



COLLABORATIVE EUROPEAN DIGITAL ARCHIVE INFRASTRUCTURE

Project Acronym: CENDARI
Project Grant No.: 284432
Theme: FP7-INFRASTRUCTURES-2011-1
Project Start Date: 01 February 2012
Project End Date: 31 January 2016

Deliverable No. :	4.4
Title of Deliverable:	Report on digital methodologies
Date of Deliverable:	April 2015
Revision No.:	
WP No.:	4
Lead Beneficiary:	UOB
Author (Name and email address):	Pavlina Bobic, p.bobic@bham.ac.uk Jakub Beneš, j.benes@bham.ac.uk Emiliano Degl'Inocente, emiliano@sismelfirenze.it
Dissemination Level:	PU = public
Nature of Deliverable:	R = report
Abstract (150 words):	This report draws upon the analysis of a number of successfully completed project activities to show the domain specific value, use, requirements and expectations of researchers with regard to digital methodologies. It reevaluates the results of the previous Reports (D4.1 and D4.2), findings of the two domain workshops (2013 and 2014), results of the Summer School on researching the First World War in a digital environment, and the main lessons learned for CENDARI at international digital humanities conferences in 2014, focused predominantly on modern history and the significance of new digital research infrastructures (VRE). The report also includes the responses received from the Summer School participants, who learned how to conduct comparative/transnational research of dispersed

	<p>archival resources (pertaining to repositories across the world) with the aid of cutting edge digital tools. WP3 activities (“Training and Transnational Access Coordination”) thus overlapped with and complemented WP4 (“Methodology and Archive User Requirement”) activities. Historical inquiry in a digital environment has certainly changed the traditional research methods and has enabled a vigorous re-examination of primary resources pertaining to both WWI and MM domain; however, it also brings to the fore a heightened need to bridge the divide between historians and archivists as generators of archival metadata. This makes the CENDARI VRE an important nexus for contextualizing (archival) data, new research initiatives and points of access to digital surrogates.</p>
--	---

CENDARI

REPORT ON DIGITAL METHODOLOGIES

WP4

PAVLINA BOBIČ with JAKUB BENEŠ, UOB
and
EMILIANO DEGL'INOCENTE, SISMEL

April 2015

TABLE OF CONTENTS

1. Executive Summary.....	3
1.1 Aim and Context.....	3
1.2 Method.....	3
1.3 Structure of the Report.....	5
2. Use and value of digital methods in the First World War studies.....	6
2.1 Introduction.....	6
2.2 Use of DM in WWI studies.....	6
2.3 Researching the First World War in a digital environment: CENDARI Summer School 2014 feedback.....	12
2.4 CENDARI key point presentations at DH conferences (2014).....	14
2.5 Value of DM and CENDARI platform in WWI studies.....	17
3. Digital methods and MM studies (section by SISMELE).....	19
3.1 The use of Digital Methodologies in Medieval Studies.....	19
3.2 The Value of DM in Medieval Studies.....	27
4. Digital Methods, historical research and DH.....	34
4.1 Introduction.....	34
4.2 WWI and MM studies: comparison.....	35
4.3 Historical debates in DM: current trends.....	36
5. Conclusion.....	39
6. Cited Bibliography.....	39
7. Appendix.....	40

1. Executive Summary

1.1 Aim and Context

This report on digital methodologies feeds into the deliverable D4.4 (“Report on Digital Methodologies”) to comply with the requirements of Description of Work (DOW) of the project “Collaborative European Digital/Archival Infrastructure”. It centres on the “use and value of digital methodologies to research in the two domain areas, and digital humanities”.

1.2 Method

The report draws on the results from a number of project activities completed to encapsulate the domain-specific value, use, requirements and expectations of researchers with regard to digital methodologies. In particular, it synthesises the results of the previous reports (D4.1 and D4.2), conclusions of the two domain workshops (2013 and 2014), findings of the summer school on researching the First World War in a digital environment, and major lessons learnt for CENDARI at digital humanities conferences in 2014. The report also includes the feedback received from the summer school participants who were presented with new digital research methods through a mixture of lectures and hands-on sessions. In this manner WP3 activities (“Training and Transnational Access Coordination”) overlapped with and enriched WP4 methodological explorations (“Methodology and Archive User Requirements”).

This report explicitly derives from the following key WP4 activities:

a) Overview of archival usage and practice, extent of current use of DM (D4.1)

The report brought forth the results of the desk research and interviews held with researchers to highlight existing research practices and introduced innovative ideas for the development of the CENDARI interface. Interviews with modern/WWI historians and medievalists (26) focused primarily on research practices, transnational history and the expectations of historians with regard to

DH. The report established the significance of search tools, digital aids for the planning of research trips, functions of social networking, and tools to help overcome the problems of archival fragmentation. It also produced a number of suggestions on how to enhance the CENDARI research infrastructure.

b) Use cases based on user scenarios, epics and WWI workshop (D4.2)

Deliverable 4.2 made use of specific user scenarios collected in a variety of contexts to define, particularly in methodological terms, how historians in WWI and MM domains expect to use the CENDARI infrastructure. The report intended to inform the technical development of CENDARI's enquiry environment.

c) User scenarios and user stories

Historian-users from the CENDARI community were asked to outline their specific research topics as well as broadly define the technical aspects needed or desired to carry out their research. These were then used as the examples of user stories, which described specific tasks within the research workflow, attempted ways of pursuing that task as well as the desired results. The partners leading WP8 used these to define technical system requirements. In addition, the report underlined methodological implications from the perspectives of medieval and WWI scholars.

d) Prototype projects

CENDARI implemented two domain-specific prototype projects that further refined the concept of user scenarios and stories. These were produced on the lines of deepened research scenarios with the addition of specific technical functionalities required to implement them. The prototype projects were led by CENDARI experts in the Medieval Manuscripts and WWI domain areas and were intended to support interface development and showcase the possibilities of the research infrastructure.

e) Domain specific workshops

Two interdisciplinary workshops brought together historians, archivists and e-scientists to discuss the questions of how to adjust the CENDARI research

platform with the priorities in each domain field. The WWI workshop was organised at King's College London in early November, while the MM workshop took place at SISMEEL in Florence in March 2014.

f) Participatory design sessions

WP8 and WP9 partners organised three participatory design workshops (Oct. 2012, Jan. 2013, Oct. 2013) that included WP4 partners. These surveyed domain specific research practices and gave participants the opportunity to video-brainstorm their ideas for functionalities of a digital research infrastructure.

1.3 Structure of the report

The report is divided into three main parts. The first part (Section Two) focuses on the current use of digital methods in the First World War studies drawing in particular on Reports D4.1 and D4.2. Its aim is to highlight the value of digital methods to First World War studies by referring extensively to the domain workshop results, summer school feedback and the key points extracted from the digital humanities conferences and CENDARI presentations, which at the same time subsume the latest findings of WP4 leaders or their partner groups. The second part (Section Three) assesses the value attributed to digital methods in the studies of Medieval Europe and has been contributed by SISMEEL (Florence). The third part (Section Four) compares the value of digital methods to the two domains and further highlights the current historical debates in digital humanities.

2. Use and value of digital methods in the First World War studies

2.1. Introduction

Historians of the First World War are confronted with a vast number of primary resources, which can help them significantly expand their research scope as well as deepen interdisciplinary insights and introduce novel historiographical approaches. The conflict that erupted in 1914 is still the subject of on-going debates on continuity and change, placing the war either at the epicentre of shattered “hopes and self-confidence with which the century began” (Michael Howard, 1998) or at the start of the short 20th century lasting from 1914 to 1991 in which fascism, communism and liberal democracy clashed for world supremacy (Eric Hobsbawm, 1994). Over the past decades all aspects of the First World War have become the subject of intense scrutiny in western historiography: experience, memory, tactics, operational method, strategy, gender, empire, race, consequences of, and origins of the war. The vastness of physical (primary) material on this first global modern conflict and recent trends in cultural history of the war have underlined the need for transnational and/or comparative investigation of archival sources that are dispersed in hundreds of repositories organised according to various standards and reflect the often-stark divisions between national or administrative bodies that produced most modern primary documents.

