errors also grow*. I found steep follies written about me, and if no-one pointed me at them, they would stay there still.

The rich are grown people, they can compare the information. I look at Wikipedia in Italian, I'm not sure that the news is correct, then I go to check the English version, then yet another source, and if all three tell me that this gentleman died in 371 AD I begin to believe it.

The poor picks the first data that he gets, and that's all folks. So Wikipedia, as the whole Internet, has the problem of *filtering the news*. It keeps both false and real news; but the rich knows filtering techniques at least for the areas they know how to check. If I have to do a search on Plato, I have no problem in pointing immediately out the sites written by madmen, but if I make a research on stem cells it's not certain that I can identify the wrong sites.

So there's this huge problem of filtering. Collective filtering is useless, since it could yield to fluctuations. I noticed that in a certain period of Berlusconi's triumph people went looking for information about me on the right-winged books and placed them in Wikipedia: as correctness prevents me from changing them directly, I left them on. But obviously it was an entry made by the winners of the moment.

The collective control is therefore useful up to a certain point: it is conceivable that if one gives a false length of the equator, sooner or later someone comes and corrects it, but correction of more subtle and difficult issues is more complicated.

And it seems to me that the internal control is minimal, that is, it cannot control the millions of news flowing in. At most, it can check if a madman

wrote that Napoleon is a racehorse, but there's not too much it can do.

R.: *There is a particular filter. The community organizes itself in groups of people who spontaneously look at the list of **Recent Changes** and try to pay attention to serious errors, vandalisms, people who delete paragraphs, etc. There are various quantitative software tools that help.*

Eco: There's assistance in case of insults, right. But those are the big things.

R.: *You are right, they are the easiest ones. Dealing with subtler things is much more complicated. Some research supports the finding that the more there is a community of people (a group of people) interested in a topic, the better. Indeed, such people save pages in a personal favorite list (*watchlist*, as they are called in Wikipedia jargon). For example, if I have your page in my watchlist I got a report when someone changes it, and I can check what happened: through a *diff* mechanism, a sort of collation, I see what has changed and I can see if it is correct or not.*

The principle of Wikipedia, in a certain sense, is that most people there are, the more they are interested, the better the work is done. This is a bit of a paradox. There has been some research on it, the last I remember was in February 2007, from HP Labs in Palo Alto. It was purely quantitative and statistical, based on the English Wikipedia; it found out that the pages with more changes are on average those with the highest quality. The more people there are, the better it is. Then there's the actual problem of the long tail. There are so many pages just fairly important or problematic or debatable. The page about you, for example, may fall into this set of pages, and besides it's a biography. Biographies of living are the most problematic, because of recentism (adding useless tidbits which happened just now), identification of the sources, etc. All biographies are generally a problem, although in the case of historical figures there is more agreement. I find it interesting to look at the pages which in theory should be the most problematic, with issues such as creationism, intelligent design. In the English language Wikipedia they are abysmally log, because people often quarrel not just over whole paragraphs, but over individual words, or the starting sentence. It's better to have more eyes, like the "wisdom of crowds" theory by Surowiecki, that says that when there are 4 parameters (independence, diversity of opinion, aggregation, de-

centralization), *on average, judging by a crowd is better than that of the experts.*

Eco: I don't quite agree with this. I am a disciple of Peirce, who argues that scientific truths are, ultimately, approved by the community. The slow work of the community, through revisions and errors, as he put it in the nineteenth century, carries out "the torch of truth". The problem is the definition of truth.

If I were forced to replace "truth" with "crowd", I would not agree. If you make a statistical analysis of the 6 billion inhabitants of the globe, the majority believes that the Sun revolves around the Earth, there's nothing to do. The crowd would be prepared to justify the wrong answer. This also happens in a democracy: we are noticing it these days, the crowd votes for Bossi. To carry on his coup d'tat, Napoleon III broadened out the electorate and included the peasants, because the crowd of the countryside was more reactionary than crowd of the cities.

We must therefore find another criterion, which I think is the *motivated crowds*. People who work on Wikipedia are not just an aristocracy, just professors, but they are not the indiscriminate crowd either: they are the part of the crowd who feels motivated to work with Wikipedia. Here it is: I'd replace the theory of "wisdom of crowds" with the theory of the "wisdom of the motivated crowds." The general crowd says that we should not pay taxes, the motivated crowd says that it's fair to pay them. In fact, it is not a digger or an illiterate who contributes on Wikipedia, but someone who already belongs to a cultural crowd for the very fact of using a computer.

R.: *I found very interesting the sentence you quoted from Peirce, which you used also in an article^A from the newspaper "Unit". You said you should switch the word "truth" with "crowd"; I'd rather think about definition of "community". Who's the "community"?*

Eco: Peirce was thinking at the scientific community, of course; especially at his time, it was definitely separated from the common crowd.

R.: *I think that in a world like ours it's more difficult to discriminate between "real" experts: often a degree does not make a real expert from an*

^A<http://espresso.repubblica.it/dettaglio/ho-sposato-wikipedia/2108845/18>

expert. Maybe there are some hardcore fans who have more expertise on a topic than the alleged experts. There has been on Wikipedia, at least at its beginning, a shift of the meaning of authoritative, from mighty to influential. Wikipedia community is much “flatter” compared with the academic or scientific hierarchy, there is no hierarchical authority like we are used to. What “you” say at one time in a given context, and everything you said before, vouches for “your” authority. Many times people listen to the content of a message, without looking at the sender. In such a sense, this can lead to a different mechanism that carries on the torch of truth. A mechanism where we look at the information, at the message, not at those who sent it. Even in science and the scientific community there are fashions, forcing (think at our barons): in fact, we study the sociology of science. In this sense, Wikipedia perhaps led, or perhaps gave some glimmer of something new. I do not know if you agree with this.

