Virtual Research Environments
MADDLAIN Final report

May 2017

Stéphanie Paul

1. Overall aims and methodology

The development of new technologies had a dramatic impact on research practices in the Humanities. On the one side, new forms of communication through the internet foster collaboration within wide-scale international and interdisciplinary projects. On the other side, tools and methods developed in the wake of the Digital Humanities support various stages of the research lifecycle, from information retrieval to data analysis and dissemination of findings. These infrastructures not only make the work of researchers easier, but also raise new research questions, and thereby promote scientific innovation.

Researchers in the Humanities form a significant part of the audience visiting archives and libraries. It was therefore only natural to include within the MADDLAIN project a section focussing on the needs and requirements of this particular user group in terms of digital tools and services. At the core of this section of the project were two main questions:

- How can the CegeSoma, the State Archives, and the Royal Library **improve their current services** to better meet the needs of researchers in the Humanities?
- What kind(s) of **virtual research environments** (VREs) can these institutions develop to encourage the dissemination of knowledge, promote their collections, and foster scientific collaboration, while fulfilling their traditional missions?

The methodology adopted for this section of the MADDLAIN project can be summarised in two main steps:

1. **State of the Art**. An extensive literature review on research infrastructures and virtual research environments was conducted in order to elaborate a working definition of VREs and gain an overview of existing platforms and projects. Those were reviewed in order to identify best practices and potential issues. Simultaneously, a number of research projects promoted by MADDLAIN institutions and aimed at developing digital tools and environments were reviewed. 7 semi-structured interviews conducted with staff members who were involved in these project provided insight into the challenges associated with this kind of developments, as well as into their use of digital tools in their day-to-day research practices.
2. **User requirements analysis**. The aim of this section was to identify the target audience – i.e. draw the profile of researchers who use the collections of MADDLAIN institutions (see section 2 of this chapter) – and their needs. For this purpose, a series of 15 semi-structured interviews were conducted with

researchers from Belgian universities and research institutes. A literature review on research practices in the Humanities and on similar surveys was conducted in order to help drawing up a questionnaire and to view findings in a broader perspective.

A final section of this report will present the main conclusions of the study and suggest recommendations to optimise online services in support to scientific research.

2. Researchers as an audience of MADDLAIN institutions

According to their mission statements, the CegeSoma, the State Archives, and the Royal Library consider scientific research as an important part of their actions towards promoting the collections that they preserve. This section will explore in how far professional researchers represent a specific audience of these institutions. In this respect, no formal documentation provides much information about the estimated proportion of researchers who visit their reading rooms and websites, or about their academic profiles. During interviews conducted in 2015 as part of the MADDLAIN project, staff members were asked to give some insight into their users' profiles. Their views can be corroborated to a certain extent by the findings of MADDLAIN's main user survey[1], as well as by the data collected from reader's cards and other sources provided by the institutions.

2.1. The CegeSoma

The CegeSoma defines itself as a "research and documentation centre (…) aiming to function as a platform for scientific and social activities, where both researchers and a generally interested public are involved." One of its main role is to support, conduct, and disseminate scientific research on 20th-century conflicts. In this respect, the CegeSoma strives to be "a meeting place and portal for research and researchers in Belgium in the field of contemporary history" and to be a part of research and academic networks both in Belgium and abroad.[2]

About 47% of respondents to the user survey declared using the collections of the CegeSoma for professional reasons. Among those, about 42% are researchers at a university or at a research institute. According to staff members, these are mostly historians, art historians, and scholars specialised in information and communication sciences.

[1] The results from the User survey concerning the user profiles should be considered as merely indicative, since random sampling was not used in the selection of the participants and we cannot know whether the percentage of respondents is fully representative of their user group.
[2] http://www.cegesoma.be/docs/media/Divers/missionstatement_ang.pdf [accessed 2017.05.18].

2.2. The State Archives

The mission statement of the State Archives includes the valorisation of their curated collections through scientific research on archival work and on the history of Belgium. One way to achieve this goal is to participate in various research projects in collaboration with Belgian universities and other scientific federal institutions.[3]

Professional researchers do not represent the majority of the public of the State Archives. About 24% of respondents to the user survey use the collections for professional reasons, and among those, about 44% are researchers at a university or at a research institute. However, it should be noted that the picture considerably differs depending on the location. Staff members pointed out that professional researchers formed the main category of visitors to the General Archives in Brussels, whereas they rarely visit the repositories in the provinces, a representation which is corroborated by previous studies on user profiles. This is due to the type of collections held in the provinces, which are more of interest to genealogists and amateur historians doing research on local history. As an exception to the rule, the State Archives repository in Ghent welcomes a high number of academics (university students, doctoral or postdoctoral researchers, professors), which can be explained by the proximity of Ghent university and collaborations between the staff of both institutions.[4]

As far as disciplines are concerned, the data collected from reader's cards do not provide a full picture. According to staff members, the main discipline is history, followed by art history, legal history, and political sciences.