2.2. Use of DM in WWI studies

In view of the persistent source fragmentation, the first task for transnational and comparative historians of the First World War is to locate, connect and organise sources in relevant countries, repositories, and formats. The initial stage of this research requires work with multilingual sources, which are usually obtainable in analogue form. The first stage is also the most decisive and informative stage and has been hitherto largely overlooked by the developers of digital tools, assuming that research begins with a clearly defined volume of

texts. This likewise reflects current concerns with regards to on-going digitisation projects but also guides the cutting edge efforts to sustain the development of digital research platforms.

The first report on digital methodologies (D4.1), based on an interview conducted with 26 historians of medieval and modern Europe, established that digitisation and the digital objects that it produces represent one key aspect of digital humanities. It is, however, still a matter of question as to how crucial they are to scholars in humanities (cf. Terras and Warwick et al, 2012). Furthermore, the disciplinary implications of large amounts of digitised materials remain debatable (Rosenzweig, 2003). Nevertheless, our survey has shown that digitisation and its items stand out in researchers' experience of, and expectations from, digital humanities. The report (D4.1) revealed the interviewees' view of digitisation of documents, particularly in the first stage of their planning of research trips. The large majority of interviewed historians made extensive use of photography in archives, while later strategies to organise collected data proved to be less uniform. There was a strong consensus on the importance of networking with peers and closer communication with archivists. This was even more crucial for historians working on transnational research topics, which are made increasingly complex by linguistic barriers and the fragmentation of archival holdings.

In the survey (D4.1) the First World War historians invariably stressed the benefits of digitisation of finding aids. The report also revealed that historians in fact made little use of the range of digital tools at their disposal and still needed to be persuaded to recognize value in them. Historians likewise asserted that recent digital developments changed research practices but they nevertheless attached deeper significance to developments within historiography and interdisciplinary research rather than digital methods.

The user requirements derived from the First World War workshop in London (Nov. 2013) further opened up and honed the question of the use and value of digital tools in current historical research. Historians who participated in the

discussion unequivocally emphasised an advancing historiographical trend towards pronouncedly transnational and comparative topics based on multilingual sources, while recognizing the need to focus on the hitherto under-researched zones in south-eastern and eastern Europe. The requirements expressed the need to overcome the imposing barriers to pursue such research, emphasising collaborative work, access to heterogeneous sources across national and institutional boundaries, and efficient means for visualising data collected. It has been widely acknowledged that cultural history strengthened WWI scholarship in recent decades but traditional approaches to problems raised by topography, military movements, and diplomatic (international) relations remain relevant.

In terms of the actual (or envisaged) use of new digital tools, First World War historians presented their requirements by providing detailed scenarios for pursuing original research directions in a digital environment. Historians expressed an interest in multinational search and retrieve functions, as well as in a dynamic online space for working on their projects, analysing data, and collaborating with others. Due to the vastness of first hand material held in dispersed repositories, the envisioned digital environment is directed towards the initial research stages of finding, storing and organising data.

Historians agreed on the need to build a research digital environment but also accentuated the demand to build trust with the potential user community. Trust could in their view be fostered through experts in the field overseeing the environment's development with regard to how historians will use it. It was also suggested that the editorial and curatorial decisions are exposed at the original repository and online while privileging exhaustive metadata over allegedly comprehensive digital surrogates in facilitating access to archival sources. Historians likewise proposed to integrate traditional workflows into the digital environment but at the same time emphasised the importance of citable publications to emerge from work in the digital environment.

The hindrances posed by transnational research provide challenges to IT developers in their attempt to build the optimal platform to meet historians' research needs. The WWI workshop in London identified the most pronounced user requirements ranging from a dynamic note-taking environment connected to data; visualising data; connecting with fellow researchers; personalised virtual research room; enabling cross-paths between digital and analogue methods; credibility of digital initiatives; online publications and integrating sources for teaching and research. According to the workshop results (D4.2), scholars see the best potential use of digital tools when working on multilingual sources, which, given the difficulty of finding the highly fragmented resources, call for closer collaboration between IT experts, archivists, librarians, and historians themselves.

On the basis of Reports D4.1 and D4.2 we may claim that new digital technologies have revolutionised the early stages of historical research. It is the stages of "finding" and "finding out" that are crucial for any researcher and where the impact – and value – of the "digital turn" are most discerned. Search functions of Google, Google Books, JSTOR and digital newspapers archives have opened up an immense new field of vision for history. For example, the ability to find all the mentions of a name in millions of pages of OCR-scanned newspapers allows researchers to scrutinise the lives of people in ways they never could before. At the same time, historians have access at their disposal to secondary sources for examining the context of a person, event or organisation, which would have previously required multiple trips to multiple archives, sometimes in different countries (cf. Beneš, 2014). This instant access to topic-identified secondary sources and to term-identified sources have revolutionised the discipline. While historians used to be trained in national academies with their nationally centralised record bases, it is now much easier to pursue research beyond borders.

Digital Turn: Archival Metadata and New Perspectives on Sources

The importance and use of digitised archival metadata is clear. Being able to see the extent and availability of archival sources as well as having access to basic information on their character online is just as “radical” a shift as being able to find people, organisations, places and events in sources at the click of a button. One of the most crucial parts of any research pertains to determining the scale and feasibility of a project, which is essential for planning researcher’s own scientific enquiry. Seeing archival data across the country and in multiple countries makes it much easier for historians to conceive their research strategy. In addition, archival metadata records allow researchers to see what is not there. This may seem a paradoxical reason, but it is paramount. While a researcher can detect a whole range of historical actors who were invisible before because it was too complicated to find them, the apparent but deceptive comprehensiveness of digital searching can render other people and groups more invisible. Also major asymmetries of resources exist from institution to institution and country to country.

Without diminishing its relevance, the archival metadata puts the actual extent of digitised material in perspective. In archival systems, they are also organised according to creator – usually an administrative body. This immediately highlights the limits of the data contained in the collection. The description itself often alludes to missing material too. Recognising and problematizing blind spots belong to good metadata as well as to solid historical research. If introducing a new perspective is the principal benefit of the “digital turn” to historians, then the scholars need tools that will allow them to focus and reformulate their ideas.

CENDARI as a Virtual Meeting Point for Historians

One of the major goals of the CENDARI project is to become a nexus for researchers’ own material that they gather and digitize or what they find online. The CENDARI platform offers a number of possible productive synergies by way of connecting individual researchers, groups of researchers, researchers and software developers, researchers and data curators in libraries and archives, as

well as data curators themselves. It is predominantly oriented towards historians who can rely on very little digitised data in their research. Many scholars have hitherto been sceptical about the digital humanities because very few sources have been digitised.

It has been ascertained that even very large digitization projects escape complete comprehensiveness: for instance, the National Library of the Netherlands has initiated massive newspaper project but has only about 10% of Dutch newspapers digitized. The latter case raises a number of crucial questions on the selection criteria for the resources to be put online. What topics should be covered and how should categories be defined? Moreover, would digitization significantly expand researchers' knowledge? This exposes further issues that are also the main research issues: How to turn the corpus of digitized archival texts into a connected network of information? How to exploit the obtained network structure for interactive exploratory search? Explicit connections are almost always missing in the newly established digitized primary resources, which is arguably the biggest problem facing digital cultural heritage collection.