Eco: Wikipedia has two unrelated functions in my opinion. The first one is to allow fast search of information, and it is just the multiplication of *Garzantine* [a popular Italian series of compact encyclopaediae], period. The other, and this is what we are talking about now, is whether the control from below can be many times more successful than the control from above. Since the world is full of expert idiots, certainly it can be.

Just an example: some days ago I was correcting an essay about Benedetto Croce. Croce, building on its authority, spread false ideas for 50 years in Italy, and everyone in Italy have accepted them, without considering that he knew nothing about art. He was the aesthetics master for two or three generations without having understood anything about art. So you see, sometimes the authority... Responses of artists, children, students would have been really more useful. This control by the mass can, produce a development in the long run, as Peirce said.

But I keep saying that I am increasingly exposed to the risk of my inability to filter the news. Lately I started writing down some false information, some errors that one can find in Wikipedia. In the same article, for example, there were two contradictory reports, a sign that there had been an amalgam.

R.: *So you don't edit the pages of Wikipedia?*

Eco: Not in that case. I don't edit pages, except for the page about me when I found it written that I married the daughter of my publisher, since *as a matter of fact* I didn't. Poor soul, she ran such a risk! *[laughs, N.d.R]* Another time I was indicated as the eldest of 13 brothers.

R.: *That was your father, right?*

Eco: Yes. If the error is of another person, I don't see why I should waste my time to correct it. I am not the Red Cross. *[laughs, N.d.R]*

Thus, I actually noticed that there was a contradiction, within the same article. The problem is that I'm good, I can notice the error, because that's my job; another person, less competent, could read just half of it and take the first version.

R.: *I just wanted to understand if you did not edit and correct errors for a matter of time, or if there was another reason. Maybe you did not want to be recognized...*

Eco: Of course, it's a matter of time. When I write, I go on Wikipedia 30-40 times a day, because it is really helpful. When I write, I don't remember if someone was born in the VI century or the VII; or maybe how many n are in "Goldmann"... Just few years ago, for this kind of things you could waste do much time. Nowadays, with Wikipedia and Babylon, which checks the spelling, you can save it a lot.

R.: *According to you, how much is the wiki model "exportable"? Here the term wiki is indicating a strong collaboration, in which there is collaborative editing. That means having a text that is collectively edited. Furthermore, this implies often the text itself to be free, meaning released with a free license. This mechanism obviously is related to several issues: the issue of the filter, the issue of bottom up vs top-down process, as well as the issue of having a community of peers with different values and motivations than a scholar community.*

The wiki world developed different projects: for example, Wikiquote, a free quote compendium (many are yours), or Wikisource, a wiki digital library.

In your opinion, this experience is exportable to other writing mechanisms not aimed to collective knowledge production, as Wikipedia is? Wikipedia, in fact, was born as an encyclopaedia developed within a wiki, a specific soft-

ware, and has been a great success, against all odds. They tried several times to build similar projects: the *Los Angeles Times*, once, tried to aggregate collective editorials, it was a failure. Thus, it seems that some projects can be collaboratively developed, meanwhile some others don't.

Eco: You are now talking about collective collaboration. Well, there are few things tha Internet provides: the first are mere data, as the train schedule that no one can correct. The other, are encyclopaedic information, that can always be corrected, because the author could be wrong or simply haven't said all yet. The third, are texts: should I edit other's texts? Moreover, there is the whole universe of blogs and Facebook; but it doesn't matter right now, they are people talking to each other, conversate.

In these very days I had to debate on Ipazia: I looked for some information on Internet, and I found interesting and less interesting texts. But they are texts. Internet provides us classical and contemporary texts, but if they are wrong or I do not agree with them I surely don't edit them. I cannot say "Your opinion on Aristotle is wrong".

R.: *Sure. So I correct myself: someone's interpretation is his own, nad it has a value in its own.*

Eco: More, it is signed. In fact I found many interesting documents that are not signed, I never understood why.

R.: *I had this idea, regarding texts. It is what in the scholar environment are called collaboratory digital library, namely digital libraries for philologists, Italian and Middle age studies scholars. . .*

Eco: Do you mean bibliographies?

R.: *Not necessarily. For example, I was thinking at the Perseus Project⁵, a Tufts University project. The scholar community is provided with tools to work on ancient Greek texts, as linguistic analysis tools, collations, statistical analysis. A project where people collaborate together for a critic edition. . .*

Eco: I stumbled upon some of them. Actually regarding Ipazia, I found a project where different scholars collaborate to translate a text from Xth century.

R.: *What is your opinion about scholar collaboration in humanities?*

⁵<http://www.perseus.tufts.edu/hopper/>

Eco: This is another topic yet. Congresses were made by ecdotic scholars to investigate this topic. These are truly auto-controlled communities.