2.3. The Royal Library

The mission statement of the Royal Library includes the responsibility to conduct scientific research in the areas of information sciences and librarianship, as well as to participate in research projects on cultural heritage which focus on their collections.[5]

About 42% of respondents to the User survey declared using the collections of the Royal Library for professional reasons, among which about 39% are researchers at a university or at a research institute. According to staff members, the public of the Royal Library mainly consists of university students and academics. Whereas the former often come to the library to study without really using the collections, the latter are especially interested in the specialised collections (Newspapers, Manuscripts, Maps and Plans, Prints, Coins and Medals, etc.).

[3] http://www.arch.be/index.php?l=en&m=about-the-institution&r=our-mission#6 [accessed 2017.05.18].
[4] Verachten 2011; Deschamps 2010. These reports are based on the information provided by readers on their registration form. During the years 2004-2008, 33.85% of visitors to the General Archives were professors/teaching assistants in universities, 22.64% for the State Archives in Ghent, against 8.52% in total (all the repositories).
[5] http://www.kbr.be/en/mission-and-vision [accessed 2017.05.18].

As far as disciplines are concerned, they are mainly researchers in the Humanities and Social Sciences, partly because the Royal Library has been pursuing an acquisition policy restricted to these subjects. In this case, the data collected from the institution allows to obtain a more detailed picture of their research areas. If they wish, researchers may access, upon request, a special reading room within the library. To do so, they are only required to send an email in advance stating the purpose of their visit. Although this data is not fully representative of all the researchers who use the collections of the Royal Library, since they do not necessarily request access to this space, it still provides an overview of users' various research profiles.

Figure 1: Visualisation of disciplines of researchers using the work room at the Royal Library (Voyant Tool)

2.4. Summary

The MADDLAIN institutions all consider scientific research as an important way to promote their collections. To this end, they are involved in various research projects and hire a number of researchers who should therefore also be taken into account as users. Although researchers do not necessarily form the major part of the audience using the collections of the three institutions, they form the major part of the audience using the collections for professional reasons, in addition to being extremely regular users. Historians are the main category of researchers who use archives as evidence for their research, whereas the professional audience of the Royal Library is more diverse, albeit mostly falling within the scope of the Humanities and Social Sciences. As far as academic positions are concerned, the data collected from the Royal Library suggests that whereas all stages of a researcher's career are represented, PhD students are a more prominent group. The data from the CegeSoma and the State Archives does not give any indication on this aspect.

1. An evolving scientific landscape

Over the past two decades, many initiatives and projects were launched towards the development of virtual research infrastructures at a variety of scales. These result from significant changes occurring within the scientific landscape. On the one hand, advances in technologies are creating an unprecedented volume of scientific data to be processed. This phenomenon is known as data deluge or infobesity.[6] Consequently, one of the main challenges that scientists face nowadays is not so much to acquire data, but to find new ways to analyse it (Voss and Procter 2009). Furthermore, the development of new analytic tools brings a fresh perspective to research data and leads to ask new research questions (ACLS 2006).

On the other hand, the research landscape itself has changed dramatically. To tackle complex research tasks and big data analysis, large-scale international and multidisciplinary collaborations between several research teams increasingly become standard (Llewellyn Smith *et al.* 2011). Advances in communication technologies facilitate such collaborations, and virtual collaborative environments, for instance, are required for research teams to carry out their work in an efficient way, for instance to share their data and tools. At the individual level as well, web-based forms of communication and digital tools support day-to-day research activities and administrative tasks (Brunvand and Duran 2010).

1.1. National and international initiatives on e-research

From the years 2000 onwards, national and international programmes were launched to ponder the consequences of changing patterns in science and to explore ways to build large research infrastructures and virtual research environments.

In this respect, the UK appears to be a pioneer. In 2001, the UK Research Council launched its e-Science programme to develop grid technologies, research software and e-infrastructures to support research in all disciplines.[7] At the core of this project were concerns in terms of "producing, managing and preserving vast amounts of digital data", "accessing an ever-increasing range of electronic resources", "supporting virtual

[6] The website "How much information?" publishes a study conducted in 2003 as "an attempt to estimate how much new information is created each year": http://www2.sims.berkeley.edu/research/projects/how-much-info-2003. The Economist published in 2010 a Special report "Managing information on the abundancy of data": http://www.economist.com/node/15557443.
[7] http://www.rcuk.ac.uk/research/xrcprogrammes/prevprogs/.

communities of researchers", and providing "unprecedented network, grid and computational capacity", as well as "safe and secure access to resources" (Pothen 2007). The e-Science programme is believed to have made significant contributions in terms of software and hardware building, access to information and data curation, enhancing training and cross-disciplinary collaborations. It also led to the development of the National Grid service and the Digital Curation Centre (Atkins *et al.* 2009).