In this regard the CENDARI WWI prototype project on the green cadres – a loose organization of Austro-Hungarian army deserters that terrorized authorities in 1918 and pushed for a certain kind of social revolution in rural areas – provides a good example of a project that is difficult to pursue in the existing digital landscape. The topic discloses a number of frequent obstacles posed by scattered, multilingual, undigitized sources in a range of institutions, some of which are “hidden” and located in at least 7 countries. An enquiry environment designed with the challenges of this initial case study in mind could, however, offer historians of the First World War the tools to push the boundaries of current research methods and foci. The real challenge in this regard is transnational and comparative nature of research that illuminates the very meaningfulness and function of CENDARI platform and its decentralized, interdisciplinary, and multinational character.

There has been little consensus so far about the nature of the DH field itself and its interventions in humanities disciplines or software developers. The question that asserts itself is: how substantial are (or should be) such digital “interferences”? The meaningfulness of debate about “digital humanities” could cease once the entire scholarly community starts using new digital tools and methods. There is a broad view among researchers that access to existing and ever growing mass of digital resources needs to be facilitated. There is also a widespread desire to raise the bar for scientific knowledge that is accessible online, while preserving the democratic features of the world wide web. The CENDARI project in this view aims at contributing to these common goals. The most difficult challenge, however, lies in the attempt to integrate multilingual sources in various formats and numerous countries, which at the same time brings to the fore legal issues concerning the connection of sources across state borders in various nationally administered archives. In addition, there are issues with conflicting metadata formats that indicate differences in the basic categories and approaches used to describe data in archival and library holdings.

These problems are contingent upon both source-related and technical issues and numerous digital projects, including CENDARI, are developing their own APIs (Application Programming Interfaces) to meet the challenge. The goal and innovative value of CENDARI VRE is, however, to allow researchers to process sources in new ways, which places it among the most advanced and ambitious DH projects. The various new tools and especially cutting-edge modes of visualizing data and text mining sources are finally expected to produce – or give an impetus to – exciting new insights in humanities research.

2.3. Researching the First World War in a digital environment: CENDARI Summer School 2014 feedback

The CENDARI Summer School on researching the First World War in a digital environment provided an appropriate ground to introduce and test the existing CENDARI digital tools for historians and to explore new research possibilities (as

well as an increased value of digital humanities) presented by the latest WWI online projects.

The Summer School was held between 21 – 25 July 2014 at Freie Universität in Berlin. 20 young scholars from Europe and the USA were selected to attend the programme, the aim of which was to highlight the question of how to conduct transnational/comparative WWI research in a digital environment and how to work collaboratively within a digital environment in order to substantially enhance it. The participants were invited to attend seminars by leading scholars in DH; broaden their perspective on the material record on the practice of history; get engaged in the discussion and the production of WWI Archival Research Guides and, finally, to gain familiarity with powerful new tools in research oriented information technologies. The participants were attracted to the programme for a variety of reasons: according to the feedback forms received, they appreciated the “subject, digital history, people”; it increased their “awareness of DH as an important yet little known area”, while giving them the opportunity to learn about “new ideas and methods in a WWI-specific context”. For another participant the Summer School programme was interesting for its “combination of history with new technologies”, thus offering a new insight into how to facilitate digital tools in historical research, which is especially relevant in transnational approach to specific historical questions. Digital tools were also recognised as a useful component to “didactics” and as aids to shape researchers’ own project ideas. The feedback forms showed that the Summer School programme met most of the participants’ expectations but at the same time raised many questions on how to use new digital possibilities in the longer-term future. The participants were particularly eager to expand their knowledge on how to do transnational research more productively; with the help of the research guides introduced in one of the sessions, one scholar was already able to find new material and gain a more thorough understanding of how archivists work and also of how a digital environment can help resolve some research dilemmas. High expectations were expressed with regard to the VRE and the idea of virtually connecting historians in order to support each other. The Summer School succeeded in offering practical lessons on how to engage actively with

new digital projects and how to make them applicable to various research contexts. In result, the VRE and Zooniverse were perceived as “very useful tools for every historian”, whereas the ARG exercise gave the appropriate framework to present specific research problems and make them visible within the scholarly community. The production of an ARG was described as “easy and fun” but it also raised a number of thought provoking questions about their actual usability, authorship, and indispensability in the later stage of scholarly research. In contrast to the idea that the open data and raw data allowed the new field of identification of research material, a view was expressed that the importance of metadata is largely “over-estimated” by non-historians.

2.4 CENDARI key point presentations at DH conferences (2014)

In order to present unique features of digital tools for historians and to further enhance virtual research environment (VRE) strengths within CENDARI platform, WP4 partners participated in discussions on current digital research methodologies at a number of international conferences. These focused predominantly on WWI (modern history) domain and covered the following issues:

1. Francesca Morselli (TCD), ‘CENDARI, the Collaborative European Digital Archive Infrastructure’, **E-Cult Dialogue Day**, KAMU Art Museum Talinn, 14 May 2014. The aim of the workshop was to identify the technological requirements of museums and cultural heritage owners and present existing technology solutions addressing such needs. The targeted audience were museum and other cultural heritage representatives and technology providers to discuss feasible solutions to enhance the accessibility and experience of cultural collections, eventually leading to mutual “benefit models”.

2. Nadia Boukhelifa (INRIA Paris), ‘A Virtual Research Environment for Historians’, **Computer Science and Digital Humanities**, ISC-PIF (Paris Region Institute of Complex Systems), Paris, 4 July 2014. The presentation focused on the principles and goals of the CENDARI VRE that include support of the research

process for historians, collecting notes, collaboration and sharing of documents, data collection as well as fostering ideation through visualisation and analysis, and enriching a common repository of historical information.

3. Francesca Morselli (TCD), 'CENDARI: Historical Research in a Digital Research Infrastructure', **Are we there yet? Functionalities, synergies and pitfalls of major digital humanities infrastructure: Digital Humanities Conference 2014**, Lausanne, 8 July 2014. The workshop addressed the question of the objectives of digital infrastructure projects and their intended users. It aimed to tackle the issues of functionalities and outcomes digital infrastructure projects try to provide, and how they serve the overarching goal of supporting and transforming humanities research. The CENDARI in this context questioned the extent to which the needs of humanities researchers are considered and how to involve the digital humanities research community. The key themes revolved around the following issues: Are there any potential synergies, and actual collaboration, with other infrastructure projects? Conversely, are there any overlaps? What are the main lessons learned so far? What are the pitfalls and potential failures, and what improvements could be achieved? The CENDARI argued for developing a number of innovative elements, which consequently exposed different levels of risk to be taken into consideration. On a content level, CENDARI is a research infrastructure that needs data in order to demonstrate its potential; however, since the project is not an aggregation infrastructure per se, some difficulties appeared with engaging with hidden and small archives. On a research-community level, the project has identified some pitfalls in the tension between the cherished norms of analogue research culture and the new avenues introduced by technological advances (e.g. digital publishing, sharing of research information). On a strategic level CENDARI has established a growing need for matching funding opportunities to the needs of the cultural heritage institutions in order to make the relationship between research infrastructures and heritage institutions more effective.

4. Jakub Beneš (UOB), Evanthia Dimara (INRIA), Alex O'Connor (TCD), 'Researching the First World War in CENDARI's virtual research environment',

Digital Humanities Conference 2014, 7-12 July 2014. In this poster presentation the CENDARI partners showed the development of a VRE based on the requirements of users from the WWI side of the project's targeted research community. It centred on WWI prototype project, which was used as a showcase for CENDARI's ability to support both the analogue and digital needs of its user communities. The presentation argued for a technology that does not interrupt historian's workflow but rather proposed a smooth and on-demand integration of intelligent tools to enable the researcher full control of his project.

5. Aleksandra Pawliczek (FUB), 'CENDARI: Collaborative European Digital Infrastructure', **The Annual Convention of German Archivists**, Magdeburg 25 September 2014. The presentation explored the notion of "digital revolution" and the novel methods and knowledge promoted by CENDARI. It also called for an explicit synergy to achieve access to data, scholarly analysis and shared expertise among researchers, archivists and e-scientists.