R.: *Communities of practice.*

Eco: Yes, but where we know that a single scholar belongs to a single university, we know where he comes from. In this case, it happens something similar to when people used to collaborate in writing a book, and they needed to take the train once a week to meet and discuss. It is a collaborative team work that is controlled by someone. It is not the wisdom of the crowds. It is simply the scaling and the simplification of a collective research work that once required filthy travels and nowadays it can be done online daily... I'd rather call them uncontrollable and controlled communities.

R.: *It is very interesting when these controlled communities (which are granted and comes from determined institutions) do not have a hierarchy or a chief, but they auto-control themselves. For examples, I was thinking at a project for the Italian studies community, which could be granted from institutions and still let the community free to auto-control itself. In your opinion, is this auto-organization possible also in these scholar communities?*

Eco: I recall a conference in Bologna, about ecdotic studies, that was dedicated mainly to digital humanities projects and text research environments and functionalities. Evidently, this was an headless community, auto-controlled and headless. But "headless" is a phrase: because in scientific communities which self-legitimate there's always someone who gains more authority: if an important philologist propose an interpretation, the others will follow.

Therefore [online collaboration within scholar projects] it is not the same thing of Wikipedia.

R.: *So you make it a question of limits, that communities more organized and communities less organized exist. The difference of a wikipedian community is to have both univeristy professors and teenagers inside.*

Eco: Let's take as example the magazine *Nature*. In the scientific world, if a paper appears on *Nature* - where a *peer review* has been carried out and there's a wide control - it's taken seriously. It's anyway possible that *Nature* can make a mistake and reject a brilliant paper: nonetheless *Nature*

is considered a center of reliability, with fringed boundaries. Because an error, or a small academic revenge, can always happen. . .

Now take me as example: with my age and my body overweight I entered the high-glycaemia phase of a type II diabetes. Once, the limit for defining glycaemia “high” was 140, today it’s 110: we all know that this new limit has been set by the pharmaceutical companies for selling their products. So, 140 is risky, maybe 110 is too low, one can get along with, say, 120. Maybe in a decade the limit will be adjusted to 120, or they’ll decide that 110 is good in terms of preventive medicine. *[he laughs]* We realized that the swine flu was partly rubbish, spiced up the vaccine manufacturers. We realized it late, billions have been spent, we realized that much less people than expected died, that they maybe overstated it.

In a way or another things fall into place: these are the controlled communities, not anarchical, but with a fringed authority. That have nothing to do with Wikipedia, where the anarchy is bigger.

R.: *This difference of fringes, limits and, in a sense, also of scale is very interesting. In a type of community the collaboration is truly anarchical, in the other there’s an adjustment. . .*

Eco: There is an adjustment. Galileo, Tycho Brahe and Keplero at the end agreed that Keplero was right. The infinitesimal calculus has been discovered both by Newton and Leibniz but at the end everybody agreed for Leibniz. *[he laughs]*

They might have been the wrong choices, but they’ve been made that way.

There wasn’t any authority, the emperor, who decided it. It’s been a collection of habits and applications.

R.: *In your opinion there is difference between hard and soft sciences in approach towards collaboration?*

Eco: Right now, yes, We all know this. In hard sciences there is measurability of data that is absent in soft sciences, unless soft sciences are a parody of hard sciences, as in analytic philosophy.

R.: *Talking about the collaboration, before, you said “this is fabulously interesting, but not surprising.”*

Eco: Sure. The Accademia del Cimento began first! And without the

Internet. *[he laughs]*

R.: *But now the scale is different.*

Eco: First there was a few of them in Florence, then a few more at the Royal Society; now it's a crowd.

R.: *A crowd that can collaborate with Thai people, with American people, in a synchronous or asynchronous way, in an ubiquitous place like the internet. The possibilities are different. Back to the previous point, in Wikipedia also we can notice a cultural difference between the articles about technology, science, maths, physics and the articles about humanistic topics. Humanistic articles are much less (philosophy, history, literature). This in Wikipedia. Within the academic communities, in a similar way, there's a different impulse to the collaboration. In the "soft" sciences, the authorship, the authoritativeness and even the interpretation, matter more.*

Eco: For what are soft sciences, there is absolutely less impulse to collaboration. There is much more interest to be the main character of an idea, than being just a "water carrier".

That's for sure. A scientist in these cases is used to not being mentioned and to know that however is carrying forward a fundamental research. In soft sciences, this happens only to the exploited student who is sent to gather data that the professor will sign and profit by.

That's an old story, there's no escape from that...

R.: *It would be important to understand if this is a "natural" or cultural consequence. Could this Humanities approach really change?*

Eco: I don't believe so. Think about ancient Greece. Plato and Aristotle, being one the other's disciple, developed two opposite philosophies. On the other hand, Euclid came and it is still discussed, his fifth postulate survived for two thousand years.

R.: *Thus it's a natural issue. Right?*

Eco: Science is *cumulative-destructive*, it stores what it needs and throw away what it doesn't require. Humanities are totally cumulative, they don't throw away anything: in fact, there is always a return to the past.

On the other hand, they are totally destructive in the way, as Maritain stated regarding to Descartes, "a philosopher is a novice in the Absolute". For Descartes, everything that philosophy stated before him was false. If a

mathematician did that, it would be the end of mathematics.