In 2006, the Office of Science and Innovation (OSI) commissioned an e-Infrastructure Working Group to publish a report on Virtual Research Communities. These are defined as "a group of researchers, possibly widely dispersed, working together effectively through the use of information and communications technology". As part of the conclusions of this report, the working group strongly encouraged to continue Virtual Research Environments development programmes, envisioned here as a support of the Virtual Research Community and defined as "a set of online tools, systems and processes interoperating to facilitate or enhance the research process within and without institutional boundaries".[8]

From 2004 to 2011, the Joint Information Systems Committee (JISC) funded such a programme to develop virtual research environments.[9] VREs' purpose was defined as a way "to help researchers from all disciplines to work collaboratively by managing the increasingly complex range of tasks involved in carrying out research." Divided in three phases, the programme explored the definition and technological solutions for VREs, implemented pilot projects, and eventually developed ten VREs and tools, which laid considerable groundwork for the development of VREs afterwards (Carusi and Reimer 2010). As a result, several reports and a how-to guide were published.[10] In 2016, JISC launched a co-design consultation around six challenge areas for future development, among which was the question "What should a next-generation research environment look like?", thus pursuing previous work on this topic.[11]

In the United States, the so-called Atkins report published in 2003 is a landmark in terms of evaluating needs and opportunities to build cyberinfrastructures to facilitate scientific research. Cyberinfrastructure is here defined as a "layer of enabling hardware, algorithms, software, communications, institutions, and personnel" lying between "integrated electro-optical components of computation, storage, and communication" and "software programs, services, instruments, data, information, knowledge, and social practices applicable to specific projects, disciplines, and communities of practices." Such a layer aims to "provide an effective and efficient platform for the empowerment of

[8] http://www.nesc.ac.uk/documents/OSI/vrc.pdf
[9] The description of the JISC programme has been archived at: http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/programmes/vre.aspx
[10] "Implementing a virtual research environment. Understanding the tools and technologies needed by researchers": https://www.jisc.ac.uk/full-guide/implementing-a-virtual-research-environment-vre.
[11] https://www.jisc.ac.uk/rd/get-involved/what-should-a-next-generation-research-environment-look-like

specific communities of researchers to innovate and eventually revolutionize what they do, how they do it, and who participates" (Atkins *et al.* 2003). Consequently, the National Science Foundation funded from 2004 to 2011 the TeraGrid Science Gateways Programme, in order to develop gateways for science and engineering based on the Grid technology (Wilkins-Diehr *et al.* 2008).

In 2002, the European Commission established the European Strategy Forum on Research Infrastructure (ESFRI) to answer the challenge of European competitiveness in a globalized world and to encourage scientific innovation. The aim of this Forum is to "support a coherent and strategy-led approach to policy-making on research infrastructures in Europe, and to facilitate multilateral initiatives leading to the better use and development of research infrastructures, at EU and international level".[12] In 2006, ESFRI published a roadmap to assess scientific needs in terms of large-scale research infrastructures in Europe. This roadmap has been regularly updated since, the last version dating from 2016.[13] It adopts an all-encompassing definition for research infrastructure, which, "including the associated human resources, covers major equipment or sets of instruments, as well as knowledge-containing resources such as collections, archives and databases. Research Infrastructures may be 'single-sited', 'distributed' or 'virtual' (the service being provided electronically)."[14] Aside from ESFRI, several projects are funded within the section "e-infrastructure" of the European Commission Framework Programme for Research and Innovation Horizon 2020. For instance, the call for 2014-2015 invited applications of projects to develop VREs, which "are expected to result in more effective collaboration (…), higher efficiency and creativity (…), higher productivity (…) and to accelerate innovation."[15]

Aside from these large-scale, sometimes transnational, programmes, smaller national initiatives arose. Since 2008, the German *Allianz der deutschen Wissenschaftsorganisationen* is leading a priority initiative on "Digitale Information", which includes a section on the development of "Virtuelle Forschungsumgebungen", a German translation for virtual research environments.[16] The same year, the German Research Foundation started a funding scheme (DFG-VRE) to support the development of integrated environments to provide researchers with access to scientific data and information, as well as enhancing new forms of scientific communication (Lipp 2009). More recently, the DFG supported a two-year project (2013-2015) coordinated by the Research and Development Department at Göttingen State and University Library, entitled "Metrics of Success for the Planning and Sustainable Operation of Virtual Research Environments (DFG-VRE)" (Buddenbohm *et al.* 2015).[17] As of December 2015, the DFG-VRE programme has been integrated into the new funding scheme "e-Research

[12] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-background.
[13] http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap.
[14] http://www.esfri.eu/esfri_roadmap2016.
[15] http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-infrastructures_en.pdf.
[16] http://www.allianzinitiative.de/.
[17] https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/dfg-vre-1/.