6. Aleksandra Pawliczek (FUB), "Connecting Research Practices and Research Communities across Borders", **European Social Science History Conference**, Vienna, April 2014. The author highlighted CENDARI's engagement in defining user requirements and emphasis on the close cooperation in developing tools and interfaces with the communities of medieval and modern historians. CENDARI envisaged integration of user requirements in the development process with the aim of ultimately providing a highly applicable VRE to relevant user communities. The CENDARI Archival Research Guides were presented as an access point to collaborative work and exchange of information on historical resources, the ARGs being part of the virtual research sphere.

7. Francesca Morselli (TCD), 'The CENDARI Archival Research Guides', **A public history in a digital world: The revolution reconsidered**, IFPH Amsterdam, 23-25 October 2014. The presentation examined the methodology of the CENDARI Archival Research Guides by framing them within the field of historical research and DH and investigated possible use by historians and public historians.

8. Jakub Beneš (UOB), 'Seeing Archival Data in Online Portals: Challenges and Opportunities for Research', **Fachkolloquium: Access digital: Nutzen und vernetzen**, Koblenz, 25 November 2014. The online presentation of archival data has transformed historians' research practices, even if that data affords no more than a cursory overview. The paper argued that the primary importance of the digital turn for historians lies precisely in the new optics or ways of seeing sources that it has enabled. It then discussed the strategies for enhancing this vision that portals such as Archives Portal Europe and CENDARI could pursue, in particular by heightening its transnational and thematic dimensions.

9. Alex O'Connor (TCD), 'Cendari: Leveraging Natural Language Processing for Research in Historical Archives', **Exploring Historical Sources with Language Technology: Results and Perspectives**, Huygens Institute for the History of the Netherlands, 8-9 December 2014.

The use and novelty of CENDARI VRE presented in the above mentioned workshops link up with the very value that researchers attach to (or discover in) DM when exploring historical sources on both medieval and modern Europe.

2.5 Value of DM and CENDARI platform in WWI studies

The key process of finding sources for transnational/comparative historical research in dispersed national databases may currently only be obtainable through a very broad lens enabled resources like the Europeana digital library or the Archives Portal Europe. The results of the initial online query give researchers a global understanding of "what is there" (or, paradoxically, of what material is not there) and hence a fundamental idea of how to shape and proceed with the research plan. This global query, however, provides much more basic contextual information than the national databases.

The CENDARI project aims at producing digital aids to help historians determine the “further” steps in their research. The envisaged VRE is designed to guide and build ways of taking individual projects through gradual stages in research process. The VRE will allow historians to collect and organise what they have seen in their own personalised online project space. It will enable historians to annotate and enrich their collected data, thus making thematic connections that comprise their analysis. Archivists and librarians will be able to use this platform too to link data on sources across national and institutional borders drawing on their own specialized knowledge. All of these annotations will be visually distinct from the original data given by the content provider. The VRE recognizes, however, that data for many potentially important collections do not yet exist in any digital form, so it will allow the researcher to create metadata records. Finally, using semantic tools, it will be possible to see and resolve entities across all collected data and in the CENDARI data repository. In this ‘enquiry environment’, the records that a researcher has found in transnational and national archival portals and databases acquire new meaning. The platform makes the transition between the radically new ways of finding and gathering sources to practices of scholarly inquiry that is less prone to change. An increased value of DM in CENDARI - and beyond - is based on recognition that research in the digital age depends on interdisciplinary collaboration between academic historians, cultural heritage institutions, and e-scientists. In this sense it also requires a new vision of how to fit together the various projects in digital landscape, deriving mutual benefit.

According to our findings (D4.1; D4.2; Summer School feedback), the greatest value that WWI researchers attach to digital methodologies stems from their pursuit of new research directions (or initiatives) found in a digital environment. Beyond multilingual and multinational search and retrieve functions, historians are interested in a dynamic online space for conducting their projects, analysing data, and establishing contacts with other researchers in the field. Because of the specific source challenges of First World War studies – above all the vastness of analogue holdings kept in dispersed repositories – the major benefit and value of

DM remain oriented towards the early research stages of finding, storing, and organizing data.

3. Digital methods and Medieval Studies

3.1 The use of Digital Methodologies in Medieval Studies

One of the first attempts to use digital tools and methodologies in the humanities is represented by the pioneering work of Roberto Busa S.J. Busa started the computer-aided lemmatization of the complete works of Thomas Aquinas in the 50s of the XX century, but the project took more than 30 years to complete. During his attempt to build a tool to perform textual searches and linguistic analysis on the massive corpus of the Aquinas' works a number of different technologies followed one another: from punched cards to hard drives, from the CD-ROM to the World Wide Web¹. One interesting fact, worth mentioning, is that Roberto Busa was a Jesuit and a medievalist. Busa was not an exception: since then the medieval scholarly community continued to use digital tools and to produce a vast amount of digital research products, almost without interruption till the present time.

Over recent years, medieval scholars have been surrounded by the availability of a plethora of digital resources, including (but not limited to): digital collections of primary sources and scholarly literature, bibliographic databases, textual corpora and dictionaries, repertoires and authority lists, thesauri and gazetteers (i.e.: medieval authors and works, place names, manuscripts shelf-marks etc.). In particular disciplines, databases have evolved into Virtual Research

¹ For a complete overview cfr. *Informatica e Scienze Umane. Mezzo Secolo di Studi e Ricerche*, Firenze 2003. The book contains essays in italian, french and english. Busa wrote a sort of *intellectual testament* on pp. 57-72.

Environments (VREs)²: digital infrastructures providing a framework of resources and tools to support the work of the researchers.

Doing research in the Humanities (i.e.: in medieval studies) in the digital era means more and more working with digital data, tools and methodologies in a scattered and heterogeneous landscape. The year 2002 - in particular - marked the beginning of the digital age: for the first time in history, mankind produced and stored more digital information than analogue. Furthermore, the world's capacity to store information in digital format - that represented merely 1% of the total in 1986 (roughly 3 exabytes³) - increased to 25% of the total in year 2000 and literally exploded to 97% of the world's storage capacity (about 300 exabytes) in 2007⁴. Medieval scholarly research is therefore changing according to the general information trend, becoming more and more digital, moving from the medieval scriptorium to the digital *screentorium*.

Since the end of the 90s of the XX century, the increasing amount of available digital information sources, raised among historians and medievalists in particular, specific concerns on the scientific quality of the available material, on the technological quality and interoperability of the different databases and

² Blanke, T., et al. *Deploying general-purpose virtual research environments for humanities research* Philosophical Transactions of the Royal Society A: Mathematical, physical and engineering sciences 368.1925 (2010), 3813-3828.

³ Cfr.: Hilbert, M and López, P. *The World's Technological Capacity to Store, Communicate, and Compute Information* Science, 332.6025 (2011), 60-65: "The world's technological capacity to store information grew from 2.6 ("optimally compressed") exabytes in 1986 to 15.8 in 1993, over 54.5 in 2000, and to 295 (optimally compressed) exabytes in 2007. This is equivalent to less than one CD (650 MB) per person in 1986 (539 MB per person), roughly four CDs per person of 1993, 12 CDs per person in the year 2000, and almost 61 CDs per person in 2007. Piling up the imagined 404 billion CDs from 2007 would create a stack from the earth to the moon and a quarter of this distance beyond (with 1.2 mm thickness per CD)"

⁴ Hilbert, M. *How much information is there in the "information society"?* Significance, 9(4), (2012), 8-12.

eventually on the level of digital literacy required for the medievalists to access and interact with the digital ecosystem⁵.

Contemporary research in the domain of medieval culture is by design transnational, multilingual and interdisciplinary: aiming at becoming one of the leading digital platforms for doing research in this field, the CENDARI infrastructure should then be able to address a number of scientific needs coming from different disciplinary traditions in medieval studies.