R.: *Back to the theme of the “strongly collaborative” projects, where there’s a collaborative editing, what do you think of the authorship, of recognizing the intellectual property? In volunteer-based projects like Wikipedia, the problem matters less, but given that the scientific world is moving toward a more and more intense collaboration (and the humanistic world as well, although more slowly) we have to face the fundamental question of the copyright. In Wikipedia it’s been solved by adopting free licenses, and the culture of the nicknames – or no names – helps; in the academic and scientific world the culture of the name, related to important things such as a personal career, leads us to a complex problem of recognition of the intellectual property.*

Eco: This is certainly coming out, also in the world of books; I think that in 50 years we’ll have a very deep mutation. We’ll probably have a cultural situation similar to the one in the Middle Ages, where comments and comments were produced, and the authoriality was lost. Then, from the Romanticism on, the authoriality became excessive. But I cannot say up to which point we can reach a total anonymity. Although it can look democratic, the total anonymity gives the idea that just one and only truth exists. Can we have a moment in a future where Wikipedia itself, on certain articles (not the one about the pitagoric table, of course), can open sections called “Conflicts” where – signed – can appear different thesis in opposition? In spite of the always present denying madman, we’re certain that Napoleon died in Saint Helens. That Pius XII did the right thing during the Holocaust, it’s an open debate. What does Wikipedia do? It says that Pius XII did not enough (irritating millions of catholics)? It says he did (irritating millions of non-believers)? Or does it open an appendix in which a series of authors, each assuming responsibility for their words, expose in twenty rows the conflict of interpretation?

R.: *(Unfortunately without the internet we cannot check the article about Pius XII and the Holocaust), Wikipedia, that is following the principle of the neutral point of view, which is not the truth, but an unbiased point of view that can always be perfected, usually publishes a version of an article that includes the critics to that version. Following a principle of synthesis there*

are – unsigned – sections that might embody what you just said. As example, the page about Silvio Berlusconi is problematic.

Eco: I never went to see it.

R.: *It's almost always semi-protected, being it often theathre of quarrels. Anyway, we try to report both the positions. Of course there's an hierarachy: there's always a dominant postition, introduced first, but followed by the second one. An example coming to my mind is the article about Beppe Grillo, where there's his biography and then a section of critics, with the sources duly cited. The aim, then, is to report and synthetize on a page (or more, should it grow too much) what other people said. Wikipedia integrates, it's a being that feeds from the outside, because it's a teritiary source, not a primary source, and this is often forgotten. I don't know whether a traditional encyclopaedia defines itself as primary or tertiary. For Wikipedia the sources are somewhere else, we just take from them. We cite them; if they say wrong things, we just cited them. There's always bias in the choice of what to cite, where to cite and how to cite but, at the end, Wikipedia tries to report – with all its limits – the reality as faceted by the other sources. The problem of the hierarchy of the page is still present, and the fact that “there are no fact, but only interpretations”. In this sens there's a very evident and aware temporariness.*

Eco: Of course, everything can change tomorrow.

R.: *Wikipedia guidelines say that “A final version does not exist”. An article is always amendable. Culturally, maybe a Treccani doesn't see its work this way.*

Eco: No, because Treccani has very signed articles. The article “Fascismo”, written by Gentile, cannot be modified; either you delete it and substitute it, or you leave it like it is.

R.: *Article that, besides, was affected by its context: today, after seventy years, we'd say that the Treccani was biased.*

Eco: Yes, that's why it remains and is not modifiable, because that is the article and is not reprinted. They create an appendix, of course. The destiny of the Treccani is to *wikipedize* itself.

R.: *Do you think they'll do it?*

Eco: With the current spend of renewal of the culture, if an ency-

yclopedia doesn't go online for being updated month by month, is doomed forever. Even when it talks about Parmenides, because even tomorrow a book casting a new light on him can be published... but never mind of Parmenides. Take "Aeroplane", as example: who knows what the article about the "Concorde" said before the Concorde crashed.

R.: *Besides, Treccani tried to "wikipedize" itself. They opened some articles, asking the users to send some edits in...*

Eco: It's a proposal for the *Dizionario degli Italiani* [a collection of biographies], but it's being withdrawn. Since writing articles is too expensive, they asked the readers to "donate" some, not considering that revising those donations takes such a group of editors that the costs are higher than simply paying for the articles.

R.: *It doesn't look easy at all to find a solution to join a traditional model (authors, publishers, redactions) with something as anarchical as Wikipedia. One survives because takes it all: it always has an input from the users, because it picks everything up. And it's free, for everybody: for those who make it and for those who read it. The other has a traditional model that cannot quantitatively stand the comparison and is hardly looking for an equilibrium, taking external contributions (without motivation, taking them and giving back absolutely nothing). Wikipedia is all for free, it's a mutual donation, there's a strong ethic thrust. The clash between these two worlds is not at all a banal question.*

Changing topic, I found it interesting that the process of the "classification" did not have a big success in the collaborative environments. Except for the folksonomies, with people tagging sites and photos, on Wikipedia as well the categorization of the topics is definitely incoherent and incomplete. It's noteworthy that a process as important as the classification/categorization tends to be authorial, personal. In the internet I find few examples of complex categorizations done with a collaborative approach.

Eco: I'm not sure I understood well what you said, but if I did, all of this depend on the fact that – apart of botanic and zoologic taxonomies – a global classification does not exist, but only a local does. In my last book "From the tree to the labyrinth", I wrote a 100-pages essay exactly about the history of the classification, from Porphyrius' tree to what we today dumbly

call “ontologies”.