Technologies". The French government periodically updates a roadmap to present national large-scale research infrastructures, be they material or virtual, the last version of which dates from 2016. These infrastructures aim to support researchers in processing their sources, as well as to create new modes of scientific publishing.[18] In the Netherlands, SURF, a collaborative organisation for ICT in education and research coordinates projects to build clouds, computing and data infrastructures, virtual research environments, and so on.[19]

1.2. The challenge of open science

In addition to support researchers evolving practices and facilitate scientific collaboration, research infrastructures aim to address a major challenge for the future of science, which is open science.[20] Significantly, open access, i.e. free access to scientific publications funded by public funds, is clearly one of the main challenges expressed in a 2006 report on cyberinfrastructures in the Humanities and Social Sciences, which considers scholarly publications as "public goods" (ACLS 2006: 30; 37). Along the same line of thought, open data implies that scientific data should be freely available for reuse, and building virtual research infrastructures would help to meet this challenge. As Voss and Procter put it, there is an increasing "need to document the research process more thoroughly, keeping and curating the resources consumed and generated so that they are discoverable and re-usable by others" (Voss and Procter 2009: 175).

2. Research infrastructures in the Humanities

For the last two decades, many digital humanities projects aimed at building databases and digital collections of documents. However, most of these projects do not go beyond the purpose of serving the research they were originally built for. Moreover, a lot of resources were directed at developing technologies that perhaps had already been developed elsewhere, hence a tendency to reinvent the wheel (Dombrowski 2014:327). This results in a compartmentalisation of most projects, due to a lack of interoperability, and in a duplication of efforts. In response to these issues, the building of consortiums and networks as well as the development of large-scale research infrastructures aim to bring greater standardization and mutualisation of skills and resources.

The debate around the development of research infrastructures in the Humanities started as early as the years 2005-2006 with the publication of two special reports: the British Academy's 2005 *E-resources for Research in the Humanities and Social Sciences – A British Academy Policy Review*, or in 2006, *Our Cultural Commonwealth. The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the*

[18] http://www.enseignementsup-recherche.gouv.fr/cid70554/la-feuille-de-route-nationale-des-infrastructures-de-recherche.html.
[19] https://www.surf.nl/en/innovationprojects.
[20] See for instance the 2016 publication by the European Commission "Open Innovation, Open Science, Open to the World".

Humanities and Social Sciences for the United States, which particularly emphasises how technologies could contribute asking new research questions on cultural heritage collections (p. 22). In 2011, the European Science Foundation contributed to the debate by publishing a report on *Research Infrastructures in the Digital Humanities*.

Simultaneously, multiple projects and initiatives were launched to further investigate possibilities to develop research infrastructures and virtual research environments in the Humanities. In the United States, Bamboo was an ultimately unsuccessful project funded by the Andrew W. Mellon Foundation "in order to enhance arts and humanities research through the development of infrastructure and support for shared technology services" (Ibid.: 326). The European Commission started a series of projects, initiatives and consortiums as part of the European Science Forum on Research Infrastructure:

- DARIAH (Digital Research Infrastructure for the Arts and the Humanities) defines as "a network of people, expertise, information, knowledge, content, methods, tools and technologies". It is organized around several working groups pertaining to four Virtual Competency Centers, one of them focusses on e-infrastructures, including data repositories, tools and digital research environments.
- CLARIN (Common Language Resources and Technology Infrastructure) "provides easy and sustainable access for scholars in the humanities and social sciences to digital language data (in written, spoken, or multimodal form), and to advanced tools to discover, explore, exploit, annotate, analyse or combine them, wherever they are located." It is organised around a network of repositories and research centres. As part of services provided, a VRE entitled Virtual Language Observatory was recently implemented.
- E-RIHS (European Research Infrastructure for Heritage Science) "supports research on heritage interpretation, preservation, documentation and management". It forms a network of institutions related to cultural heritage and aims to provide scholars with access to infrastructures, methodologies, data and tools. The project is currently in a preparatory phase and will be implemented in 2022.
- CENDARI (Collaborative European Digital Archive Infrastructure) aims to develop a digital toolkit for historians working with archives. Focussing on the medieval period and WWI, the project developed an archival directory enabling researchers to access historical sources and a working space, called note-taking environment, allowing scholars to upload and organise their documents, and annotate images and texts collaboratively or privately.
- EHRI (European Holocaust Research Infrastructure) developed a portal to support research in the Holocaust by providing searchable information on related archival sources and facilitating collaboration between scholars.

In Belgium, the Ghent Centre for Digital Humanities is currently leading a project (VRE-SI) funded by the Research Foundation Flanders (FWO) under the auspices of the DARIAH network, towards the development of a virtual research environment in close

cooperation with five pilot research projects in digital humanities. The aim is "to offer a sustainable portfolio of services enabling digital scholarship in the arts and humanities in Flanders, Belgium and beyond."