Since we need to address a vast number of scientific needs coming from the scholarly community, SISMEL established connections with other relevant EU projects and actions in the same domain such as the IS1005 COST action - *Medieval Cultures and Technological Resources* and the *Text and Manuscript Transmission of the Middle Ages in Europe* - TRAME⁶ project.

The context provided by the IS1005 COST action was relevant because it grouped together major research institutes from 25 different European countries including SISMEL, FEF, Warburg Institute, Institut de recherche et d'histoire des textes, K.U. Leuven, Ghent University, Cyrillo-Methodian Research Centre at Bulgarian Academy of Sciences, Mittellateinisches Seminar der Universität Zürich, Institute of Philosophy of the Academy of Sciences of the Czech Republic - Institute for Classical Studies, Heidelberger Akademie der Wissenschaften, Universität Bonn, Institut für Germanistik, Skandinavistische Abteilung, Københavns Universitet, National - Kapodistrian University of Athens, Universitat Autònoma de Barcelona, Helsinki School of Advanced Studies, Académie des Inscriptions et Belles Lettres Comité Du Cange, University of Haifa, Institute for Icelandic studies, National and University Library of Iceland, Fondazione Ezio Franceschini, University of Bergen, University of Oslo etc.).

⁵ Cfr. Degl'Innocenti, E. *Trame e fili di Arianna nella massa digitale* in *Archivi Informatici per il patrimonio culturale. Convegno Internazionale*, Roma 2006, pp. 155-71.

⁶ For a detailed overview cfr. the Report on the CENDARI Experts Seminar *Medieval Scholarly Research and the Digital Ecosystem* (Florence March 20-22, 2014).

This network - through its activities - helped to capture a number of actual scientific questions and needs that we considered in this overview, and to develop a more articulated idea on the role of medieval culture in the digital domain. In a long-term development perspective, the collaboration with major actions, projects and infrastructures (both at a national and international level) will be essential: SISMELE and FEF where both involved in the COST IS1005 action. A deep level of interconnection and collaboration between the IS1005 and the CENDARI medieval community has been an added value since the beginning of the project: in July 2013 the 1st CENDARI Summer School took place in Florence, as a joint initiative with the mentioned IS1005 Action⁷.

The TRAME⁸ project was also relevant as a starting point for the design of the medieval section of the CENDARI environment, since it was already addressing some of the issues raised by the scholarly community of medievalists, coming from the same ground (i.e.: the research institutions).

SISMELE also tried to capture the necessities coming from the scholarly community leveraging on the results of workshops, conferences, roundtables and questionnaires. In May 2014 SISMELE organized together with the French Biblissima project and the IS1005 Cost action, a workshop in Paris: *From the Middle Ages to the Present Times. Towards a Digital Research Infrastructure and the European Agency for Interoperability*. During this workshop, the organizers posed questions to the participants for this specific aim. The scheme of the questionnaire was very simple:

1. According to you, what are the past or on-going major projects in Digital Humanities in your regional / national / European context?
2. What about interoperability in your research / library area?
3. What would be your needs in the short or medium term for your own data?

⁷ <http://www.cendari.eu/research/summer-school-2013/> (last access, 07/03/2015)

⁸ On TRAME see also the above Report on the CENDARI Experts Seminar *Medieval Scholarly Research and the Digital Ecosystem*.

- a. working on my data to make them (more) compatible with Linked Open Data
- b. an aggregator/operator who would harvest my data and align them with other relevant data (ontology/transformation)
- c. a usage/dissemination tool (web portal/tool for researchers etc.)
- d. a specific project on potential common data (authority files, thesauri, vocabularies in specific knowledge area)

Whilst the first two questions were focused on specific national and regional situations, the third question has revealed a core of shared issues and needs across EU research institutions, such as:

- foster data interoperability both at individual/institutional and general level;
- implement semantic technologies and tools to create a medieval digital knowledge space for scholarly research;
- develop meta-search and semantic tools with a *from data to knowledge* approach
- elaborate mappings between existing databases, in order to help the interlinking, interoperability and foster data sharing;
- promote the integration between different authority lists, thesauri, gazetteers and related entities, with the following priorities:
 - manuscripts shelf-marks;
 - anthroponyms;
 - toponyms;
 - titles of works;
 - lemmas;
 - meanings;
 - events
 - objects

Within the broader medievalists' community the workshop spotted at least 2 different - nearly opposite - approaches in employing digital methods and tools: on the one hand, they have the basic need to have access to every possible digital

content available (i.e.: manuscripts, texts, bibliography, catalogues, etc.). It represents an early phase of the link between traditional scholarly methods and digital methods for medieval studies: at this stage, the solution for scholarly needs is the digitization of documents and their publication on private and public access platforms. These contents can be presented with or without structured metadata and the user can search among them using more or less refined search systems (browsing, lists, etc.).

On the other hand, a number of digital medievalists started feeling the need for more complex environments with interoperable DBs bridging different research contexts (e.g.: Latin and vernacular) and linking different kind of information: an example is represented by SISMELE and FEF projects and DBs integrated in the Mirabile platform. In Mirabile⁹, repertories, analytic bibliography, external and internal descriptions of medieval manuscripts as well as information on medieval authors and works are available and accessible for the users through a highly integrated graphical user interface.

Connecting Data

Medieval & Modern Latin and Vernacular Cultures
Bibliography: + 350.000 records
Thesauri: + 320K medieval mss. Shelf marks, + 15K names; + 80K variants
Medieval manuscripts descriptions: internal (philological, paleographical) + external (codicological/paleographical). Identification of texts, authors, incipit / explicit, ...
Inventories of medieval libraries: + 9K records with 3000 anthroponyms, 1200 toponyms, more than 16.000 digital reproductions of the printed editions
Thematic DBs. Hagiography: 1700 texts in prose and poetry on +500 saints in +1000 manuscripts, including vulgarizations and vernacular hagiographic texts
Thematic DBs. Gregory the Great: +8400 manuscripts found in European and extra-European libraries

⁹ <http://www.mirabileweb.it> (last access, 07/03/2015)

Connecting Data

Medieval & Modern Latin and Vernacular Cultures
Medieval Digital Latin Library: Medieval authors writing in latin from VI to XVI century: +600K digital images with metadata, more than 100 digital texts in critical edition
The Digital Plutei Collection of the Biblioteca Medicea Laurenziana in Florence: +4K mss. fully digitized with descriptive metadata. 3 printed catalogues fully digitized with descriptive metadata: More than 1.500K digital images
Lexicographical data. The Italian Online Dictionary: compiled from the texts contained in the OVI databases: 26,500 entries already online, with an average of ca. 2,000 new entries per year. The TLIO corpus: 21,862,472 occurrences in 1995 texts and 443,784 different written forms; The Old Italian Corpus: 23,160,300 word occurrences in 2,314 texts (467,190 different written forms)

Since it seemed to be pointless to try to address every possible research interest and need, the attention in the CENDARI project was focused on 3 scenarios, based on everyday research practice in medieval studies:

Scenario 1: scholars working on medieval authors and texts

Scholars that are mainly concerned with the accessibility of research tools (i.e.: both existing and new databases) providing information on authors and texts from the medieval period such as repertories, authority lists and bibliographies.

In this perspective both the Latin and the vernacular worlds are to be taken in account to allow direct access to all possible medieval authors, dealing with multilingual contents issues and access. Concerning the concept of works (i.e.: texts) scholars should be able to find *oeuvres*, opera and kinds of opera, the genres, linked with literary, diplomatic, archaeological sources etc. The CENDARI research environment should allow a circular data structure to pass from digital object to digital object (across different data-types and languages) and to switch from a context (linguistic, conceptual, etc.) to another. It should include texts,

dictionaries, images (referring to secondary literature and also bibliography) and allow the scholar to do the path back and forth. The need to address and compare different traditions (eg. Arabic world), timelines and geographical data are also to be included.