The problem here is that centuries have been spent in trying to make a total classification, but it’s impossible, it’s always local and in perspective. Consequently, it can be authorial and not collective. It’s a goal attainable in certain fields only, as example animals and trees, as they are universes somehow finite. And it leaves anyway big problems in the classification of the insects. And there’s the famous example of the ornithorhynchus, for which it took them eighty years, but they found an agreement, all together.

Animals then are finite and – one way or another – can be categorized. In those cases where elements are more disperse, instead, the total and collective categorization is impossible.

R.: *Back to something more trivial, in “Sei passeggiate nei boschi narrativi”, you talk about the book “Sylvie” by Nerval as a sort of destiny book, that you studied hundred times.*

The idea that one book corresponds to one person is very striking. Do you still believe in it?

Eco: Yes, although probably there’s more than one book for each person. Yes, I do. But this question is like “Why did you care about the Middle Ages?”, that it’s like asking “Why did you marry that one and not another?” *[he laughs]* If you’re interested, I made the translation of that book and talked about it in a collection of essays about literature. . . but this has nothing to do with the question.

R.: *Do you know the world of the free licenses? They were born in the ’80s and allow the re-using, the sharing, also the editing of the content; very important characteristics in the digital world. Wikipedia too was born within this world and releases its content under a free license. How do you see the world of the intellectual property today, in the age of the internet?*

Eco: I am very empirical. I make my living on the gains of the intellectual property, but every time I’ve been object of piracy I got off cheaply. It happened that my American publisher sued an university for having made thirty copies of a book of mine, and I protested. It’s fine for me like this, at least 3 or 4 of my books can be downloaded through eMule. . . Why am I so careless about this? Considering that I live with that, I should be worried. One answer might be that I’m earning enough this way, the other is that I

am a good democrat.

Let me make an example. When the newspaper *La Repubblica* decided to distribute books with the paper, they began with *Il nome della Rosa*, giving me a modest flat sum. And then they sold two million copies that day. I decided not to mind, I didn't earn anything but it was all right. Six months later I checked the reports of my publisher and the sale of the *paperback* wasn't changed at all. That is, those two million people were people that never would have bought my book in a bookstore. I didn't lose a sale. This means that the "space" is such big that [the piracy] doesn't look like a tragedy to me. It's the author that sells a thousand copies that gets angry if a hundred of them are bootlegged.

Up to the 17th and 18th centuries, a writer made his living from a benefactor's will. Maybe we'll return there, we won't be paid by the audience, but by a patron. Ariosto got off well, why shouldn't I? *[he laughs]*

They got off even before. Then the 18th century revolution – when the storyteller went around selling his own books – gave birth to the rights. In a sense, this democratize that work, because the writer and the philosopher did not have to lick the benefactor's ass any longer.

Well, nothing changed that much between the way that Ariosto licked the Estensi's ass and the way a lot of people lick everybody's ass. *[he laughs]*

Arioso doesn't interest us less because writes two octaves to thank the Estensi.

R.: *About books and rights, it's Google Books⁶ that lately caught big attention.*

Eco: I don't understand all these protests against Google Books. Honestly, I get angry because I can see two pages and I cannot buy the book. The publishers should be enthusiastic, I don't understand. It's like the pedestrian areas: when you close a road to the cars, all the shopkeepers protest, although it's scientifically demonstrated that such an action increases the sales.

R.: *It's a theme entwined with the one about the public domain. Both in US and in Europe there are pressures to lengthen the copyright times,*

⁶<http://books.google.com>

reducing the “area of the public domain”. There’s much fuss and much fear about the intellectual property.

Eco: Each writer lives a conflict about this: on a side he’s fine that his book is read, on the other he’s sorry that his grandchildren won’t earn anything from the rights. Now, my publisher said he’ll give the rights of *Il nome della rosa* for making an eBook for the Kindle, I think. The percentage is much lower than for normal books, but it’s all right. I personally don’t believe in it, I think that people still wants paper for reading a book, but I have no problems, it’s correct that people asking for an electronic version might have it. It doesn’t look complicated, they pay for the rights, although less because the eBook is cheaper. Either it’ll be a *smash* and you’ll sell millions, or you’ll sell few copies, and it’ll be all right anyway.

I think that everybody is overreacting, just like the publishers against Google. Google Books is for selling books, not for selling less books. It plays the same role a bookshop does, when you go to browse the books. You can buy them, or just read a couple of pages, or read the index. Just as with Google.

And there’s the trend to give more and more things for free. I cannot stand Adobe, that every year asks me to pay for reading the PDFs. In few minutes I found programs that do the same for free. I don’t understand where’s the return for the developers...

R.: *Often, nothing. Open-source software is often written by people for personal use (maybe people that didn’t want to pay Adobe’s license like you) and freely released. And if someone makes it better, the creator itself has a return. It’s a virtuous circle.*

Eco: There’s also OOorg⁷, that substitutes Word. It’s very good and it works very well.