Despite those initiatives, research infrastructures and virtual research environments in the Humanities tend to experience a slow uptake by researchers (Carusi and Reimer 2010: 12). For the last decades, the wide-scale digitisation of heritage collections allowed researchers to experiment with computationally-based methods, such as quantitative analysis, text mining, named entity recognition, and so on, but there is still some scepticism as to the relevance of the concept of "big data" for the Arts and the Humanities (Anderson and Blanke 2012: 151, 155; Anderson 2013: 6-7). Some researchers also doubt the innovative character of research questions that can be applied to this material and whether dedicating a great amount of money to fund large-scale infrastructure is a good investment (van Zundert 2012: 168). Another objection concerns the nature of the sources itself. Research in humanities is based on a diversity of documents, contexts and research methods. Sources can be used by specialists coming from different fields and context is at risk to be lost through the digitisation process (ACLS 2006: 25-26). All-encompassing and generic research infrastructures might therefore miss the point of most actual researchers' needs (van Zundert 2012). Finally, the strong conservative culture in the Humanities, where collaborative work is not the norm, may explain the reluctance to adopt these new tools (ACLS 2006: 21).

Should those challenges be overcome, research infrastructure could provide an opportunity to link the scholarly community with cultural heritage institutions such as archives, libraries, and museums. Those institutions could be key stakeholders in the conception and building of research infrastructures for the Humanities not only as provider of content, but also through their valuable knowledge and experience in information sciences and data curation (Ibid.: 18; Carusi and Reimer 2010: 6, 44; Anderson 2013; Speck and Links 2013).

3. Virtual research environments: a working definition

The concept of virtual research environments originate from virtual learning environments, web-based platforms to support teaching, now widely used in higher education (Yang and Allan 2010: 72-76). A literature survey on environments built as a support to scholarly work brings to light a varied terminology:

1. One trend is based on the notion of infrastructure: research infrastructure, cyberinfrastructure (especially in the US), or e-infrastructure.
2. Another trend emphasises the notion of collaboration: collaboratory, collaborative virtual environment, collaborative e-research community, virtual research community.

3. In addition to virtual research environment, we also find science gateway, digital library, inhabited information space, virtual workspace, virtual organisation, or virtual laboratory/observatory (especially in STEM).

While definitions often overlap, these terms do not all refer to the same reality. Collaborative e-research community, virtual research community, and virtual organisation designate the community of practice using the VRE rather than the environment itself (Carusi and Reimer 2010:13). On the other hand, digital libraries are digital collections of documents lacking the required collaborative aspect of VREs. The concept of e-infrastructure (or cyberinfrastructure, research infrastructure) goes beyond the VRE. According to the definition of the UK Research Council, "e-infrastructure refers to a combination and interworking of digitally-based technology (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks) and the people and organisational structures needed to support modern, internationally leading collaborative research."[21] Virtual research environments can therefore be one of the resources provided by such an infrastructure, as a domain-specific web-based interface that provides access to data and services (Carusi and Reimer 2010: 14; Koureas et al. 2016: 5). In short, VREs are a component of e-infrastructures, but do not equate with it.

According to the definitions found in the literature, the main features of virtual research environments can be summarised as follows:

1. VREs support every step of the research workflow, from grant application to dissemination of results. They provide services related to administration, project management, data collection, annotation and analysis, networking, etc. (Yang and Allan 2010: 66-67).
2. VREs enhance collaboration between researchers. Collaboration beyond the national and disciplinary borders is one of the main challenges addressed by VREs (ACLS 2006: 35; Voss and Procter 2009: 175; Yang and Allan 2010: 66; Carusi and Reimer 2010: 12; Anderson and Blanke 2012: 154; Candela, Castelli and Pagano 2013: GRDI78; Dovey 2015: 47).
3. VREs should be user driven. VREs are "tailored to serve the needs of a community of practice" (Candela, Castelli, and Pagano 2013: GRDI75). They can be either very specific, aimed at a particular research project, or provide very generic functionalities (Voss and Procter 2009: 176; Dovey 2015: 47). They should not be built for the technology's sake, but rather with the researchers' needs in mind (Carusi and Reimer 2010: 43; Candela, Castelli, and Pagano 2013 :GRDI79). A one-size-fits-all approach will not work, VREs need to be flexible and customizable, and to integrate a variety of tools and components (Voss and Procter 2009: 176; Carusi and Reimer 2010: 5; Buddenbohm *et al.* 2015).

[21] http://www.rcuk.ac.uk/research/xrcprogrammes/otherprogs/einfrastructure/.

4. VREs should be interoperable, a necessary requirement to conduct multidisciplinary research projects (Atkins et al. 2003: 13).
5. VREs should guarantee security, authentication, and ownership (Candela, Castelli, and Pagano 2013: GRDI76).