Scenario 2: scholars working on manuscripts and textual tradition

Scholars are mainly concerned with manuscript descriptions and digital palaeography, codicology and bibliography, scholarly editing and digital scholarly editions. In this perspective they pay particular attention to issues related to:

- mss identification (shelf-marks);
- mss datation;
- mss cataloguing (on-line cataloguing initiatives and standards);
- digital editions;
- stemmatological analysis.

Digital editions are now establishing themselves as the norm in many areas of philological endeavour, with a number of large-scale digitization and editorial initiatives under way in recent years. The use of digital technologies in the production of scholarly editions should by all means be promoted, but it is vitally important that established standards are adhered to in order to ensure maximum interoperability and long-term preservation of and access to digital data. In this perspective it is crucial to give access to a number of different resources (i.e.: texts, images, transcriptions, modern translations, morphological edition of text) and foster the development of tools for representing time (along with different chronological systems) to perform cross boundaries (time, space, language etc.) and allow a high degree of data circularity.

Scenario 3: scholars working on textual corpora and reading tools

Due to the nature of the sources and tools they deal with, scholars' needs are quite peculiar and mainly focused on digital integration of various electronic

dictionaries. Some of their needs are also focused on tools to work with online textual corpora and dictionaries, with possibilities for interaction among parallel linguistic traditions, and for advanced linguistic analysis. Due to the nature of a typical dictionary article, often presenting different types of information, this scenario could represent a driving force for the creation of a dense network of different resources. This information can include:

- one or multiple passages quoted from source texts, that provide the possibility for cross-linking with digital text corpora where these are available;
- bibliographical references to primary and secondary sources and to other dictionaries, that can serve as a basis for linking with electronic bibliographies
- in less frequent cases, mentions of manuscripts, which could be connected with digital manuscript inventories.

The possibility to create interactions between textual corpora and reading tools on the one hand and other types of on-line resources – such as bibliographies and manuscript inventories – may also be explored.

Conclusions

The feedback coming from the scholarly community proved that to build a better digital ecosystem for Humanities Research (in particular for medieval studies) means - above all - to foster the representation of a complex body of knowledge in a connected and interoperable way and provide a platform for applying sophisticated discovery and analytical tools on research data across it, reducing fragmentation and overcoming the data-silos logic, avoiding lack of coordination, integration (e.g.: collections & manuscripts with secondary literature) and interoperability - both in technical (e.g.: metadata formats) and in scientific terms (e.g.: thesauri, vocabularies and indexes).

3.2 The value of Digital Methodologies in Medieval Studies

The medieval digital ecosystem has been traditionally a highly complex and fragmented knowledge space, often struggling with issues connected to the presence of multilingual contents, to the difficulty of accessing contents and resources and eventually to the necessity of dealing with a vast number of non interoperable - though valuable - resources using a variety of different standards and technologies. Recently, the rise of semantic tools and technologies has seemed to pave the way for a new approach towards the organization and publication of scholarly research data about medieval manuscripts, fostering collaboration, data openness, interconnection and reuse.

One of the most relevant achievements of the use of digital methodologies in medieval studies would be to allow researchers to stop making database queries and start asking large-scale research questions across different digital resources, leveraging the power of the Linked Open Data¹⁰ technologies, building bridges between different data silos (existing databases and vocabularies, etc.) and allowing the creation of a new generation of digital research tools.

Another nascent improvement is represented by the creation of tools to help users and researchers find the data they need, without having to deal with hundreds of different websites (despite their quality and relevance): very often indeed, due to lack of integration and interoperability between different digital resources, most of the relevant resources have to be consulted separately and individually. As a result, researchers around the world still face major difficulties in finding, using and sharing knowledge about medieval manuscripts, most of the times ending up using generic search engines like Google.

Furthermore, the availability of detailed descriptions (codicological, paleographical, philological, etc.) of manuscripts as physical objects and as cultural witnesses, will open up to the possibility to address a wide range of research questions, including their history (the place of origin, the date or period

¹⁰ <http://www.w3.org/standards/semanticweb/data> (last access, 07/03/2015)

of origin, the materials used, the decoration and illumination, the handwriting, the scribe, the binding, arrangement of the physical volume, and the language) and a number of cultural aspects concerning life in the medieval period (based on the interpretation of the contents of a manuscript: texts, illuminations, etc.). To allow for this kind of research a standard and interoperable format for such descriptions is required (including authors' names, titles of works, incipits, subject and concept terms, etc.).

The constantly growing number of different – but not interoperable – resources nowadays is simply too many to be managed.¹¹ This is on the one hand a great added value to manuscript research, but, on the other, due to the growing complexity of the knowledge space (filled with descriptions in different formats and multiple languages, many variant forms of names, titles and concepts) researchers are not able to fully exploit the richness of the available data to address complex, large-scale research questions (i.e., trace the relationships between different manifestations of the same manuscript; trace relationships between manuscripts and other related contents, like commentaries, translations and secondary literature, etc.)¹².

Conclusions

As both the surveys and the other feedback from the medieval scholarly community suggested¹³, for the digital turn to become reality in their domain, the medievalists will need to move towards the development of a medieval digital ecosystem for scholarly research, focusing on a small core of shared priorities (see below, Table 1: Priorities) to transform the tremendous amount of knowledge stored in existing traditional databases (i.e. all the relevant entities, such as: names, identifiers, concepts, etc. See below, Table 2: Types of

¹¹ On this specific subject cfr.: D. Weinberger, *Too Big to Know* New York, 2011

¹² Cfr. A. Pavicini Bagliani, *Medioevo Europeo. Medieval Studies And Technological Resources in The Safeguard of Cultural Heritage. A Challenge from the Past for the Europe of Tomorrow* Firenze, 2011, pp. 47-9.

¹³ See above (for the survey) and the Report on the CENDARI Experts Seminar *Medieval Scholarly Research and the Digital Ecosystem* (Florence March 20-22, 2014).

information to be included and Table 3: Controlled vocabularies to be included) into Semantic Web formats and make the resulting datasets available for reuse.

<ul style="list-style-type: none"> ✓ Integration <ul style="list-style-type: none"> ✓ Sources ✓ Datasets ✓ Secondary lit. ✓ Authority lists etc. <ul style="list-style-type: none"> ✓ Shelf ✓ Authors ✓ Evaluation and curation ✓ Data dissemination and reuse ✓ Continuous integration of new resources for <ul style="list-style-type: none"> ✓ Better discoverability ✓ KB enrichment 	<ul style="list-style-type: none"> ✓ Towards a BIG-DATA approach for the Humanities. simple resources discoverability to the establishment of a comprehensive Historical knowledge base, built on high-quality scholarly databases <ul style="list-style-type: none"> ✓ Mapped ✓ Triplified ✓ Curated ✓ Imported into the Medieval KB ✓ To create a bridge between scholarly research communities and Memory institutions in USA and
---	---

linked data integration	advanced data integration tools
permanent identifiers management	advanced data visualization techniques
permalinks (for citation etc.)	advanced H/M interaction
sharing and integration of external authority lists and thesauri (TGN, VIAF etc.): <ul style="list-style-type: none"> • http://www.getty.edu/research/tools/vocabularies/tgn • http://viaf.org/ 	advanced tools for data export and reuse
semantic and annotation tools	

TABLE 1: Priorities

topics/subjects	→				- reference repertoires
sources (mss. / documents)	→ person names	→	→ roles	→	- printed and digital editions
	→ works		→ religious order		- selected Studies
	→ anonymous texts	→	→ literary genres	→	- general Bibliography
			→ literary forms		
			→ documents type		

places	→ institutions	→	→ religious orders	→	
--------	----------------	---	--------------------	---	--