R.: *Besides, Wikipedia comes right from the open-source world, that is developing in years its own philosophy about the free culture. They are linked.*

⁷<http://www.openoffice.org>

Appendix 2

- TAG:Linguistic analysis (2)
- TAG:Annotation (3)
- TAG:Apographeme (1)
- TAG:Augmentation (1)
- TAG:Author/Reader (3)
- TAG:Authorship (4)
- TAG:Collaboratory digital library (1)
- TAG:CMPP (1)
- TAG:Catalogue services (1)
- TAG:Classics (1)
- TAG:Collaboration (8)
- TAG:Collaboratories (4)
- TAG:Community (2)
- TAG:Creative Commons (1)
- TAG:Customization (1)
- TAG:Cyberinfrastructure (7)
- TAG:DSpace (2)
- TAG:Data mining (2)
- TAG:Digital communities (3)
- TAG:Digital curation (1)
- TAG:Digital humanities (6)
- TAG:Digital incunabula (1)
- TAG:Digital libraries (11)
- TAG:Digital preservation (3)
- TAG:Digital scholarship (1)
- TAG:Dynamic corpora (1)
- TAG:E-research (1)
- TAG:E-science (5)
- TAG:Ecumenical (1)
- TAG:Collaborative editing(3)
- TAG:Fedora (1)
- TAG:Fourth-generation collections (1)
- TAG:Google Books (2)
- TAG:Granularity (1)
- TAG:Hypertext (1)
- TAG:Informatica umanistica (1)
- TAG:Institutional repositories (1)
- TAG:Internet Archive (1)
- TAG:Interoperability (5)
- TAG:Liber Liber (1)
- TAG:Mark-up (1)
- TAG:MediaWiki (1)
- TAG:Metadata (2)
- TAG:Multitext (1)
- TAG:Named entity services (1)
- TAG:Noosphere (1)
- TAG:OAI-PMH (1)
- TAG:OAI-ORE (3)
- TAG:Open Content Alliance (1)
- TAG:Open content (2)

- TAG:Open source (4)
- TAG:Parsed corpora (1)
- TAG:Perseus project (3)
- TAG:Philology (2)
- TAG:Intellectual property (1)
- TAG:Quantitative analysis (1)
- TAG:Repositories (3)
- TAG:Reuse (2)
- TAG:Scale (2)
- TAG:Semantic Web (1)
- TAG:Service provider (1)
- TAG:Services (4)
- TAG:Servizi (1)
- TAG:Sociology of science (1)
- TAG:Software (1)
- TAG:Stoa (1)
- TAG:TLG (2)
- TAG:Text mining (2)
- TAG:Transparency (1)
- TAG:User empowerment (1)
- TAG:Visualization (1)
- TAG:Web 2.0 (4)
- TAG:Wiki (4)
- TAG:Wikipedia (2)
- TAG:Wikisource (1)
- TAG:Wissenschaft (1)
- TAG:ePhilology (3)

Bibliography

- Agosti, M., Ferro, N., Frommholz, I., & Thiel, U. (2004). Annotations in digital libraries and laboratories—facets, models and usage. *Research and Advanced Technology for Digital Libraries*, 244–255.
- Albert, R., & Barabási, A. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47–97.
- Atkins, D. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. National Science Foundation.
- Berners-Lee, T. (1999, April 14). *Talk to the LCS 35th Anniversary celebrations*. (Retrieved 10:43, June 13, 2010 from <http://www.w3.org/1999/04/13-tbl.html>)
- Berners-Lee, T., & Hendler, J. (2001). Scientific publishing on the semantic web. *Nature*, 410, 1023–1024.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43.
- Blackwell, C., & Crane, G. (2009). Conclusion: Cyberinfrastructure, the Scaife Digital Library and Classics in a Digital age. *Digital Humanities Quarterly*, 3(1).
- Boyd, D., & Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer Mediated Communication*, 13(1), 210.
- Britannica, E. (2010). *wiki*. (Retrieved June 24, 2010, from Encyclopedia Britannica Online: <http://www.britannica.com/EBchecked/topic/1192819/wiki>)
- Bush, V. (1945). As We May Think. *The Atlantic Monthly*, 176(1), 101–108.

- Cammarata, V. (2007). *Innovation Oriented Workplaces The CERN Case*. University of Lugano.
- Celentano, A., Cortesi, A., & Mastandrea, P. (2004, December). Informatica umanistica: una disciplina di confine. *Mondo digitale*, 4, 44–55.
- Cerf, V., Cameron, A., Lederberg, J., Russell, C., Schatz, B., Shames, P., et al. (1993). *National collaboratories: Applying information technologies for scientific research*. Washington: DC: National Academy Press.
- Christiansen, L., Stombler, M., & Thaxton, L. (2004). A report on librarian-faculty relations from a sociological perspective. *The journal of academic librarianship*, 30(2), 116–121.
- Council, N. (2006, July). *NSF's Cyberinfrastructure Vision for 21 st Century Discovery*.
- Crane, G. (2006). What do you do with a million books? *D-Lib Magazine*, 12(3), 1.
- Crane, G., Babeu, A., & Bamman, D. (2007). eScience and the humanities. *International Journal on Digital Libraries*, 7(1), 117–122.
- Crane, G., Bamman, D., Babeu, A., Breuel, T., & Cerrato, L. (2009). Classics in the Million Book Library. *Digital Humanities Quarterly*, 3(1).
- Crane, G., Bamman, D., Babeu, A., & Schreibman, S. (2008). ePhilology: when the books talk to their readers. *A Companion to Digital Literary Studies*.
- Crane, G., Seales, B., & Terras, M. (2009). Cyberinfrastructure for Classical Philology. *Digital Humanities Quarterly*, 3(1), 1–27.
- Crane, G., & Wulfman, C. (2003). *Towards a cultural heritage digital library*.
- Cunningham, L. (2010). The Librarian as Digital Humanist: The Collaborative Role of the Research Library in Digital Humanities Projects. *Faculty of Information Quarterly*, 2(2).
- Cunningham, W. (2003, November). *Correspondence on the etymology of wiki*. (Retrieved 10:43, June 21, 2010 from <http://c2.com/doc/etymology.html>)
- Edwards, P., Jackson, S., Bowker, G., & Knobel, C. (2007). *Understanding infrastructure: dynamics, tensions, and design*.

- Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4), 469–486.
- Engelbart, D. (1963). A conceptual framework for the augmentation of man's intellect. *Computer-supported cooperative work: a book of readings*, 52.
- Engelbart, D. (1988). A conceptual framework for the augmentation of man's intellect. *Computer-supported cooperative work: a book of readings*, 36–65.
- Erlandson, D. (1993). *Doing naturalistic inquiry: A guide to methods*. Sage Publications, Inc.
- Finholt, T., & Olson, G. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science*, 8(1), 28.
- Flick, U. (2002). Qualitative research-state of the art. *Social Science Information*, 41(1), 5.
- Ford, J. (1975). *Paradigms and fairy tales: an introduction to the science of meanings*. London.
- Frumkin, J. (2005). The wiki and the digital library. *Perspectives*, 21(1), 18–22.
- Greene, J. (1990). Three views on the nature and role of knowledge in social science. *The paradigm dialog*, 227–245.
- Guba, E., & Lincoln, Y. (1988). Do inquiry paradigms imply inquiry methodologies. *Qualitative approaches to evaluation in education*, 89–115.
- Guedn, J. (2009). *Open Access*. National Science Foundation.
- Hauben, M. (2004). *History of ARPANET*. (Retrieved 23:13, June 23, 2010 from <http://www.dei.isep.ipp.pt/~acc/docs/arpa.html>.)
- Heisenberg, W. (1958). *Physics and philosophy*. New York: Harper.
- Hockey, S. (2004). The history of humanities computing. *A companion to digital humanities*, 3–19.
- Jankowski, N. (2007). Exploring e-science: An introduction. *Journal of Computer-Mediated Communication*, 12(2), 549–562.
- Koblas, J. (2006). *Wikis: tools for information work and collaboration*.

- Oxford: Chandos Publishing.
- Kouzes, R., Myers, J., & Wulf, W. (1996). Collaboratories: Doing science on the Internet. *Computer*, 29(8), 40–46.
- Krowne, A. (2003). Building a digital library the commons-based peer production way. *D-Lib magazine*, 9(10).
- Kumar, R. (1999). *Research methodology: A step-by-step guide for beginners*. London: Sage Publications.
- Kurzweil, R. (2005). *The singularity is near: when humans transcend biology*. New York: Viking Adult.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. London: Sage Publications.
- Lankes, R., Silverstein, J., Nicholson, S., & Marshall, T. (2007). Participatory networks: the library as conversation. *Information technology and libraries*, 26(4), 17–33.
- Licklider, J. (1960). Man-Computer symbiosis. *IRE Transactions on Human Factors in Electronics*, 1, 4–11.
- Licklider, J. (1965). *Libraries of the future*. Cambridge: The MIT Press.
- Lih, A. (2009). *The Wikipedia revolution: how a bunch of nobodies created the world's greatest encyclopedia*. New York: Hyperion.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. London: Sage Publications.
- Marshall, C. C., & Shipman, F. M. (2003). Which semantic web? In *Hypertext '03: Proceedings of the fourteenth acm conference on hypertext and hypermedia* (pp. 57–66). New York, NY, USA: ACM.
- Massarenti, A. (2002, August 7). L'umanesimo viaggia sul computer. *Il Sole-24 ore*, 6.
- Maykut, P., Morehouse, R., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. Routledge.
- Melia, K. (1997). Producing plausible stories: interviewing student nurses. *Context and method in qualitative research*, 26–36.
- Meta. (2010, May 12). *List of wikipedias*. (Retrieved 19:04, June 6, 2010 from http://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias&oldid=1969133)
- Musser, J., O'Reilly, T., et al. (2006). *Web 2.0: Principles and best practices*.

Cambridge.