4. Main challenges

The development of virtual research environments potentially raises several issues. One of the main challenges is undoubtedly uptake by researchers. Those can sometimes be reluctant to adopt VREs, for various reasons:

1. Researchers do not see the point of using such environments, because they are not convinced of the time and productivity gain as opposed to their traditional practices. There may also be a discrepancy between the rhetoric used by funding bodies to describe the benefits of VREs in terms of "excellence", "creativity", "innovation", "competitiveness", and the actual needs of researchers (Anderson 2013: 7). From this point of view, it is important to adopt a bottom-up, rather than top-down, approach, in order to ensure that VREs do indeed correspond to what researchers need, as opposed to build VREs for the technology's sake and assume that users will automatically come (Anderson 2013: 9-10).
2. VREs need the engagement of users in order to be viable and useful (Brunvand and Duran 2010: 115), but users are not willing to engage in an environment whose sustainability is not guaranteed. Sustainability is therefore one of the main challenge in building a VRE (Carusi and Reimer 2010: 5). Given the fast pace of innovations in technology and the long time required to plan and build such an environment, there is a high risk that VREs become out-of-date even before their implementation. In this respect, agile software development methods could allow to respond to changes faster and more easily (van Zundert 2012).
3. VREs are not user-friendly or are too complicated for non-experienced users (Dovey 2015: 50; Koureas et al. 2016: 3).
4. Researchers are uncertain of how secure VREs are. Among other things, they fear that ownership of their work will not be guaranteed Ibid..
5. Finally, issues of rights and copyright limitations may hinder access to data (ACLS 2006).

5. Classification and examples

For the purpose of this report, a classification of virtual research environments with a focus on the Humanities was elaborated (see appendix). The aim is not to provide a complete overview of all VREs in the field, but rather to show a few examples of existing projects. Four categories were distinguished:

1. Project management platforms, which are rather generic and could potentially be used by researchers from all disciplines. These usually provide researchers with storage and collaborative features.
2. Platforms focused on a particular discipline, such as digital humanities, art history, or history.
3. Platforms focused on a research activity, which are built to enable researchers to perform a specific task, such as editing or transcribing texts.
4. Platforms focused on a research project, which are highly customised environments built around a particular collection of documents and integrating particular tools to answer a specific research question. They are usually restricted to members of the research team.

6. Tools and platforms developed by MADDLAIN institutions

A number of research projects in which the CegeSoma, the Royal Library, and the State Archives are involved aim to develop virtual platforms and other kinds of digital tools. A review of those project was undertaken in the context of the MADDLAIN project, with a focus on the role of the institution in the development process (see table below). It is based on official documentation published on websites, as well as on semi-structured interviews conducted with members of staff involved in the project.

Project	**Tool**	**Content**	**Institution**	**Role**
BELGIUM WWII	website	online encyclopaedia	State Archives	development manager (outsourced)
			CegeSoma	
CARTESIUS	portal	digitised collection	State Archives	content provider
	virtual lab		Royal Library	content provider
EHRI	portal	metadata	CegeSoma	content provider
IMMIBEL	database	metadata	State Archives	development manager (existing tools)
				content provider
	website	valorisation of the project		development manager (existing tools)
JUSINBELLGIUM	database	metadata	State Archives	development manager (existing tools)
		digitised collection		content provider
TIC-Belgium	VRE	digitised collection	State Archives	content provider

Two types of projects are identified:

1. The main purpose is to build a digital tool (BELGIUM WWII, Cartesius, EHRI);
2. The tool serves a specific research question at the centre of the project (IMMIBEL, JUSINBELLGIUM, TIC-Belgium).

Most frequently, the institutions act as content provider and are not involved in the technical development. When it is the case, the development is either outsourced or based on existing software, such as Excel or CMS web platforms. The ease of use is often emphasised as an important requirement in those projects. The goals behind the implementation of digital tools are most frequently:

- Facilitating access to documents;
- Enhancing innovative and interdisciplinary research, as well as generating new research questions;
- Supporting a research project.

Other motivations include developing tools to valorise heritage material and disseminate scientific knowledge to a wide audience.

Staff members pointed out a number of challenges in the design and development of digital tools, such as:

- Technical difficulties and bugs in the development process;
- Collaboration with external service providers / dialogue between IT-specialists and researchers;
- Metadata entry as a time-consuming process;
- Lack of standardisation in archival descriptions;
- Access to scattered sources;
- Access management with regard to privacy laws, copyright, and ethics;
- Bilingualism;
- Finding the right way to communicate in order to reach the target audience;
- Short timeframe and limited funding of research projects.

The dynamics of short-term projects and lack of funding were especially viewed as an ongoing problem with regard to sustainability, future updates and developments of the tools. One of the solutions that has been implemented to ensure the sustainability of the project results is to integrate the dataset within the institution's existing infrastructure, namely search engines, whose maintenance will be guaranteed on a longer term.

1. Conducting a qualitative survey: aims and methodology

From November 2016 to January 2017, a user requirements survey was conducted among professional researchers from Belgian universities or research institutions. The survey had for main purposes to:

- Evaluate how professional researchers use the online services of MADDLAIN partner institutions (State Archives, CegeSoma, Royal Library), and what is their satisfaction with those services.
- Define researchers' needs and requirements in regards to the improvement of current services, as well as to the implementation of virtual research environments and new resources to support research in the Humanities.

Although defining the overall aims for the user research was obvious, the question of how to achieve those aims was less straightforward. For users, defining their needs in terms of digital services is not necessarily self-evident. First, researchers may not have thought deeply about their digital research practices, which are sometimes integrated into their workflow in a very intuitive way. Furthermore, they may not be aware of all the existing possibilities that can be implemented in terms of online services and tools.