TABLE 2: type of information to be included

Content	Type	Example	Notes
topics/subjects *	authority list, multilingual	Storia delle città medievali, History of medieval cities	<i>related to:</i> shelfmarks, person names, works titles and anonymous texts
shelfmarks *	authority list, multilingual, complex syntax	Firenze, Biblioteca Medicea Laurenziana, Plut. 01.17 Florence, BML, Plut., 01.17	<i>related to:</i> person names works titles anonymous texts, place names, institutions, literary forms and genres
person names *	authority list, multilingual, complex syntax	Alcuinus de York Alboinus Flaccus Pius II papa Aeneas Piccolomini Aeneas Silvius Aeneas Silvius Piccolomini	<i>related to:</i> roles, works titles, religious orders, place names, shelfmarks, institutions

		Aeneas Silvius Piccolomini Senensis	
personal roles *	authority list, multilingual, simple syntax	abbas, papa, advocatus, professor ...	<i>related to:</i> person names
religious orders *	authority list, multilingual, simple syntax	OFM, OESA, OP	<i>related to:</i> intitutions, person names
works titles *	authority list, multilingual, complex syntax	Lectura super Iohannem [Reportatio]	<i>related to:</i> person names, shelfmarks
anonymous texts titles *	authority list, multilingual, complex syntax	Anonymus Valesianus De sepultura eorum qui falso excommunicati dicuntur non turbanda	<i>related to:</i> shelfmarks
place names *	authority list, multilingual, simple syntax	Teschen → Cieszyn	<i>related to:</i> institutions
institutions *	authority list, multilingual, simple syntax	Italia, Toscana, Firenze, Santa Maria Novella, Convento OP	<i>related to:</i> place names
literary genres *	authority list, multilingual, simple syntax	Chronographia et computus, Drama comicum	<i>related to:</i> person names, shelfmarks, works titles,

			anonymous works
literary forms *	authority list, multilingual, simple syntax	Accessus, Anthologia, Carmina, Commentum, Dialogus	<i>related to:</i> person names, shelfmarks, works titles, anonymous works
document/source type *	authority list, multilingual, simple syntax	Atto di acquisto, Atto di cessione, Atto di confisca, Atto di consegna, Atto di costituzione di società, Atto di divisione dei beni, Atto di donazione	<i>related to:</i> shelfmarks, works titles, anonymous works

TABLE 3: controlled vocabularies to be included

Content	Type	Example	Notes
topics/subjects	authority list, multilingual	Storia delle città medievali, History of medieval cities	<i>related to:</i> shelfmarks, person names, works titles and anonymous texts
shelfmarks	authority list, multilingual, complex syntax	Firenze, Biblioteca Medicea Laurenziana, Plut. 01.17 Florence, BML, Plut., 01.17	<i>related to:</i> person names works titles anonymous texts, place names, institutions, literary forms and genres
person names	authority list, multilingual, complex syntax	Alcuinus de York Alboinus Flaccus	<i>related to:</i> roles, works titles, religious orders, place names, shelfmarks, institutions

		Pius II papa Aeneas Piccolomini Aeneas Silvius Aeneas Silvius Piccolomini Aeneas Silvius Piccolomini Senensis	
personal roles	authority list, multilingual, simple syntax	abbas, papa, advocatus, professor ...	<i>related to:</i> person names
religious orders	authority list, multilingual, simple syntax	OFM, OESA, OP	<i>related to:</i> intitutions, person names
works titles	authority list, multilingual, complex syntax	Lectura super Iohannem [Reportatio]	<i>related to:</i> person names, shelfmarks
anonymous texts titles	authority list, multilingual, complex syntax	Anonymus Valesianus De sepultura eorum qui falso excommunicati dicuntur non turbanda	<i>related to:</i> shelfmarks
place names	authority list, multilingual, simple syntax	Teschen → Cieszyn	<i>related to:</i> institutions
institutions	authority list, multilingual, simple syntax	Italia, Toscana, Firenze, Santa Maria Novella, Convento OP	<i>related to:</i> place names
literary genres	authority list, multilingual, simple syntax	Chronographia et computus, Drama comicum	<i>related to:</i> person names, shelfmarks, works titles, anonymous works
literary forms	authority list, multilingual, simple syntax	Accessus, Anthologia, Carmina, Commentum, Dialogus	<i>related to:</i> person names, shelfmarks, works titles, anonymous works
document/source type	authority list, multilingual, simple syntax	Atto di acquisto, Atto di cessione, Atto di confisca, Atto di consegna, Atto di costituzione di società, Atto di divisione dei beni, Atto di donazione	<i>related to:</i> shelfmarks, works titles, anonymous works

In the current state of development, all the needed information is extracted from the AIM – Archvio Integrato per il Medioevo (Integrated Archive for the Middle Ages), the internal data management system of SISMELE and FEF. All the data coming from the below databases are also available via the MIRABILE web portal¹⁴:

¹⁴ The MIRABILE web portal is available at: <http://www.mirabileweb.it> (last checked 21 April 2015).

- **CODEX: A digital census of medieval manuscripts in Tuscany.** More than 5000 internal and external descriptions of medieval manuscripts in Tuscany, including archives, libraries, monasteries, churches/seminars, museums and other cultural institutions.
- **Authority list of Medieval Authors' names:** an index of medieval authors and texts based on *Medioevo latino, Bibliotheca Scriptorum Latinorum Medii recentiorisque aevi* and *Compendium Auctorum Latinorum Medii Aevi*.
- **Authority list of Medieval Manuscripts' shelfmarks:** and index of medieval manuscripts' shelfmarks based on *Medioevo latino, Bibliotheca Scriptorum Latinorum Medii recentiorisque aevi* and *Compendium Auctorum Latinorum Medii Aevi*.

Other online resources could be added during the Ontology modeling process, according to specific research needs.

4. Digital Methods, historical research and DH

4.1 Introduction

The CENDARI project has been described as an example of a “digital ecosystem” (Gartner and Hedges, 2013) in that it conveys key attributes of balance,

engagement, interaction or self-organisation. It aims at designing an open community that continuously grows an information network into which research results are fed. The novelty of the infrastructure lies in the desire to embed the research process into its design by envisioning the archives as a nexus of information that fluctuates together with historians' research process. CENDARI's major goal is to build a research platform that surpasses the function of a mere portal to resources. In other words, the project's ambition is to construct an "enquiry environment" in which dispersed collections can be linked and analysed in new ways to introduce shared virtual research infrastructures. It is creating tools and workspaces that will enable scholars to engage with fragmented and hard to find sources while defining workflows that will allow for the construction of personalised research spaces, shared virtual research and teaching environments, annotations trails, etc.

The innovative process in DH is reciprocal: advancements in digital methodology have substantially changed (and keep changing) archival historical research, while guiding new ways to discover meaning and create knowledge. The earlier CENDARI project report (D4.1) has shown that historians across all periods consider the digitisation of online finding aids and of documents, the possibility of full-text search, and the availability of online catalogues a major asset. This new research practice is, however, not diminishing the importance of traditional methods: as one historian in our survey put it, "history is not about sharing the information, but about sources and choices you make as a historian" (Report D4.2). In the same vein the question of whether there is a clear-cut distinction between "archival and researcher-produced data" exposes a new potential for developing resources and environments for the benefit of the wider community of scholars.

As Wendell Piez's has argued, the proper object of DH is what one might call "media consciousness" in a digital age. It illuminates a certain critical attitude equal to, and continuous with, a more general media consciousness bestowed upon "cultural production in any nation or period" (Piez, 2008). What we study

in DH are not only digital media and the reciprocal cultural impacts of digital media, but also the process of designing and producing them.

The following sections will further highlight the commonalities and divergences between the two domain subjects embedded in digital workflows and look at how the specific digital methodologies – that are most promising for WWI and MM studies – have enriched or refocused debates in the DM.