- Nelson, T. (1987). *Computer lib: Dream machines*. Oxford: Tempus Books.
- Nelson, T. (2008). *Geeks Bearing Gifts*.
- Neuhold, E., & Frommholz, I. (2004). Digital libraries as experienced by the editors of the journal. *International Journal on Digital Libraries*, 4(1), 1–2.
- O'Reilly, T. (2005, September 30). *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*. (Retrieved at 18:03, June 14, 2010 from <http://oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>)
- Patton, M. (2002). *Qualitative research and evaluation methods*. London.
- Pears, R., Shields, G., & Lancaster, S. (2005). *Cite them right: the essential guide to referencing and plagiarism*. Pear Tree Books.
- Pickard, A. (2007). *Research methods in information*. London: Facet Publishing.
- Raymond, E. (1999). The Cathedral and the Bazaar. *Knowledge, Technology & Policy*, 12(3), 23–49.
- Renda, M., & Straccia, U. (2005). A personalized collaborative Digital Library environment: a model and an application. *Information processing & management*, 41(1), 5–21.
- Ridi, R. (2007). *La biblioteca come ipertesto*. Milano: Editrice Bibliografica.
- Roncaglia, G. (2002). Informatica umanistica: le ragioni di una disciplina. *Intersezioni*, 22(3), 353–376.
- Rydberg-Cox, J. (2006). *Digital libraries and the challenges of digital humanities*. Oxford: Chandos Publishing.
- Shirky, C. (2008). *Here comes everybody: the power of organizing without organizations*. New York: The Penguin Press.
- Stake, R. (1995). *The art of case study research*. London.
- Stenhouse, L. (1984). Library access, library use and user education in academic sixth forms: An autobiographical account. *The research process in educational settings: Ten case studies*, 211.
- Strauss, A. (1987). *Qualitative analysis for social scientists*. Cambridge.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London: Sage Publi-

cations.

- Suber, P. (2009, February 9). *Timeline of the Open Access Movement*. (Retrieved 22:23, June 11, 2010 from <http://www.earlham.edu/~peters/fos/timeline.htm>)
- Surowiecki, J. (2004). *The wisdom of the crowds: why the many are smarter than the few*. New York: Doubleday.
- Tammaro, A. (n.d.). Biblioteca digitale co-laboratorio: verso l'infrastruttura globale per gli studi umanistici.
- Tapscott, D., & Williams, A. (2006). *Wikinomics: how mass collaboration changes everything*. New York: Portfolio.
- Text Encoding Initiative. (2007a). *TEI: Frequently Asked Questions*. (Retrieved 19:49, June 5, 2010 from http://www.tei-c.org/About/faq.xml#body.1_div.1_div.1)
- Text Encoding Initiative. (2007b). *TEI: Guidelines*. (Retrieved 17:49, June 10, 2010 from <http://www.tei-c.org/Guidelines/>)
- Voss, R. (2007, June). *Open access publishing in particle physics* (Tech. Rep.). Geneva: CERN.
- Wagner, C. (2004). Wiki: A Technology for Conversational Knowledge Management and Group Collaboration. *Communications of the AIS*, 13, 256–289.
- Warden, B., & Wong, S. (2007). *Introduction to qualitative analysis*. (Retrieved 15:56, June 24, 2010 from <http://www.eval.org/SummerInstitute07/Handouts/si07.wongF.pdf>)
- Wardrip-Fruin, N., & Montfort, N. (2003). *The NewMediaReader*. Cambridge: The MIT Press.
- Wenger, E. (2006). *Communities of practice: a brief introduction*. (Retrieved 17:03, June 5, 2010 from <http://www.ewenger.com/theory/>)
- Wenger, E., & Snyder, W. (2000). Communities of practice: The organizational frontier. *Harvard business review*, 78(1), 139–146.
- Wiki, M. (2008, September 7). *Soft Security*. (Retrieved 10:49, June 21, 2010 from <http://usemod.com/cgi-bin/mb.pl?SoftSecurity>)
- Wikipedia. (2010a, June 18). *E-science*. (Retrieved 17:33, June 24, 2010 from <http://en.wikipedia.org/w/index.php?title=E-Science&oldid=368766879>)

- Wikipedia. (2010b, June 5). *Moore's law*. (Retrieved 17:03, June 5, 2010 from http://en.wikipedia.org/w/index.php?title=Moore%27s_law&oldid=366166522.)
- Wikipedia. (2010c, June 16). *Neutral point of view*. (Retrieved 18:35, June 15, 2010 from http://http://en.wikipedia.org/w/index.php?title=Wikipedia:Neutral_point_of_view&oldid=368361231)
- Wikipedia. (2010d, May 26). *Project Gutenberg*. (Retrieved 17:34, June 5, 2010 from http://en.wikipedia.org/w/index.php?title=Project_Gutenberg&oldid=364370296)
- Wikipedia. (2010, May 16). *Publish or perish*. (Retrieved 16:57, June 9, 2010 from http://en.wikipedia.org/w/index.php?title=Publish_or_perish&oldid=362454397)
- Wikipedia. (2010, June 14). *Semantic web*. (Retrieved 19:35, June 15, 2010 from http://en.wikipedia.org/w/index.php?title=Semantic_Web&oldid=367996192)
- Wikipedia. (2010, June 24). *Wikipedia*. (Retrieved 17:39, June 24, 2010 from <http://en.wikipedia.org/w/index.php?title=Wikipedia&oldid=369721268>)
- Wikipedia. (2010, May 12). *Wikisource*. (Retrieved 18:53, June 6, 2010 from <http://en.wikipedia.org/w/index.php?title=Wikisource&oldid=361691031>)
- Wikipedia, The Free Encyclopedia. (2010, April 16). *MediaWiki*. (Retrieved 10:33, April 17, 2010 from <http://en.wikipedia.org/w/index.php?title=MediaWiki&oldid=356386203>)
- Wilkinson, D., & Huberman, B. (2008). Assessing the value of cooperation in Wikipedia. *Arxiv preprint cs/0702140*.
- Williams, K. (2009). A Framework for Articulating New Library Roles. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, 265, 3–8.
- Yin, R. (2002). *Case study research: Design and methods*. London: Sage Publications.