One way to get around this problem has been to reflect on how researchers work, to understand in what stage of their research they use online resources, how they use them and for what purpose. In other words, it was necessary to get a better picture of researchers' user experience to identify their needs. Only then can we find ways to improve current online services and implement new ones.[22]

1.1. Conceptual Framework

Over the last two decades, scholars have attempted to conceptualise research activities in studying scholarly information behaviour in the light of new technologies.[23] Some of this work aimed to target the needs of researchers in the arts and the humanities as preliminary reflection to develop online research infrastructures and tools. Within this framework, Unsworth introduced the concept of scholarly primitives, defined as "basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation." These primitives were identified as *discovering, annotating, comparing, referring, sampling, illustrating, representing* (Unsworth 2000). A report commissioned by the *Online Computer Library Center* (Palmer, Teffeau and Pirmann 2009) further explored the concept of scholarly primitives, emphasising "the explicit

[22] For the interest of user studies in digital humanities, see: Warwick 2012; Kemman and Kleppe 2015.
[23] For a short literature review on this subject, see Benardou *et al.* 2010.

role of information in the conduct of research and production of scholarship (p. 7)." This report aimed to compile a literature review on research practices to provide academic libraries with guidelines for the development of services that would meet their users' needs. It identifies five core scholarly activities, *searching, collecting, reading, writing, collaborating*, which encompass in turn several scholarly primitives. In addition, four cross-cutting primitives, *monitoring, notetaking, translating, data practices*, can potentially be linked to all the above activities (Figure 2: Ibid., 36.).

Table 2: The five core scholarly activities and their primitives

1. Searching	4. Writing
1.1 Direct searching	4.1 Assembling
1.2 Chaining	4.2 Co-authoring
1.3 Browsing	4.3 Disseminating
1.4 Probing	
1.5 Accessing	5. Collaborating
	5.1 Coordinating
2. Collecting	5.2 Networking
2.1 Gathering	5.3 Consulting
2.2 Organizing	
	6. Cross-cutting Primitives
3. Reading	6.1 Monitoring
3.1 Scanning	6.2 Notetaking
3.2 Assessing	6.3 Translating
3.3 Rereading	6.4 Data Practices

Figure 2: Ibid., 36.

With its focus on information retrieval, this model has proved itself particularly useful as conceptual framework for conducting user research before developing online services.[24] Figure 3: Scholarly activities and online services is an attempt to demonstrate how scholarly activities can be translated into specific tasks to which correspond in turn online tools and services.

[24] For some examples of a similar model applied in the context of infrastructure development, see Anderson, Blanke and Dunn 2010; Benardou *et al.* 2010; Blanke and Hedges 2013.

SEARCHING	Direct searching	Keywords search to find a specific source or reference	Library catalogues Search engines
	Chaining	Go from one reference to another	Links Suggestions of similar docs
	Browsing	Browse through the documents of a collection	Browse interface Virtual shelves Newly added documents
COLLECTING	Probing		
	Accessing	Visualising documents	Digital library/viewer
	Gathering	Collecting documents or references in a personal collection	Download Photograph/copy Export citations
	Organising	Manage a personal collection	Ref management tool Database
READING	Scanning	Rapidly go through document content	Full-text search
	Assessing	Understand context, assess relevance Compare similar documents	Comparison tool
	Rereading	Read through documents several times	User account to save docs
WRITING	Assembling	Drawing together ideas, notes, documents	Mind mapping tools Word processing Writing tools
	Co-authoring	Write research outputs collaboratively	Collaborative writing tools
	Disseminating	Publish research outputs	Repositories Publishing platforms
COLLABORATING	Coordinating	Manage research projects	Project management tools Tasks management Clouds
	Networking	Communicate with colleagues	Social networks
	Consulting	Consult professionals for advice on sources	Forum/chat/contact form
Monitoring		Review new information on a regular basis	Newsletters Social networks RSS feeds
Notetaking		Taking notes on documents	Annotation tools Virtual notebooks
Translating		Translating documents to another language Transcribing documents in computer language	Online dictionaries Transcription tool
Data Practices		Share data	Data repository Cloud

Figure 3: Scholarly activities and online services

All tools and services are not equally relevant with respect to the traditional roles of heritage institutions. For the purposes of the MADDLAIN project, the user research should focus on those activities that concern the researchers' information behaviour. In other words, the aim is to explore how researchers in the humanities seek, gather, and manage information, and particularly how the use of digital resources and tools fit into their research workflow.

1.2. Choice of method

Ethnographic studies provide various methods to conduct a user experience research.[25] Among these, the qualitative survey via semi-structured interviews appeared to suit best the overall aims of the project. Conducting an online written survey was briefly considered, but this option was quickly ruled out. A written survey would have resulted in a higher response rate, thus providing a better representativeness, but its rigid format would not have allowed to gain an in-depth understanding of the users' information behaviour and research practices. On the other hand, semi-structured interviews based on open-ended questions provided a more flexible approach to the subject. This option allowed to cover larger themes and to leave enough room for researchers to reflect thoroughly on their practices. It also gave way to a necessary interaction between interviewer and interviewee. This was particularly useful for the topic of virtual research environments, which turned out to be rather abstract for some of the participants. In this case, interviews allowed to clarify the questions or give further explanations, whereas a written survey would probably have resulted in unsatisfactory responses.