4.2 WWI and MM studies: Comparison

According to our survey, the major value of DM for modern historians lies in enabling or accelerating new research -- as well as research techniques -- within a digital environment. This is particularly important in the view of enormity of primary resources available to WWI researchers who are invariably faced with the problem of tracing down sources in the (often) dispersed repositories. The latter is especially relevant in the early stages of defining the project, storing and organising data.

In the domain of medieval studies, however, digital resources, collections of primary sources and academic literature, bibliographic databases, textual corpora and dictionaries, repertoires and authority lists, to name the most important ones, are becoming increasingly available to researchers. This also means that academic research in medieval studies is more and more reliant upon digital data, tools and methodologies. In consequence, medieval historians can reasonably expect digital access to a wide variety of relevant sources, while WWI historians are able investigate only a rather limited amount of digitised material (or selected digitised papers normally obtainable in analogue format). However, the need for a more complex digital landscape with interoperable databases to connect different research contexts and different sorts of information binds both medieval and WWI domain in a strongly correlated digital research environment. The MM digital sphere has been traditionally viewed as a highly fragmented “knowledge arena”, faced with the problem of multilingual resources, access to them, and a huge number of non-interoperable sources using a variety of

standards and technologies. An early phase of medieval research is a combination of traditional and digital scholarly methods; WWI historians too employ the same methods to clarify their research objectives.

Due to fragmentation of knowledge stored in traditional databases, medieval historians are focused on preparing these silos for delivery via the semantic web, ready for scholarly re-use; WWI historians require instead a dynamic virtual research space for carrying out projects, analysing data, and establishing as well as fostering contacts with other researchers in the field.

Both domains are defined by transnational, multilingual and interdisciplinary research and the CENDARI infrastructure is currently addressing the technical aspects of making it one of the leading digital platforms for conducting virtual research in the two fields. They are bound together in their requirement to overcome the previously discussed research issues that make the medieval and WWI eras both very specific as well as complex areas of study.

4.3 Historical Debates in DM: Current Trends

It has been claimed that the digital age has both improved and exacerbated problems in the scholarly research process. The growing number of small resources and portals curated outside the traditional archival environment has given an incentive to numerous knots of integration that are not necessarily reused beyond their creators. On the other hand, large (national) archives and libraries have responded to the “digital turn” with a strong stimulus to digitise finding aids, catalogues, special collections and record series as well as by the creation of portals and hubs to provide searching of dispersed sources. Archives and libraries create knowledge in the way of structuring, describing and classifying their collections and are responsible for the transparent information as well as document discovery or retrieval. The “archival knowledge” is inhabited in finding aids, catalogues, classifications, and in the expertise of the archivists and the librarian. On the other hand, knowledge in humanities is

derived from a hermeneutic procedure, which claims sound argumentation and a reservoir of previously gained insights in the discipline.

Digital projects – and CENDARI as a “digital ecosystem” in particular – are intended to bridge the divide between the archival and conceptual creation of knowledge by means of designing a dynamic framework for scholarship. The flexibility in a digital environment acts as a mechanism to virtually integrate fragmented sources by placing them in historical and humanistic contexts while simultaneously accommodating and acknowledging the “knowledge populating” structured archive and library metadata (cf. Anderson et al., unpublished paper). This dynamism and flexibility are inherent features of any digital research infrastructure.

CENDARI as a digital “ecosystem” is distancing its infrastructure from the model of the infrastructure as a “portal” to archival and library sources. Moreover, it is defining its model as a meeting place that reflects both dynamism and production of new knowledge. Current digital developments seem to transpire (and generate) a deep divide in relations between archivists and historians with both groups insisting on specific – scientific – grounds and purposes.

The overcoming of “physical space” created between researchers and archivists due to their increased interaction through technology is at the core of endeavours behind the construction of individualised virtual research environment. It needs to be stressed that a VRE is designed to help researchers in all disciplines to cope with the growingly complex tasks involved in conducting research (cf. Edwards et al., 2012). In this regard “a virtual space” has the function of “containing or providing access” to information about a specific source or a category of sources (cf. *ibid*), thus acting as an indispensable communication field between researchers and their subject matter.

Joshua Sternfeld has introduced a framework called “digital historiography”, which he described as the “critical, interdisciplinary study of the interaction of digital technology with historical practice” (Sternfeld et al., 2014). In his view,

archival theory and practice need to inform and substantiate such a “critical framework” as well as further develop historiographical and professional practices (ibid.). Furthermore, the digital medium has given an incentive to historians to broaden their knowledge about archives and to see their function in producing scholarship and knowledge.

The digital turn has been decisively remoulding historical methods, theory and pedagogy. The historian of today has access to immense amounts of diverse evidence, with the chance to exploit freely multimedia sources such as digitized corpora from antiquity to the present, film and audio recordings, as well as vast digital resources such as websites, artwork and computational data. The task of historians has always been to critically evaluate what information has value as evidence. The range of digital tools available today, along with a “sheer size and scope” of online sources, call for a new “level of methodological rigour” that historians still need to exercise and become accustomed to (Sternfeld, 2014).

Digital media have been vigorously expanding new lines of enquiry. Historians and archivists are beginning to use and examine sources differing in scope, format and provenance, which has prompted them to reconsider traditional limits or modes of scientific inquiry, the “contextualising properties of metadata”, the construction of access systems, and ultimately, the engagement with new, digitally-aware audiences (Sternfeld et al., 2014).

5. Conclusion

Historical enquiry in a digital environment has unambiguously redefined the traditional research method, hitherto limited to analogue finding aids and materials dispersed in various national or state repositories. New digital methods apply most particularly to the early stages of research in that they help determine historians’ project strategy as well as subject of their enquiry. Production of new knowledge derives from vigorous (re)interpretation of first hand material and pertains to both WWI and MM domains but in the same vein calls into view their relation to archival generation of (meta)data and the need to

transcend the divide between researchers and archivists. The envisaged CENDARI VRE is acting as a necessary or highly original meeting point between digital collections and historical practice, thus contextualising the concepts of time and space, data, and points of access to digital surrogates.

6. Cited bibliography

Anderson, Sheila et al., 'The CENDARI Knowledge Framework: Bridging the Divide', an unpublished article.

Beneš, Jakub, 'Seing Archival Data in Online Portals: Challenges and Opportunities for Research', an unpublished presentation at the workshop *Access Digital: nutzen und vernetzen*, 25 November 2014, Koblenz.

CENDARI Summer School feedback forms. Internal project documentation.

Edwards, Paul et al, 'Our Spaces – Design and Deployment of a Semantic VRE', in Cudré-Mauroux, P., Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Parreira, J., Hendler, J., Schreiber, G., Bernstein, A. and Blomqvist, E. (Eds.), *The Semantic Web – ISWC 2012*, Springer-Verlag, Berlin Heidelberg, pp. 50-65.

Gartner, Richard and Hedges, Mark, 'CENDARI: Establishing a digital ecosystem for historical research', Vol. N/A N/A. ed. IEEE, 2013. p. 61-65.

Hobsbawm, Eric, *Age of Extremes: Short 20th Century 1914-1991*, London 1994

Howard, Michael and Roger, William Louis, *The Oxford History of the Twentieth Century*, Oxford 1998, 9.

Piez, Wendell, 'Something called "DH"'. *Digital Humanities Quarterly*, 2 (January 2008) (digital source).

Rosenzweig, Roy, "Scarcity or Abundance? Preserving the Past in a Digital Era." In *The American Historical Review* 108 (June 2003): 735–762.

Sternfeld et al., 'Digital Historiography and the Archives. *Journal of Digital Humanities*, 3 (2014) (digital source).

Sternfeld, Joshua, 'Historical Understanding in the Quantum Age. *Journal of Digital Humanities*, 3 (2014) (digital source).

Warwick, Claire, Melissa Terras, and Julianne Nyhan, eds., *Digital Humanities in Practice*. London, 2012.