1.3. Profile of participants

A study on the scientific audience of MADLDLAIN partner institutions concluded that it mainly consisted of researchers in the humanities and, to a lesser extent, social sciences. The recruitment of participants to the survey therefore targeted researchers in these areas, who regularly used the collections of at least one of the above institutions for their research. The survey aimed to include researchers in all stages of the academic career, from all three regions of Belgium (Brussels, Flanders, and Wallonia) with the purpose of achieving parity between Dutch-speaking and French-speaking scholars.

Possible participants were identified in various ways:

- Recommendations from promotors and collaborators of the MADDLAIN project;
- Recommendations from curators of specialized reading rooms;
- List of researchers having access to the work room of the Royal Library;
- Call for participants disseminated in a few university departments.

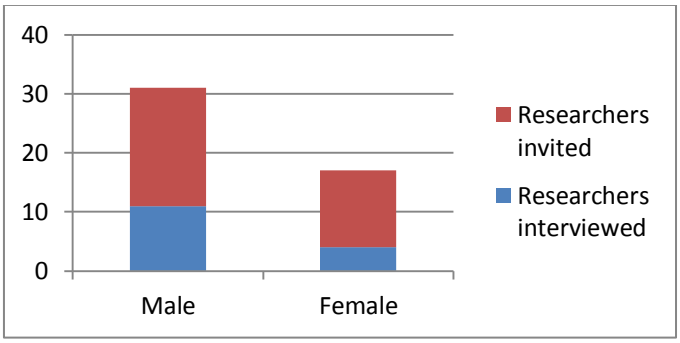
[25] The movement UXLibs in particular brought out innovative methods to study users in libraries: Priestner and Borg 2016.

The snowball technique, i.e. asking interviewees to recommend colleagues who might also be willing to participate, was ruled out here, as it presented the risk of having too many similar profiles, in addition to extend the timeframe of the survey. However, researchers were invited to suggest names in the case where they were to decline the interview request. Contacting researchers personally on behalf of someone was the most successful technique. No one answered the call for participants.

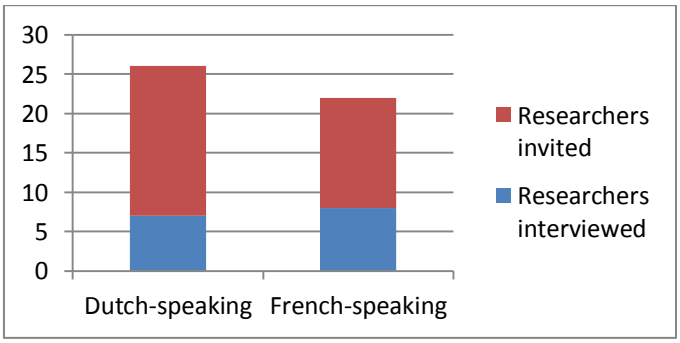
In total, 32 researchers were contacted.

- 10 did not answer,
- 4 declined,
- 2 accepted first, but then cancelled or did not follow up,
- 1 replied outside the timeframe of the survey,
- 15 accepted.

Although parity of gender and representativeness of affiliation, career stage, and discipline was a major concern of the survey, some hazards in the response rate could not be completely avoided. For instance, gender equality could not be achieved. In addition to the fact that recommendations targeted more male researchers, proportionately, those were the ones who more often responded positively (55% against 30,76%).



With respect to the native language of participants, equality is almost achieved:



As regards affiliation, the major Belgian universities are represented. In addition, the panel included one member from the scientific staff of each partner institution. As concerns the academic career, all positions are represented, although retired researchers were excluded from the sample:

- 3 PhD students
- 5 postdoctoral researchers
- 2 researchers in permanent positions
- 5 lecturers or professors

As concerns fields of study, history is substantially overrepresented:

Contemporary history	8
Modern history	2
Medieval history	1
Book history	1
Art history	1
Communication sciences	1
Literature	1

This overrepresentation can be explained by the fact that historians represent the major part of the academic audience of archives, but it should also be noted that the response rate from historians was considerably higher (61,11%) than for all the other fields included (26,67%):

	Invited	Accepted
History	18	11
Book history	2	1
Art history	2	1
Communication	3	1
Literature	4	1
Psychology	1	0
Educational sciences	1	0
Political sciences	1	0
Information sciences	1	0

As a consequence, responses will be strongly oriented towards the historical perspective, and this should be kept in mind for the interpretation of the findings.

1.4. Questionnaire

A questionnaire with mainly open-ended questions provided the basis for the interview and evolved along the way. It covered four main topics:

Research profile

This section aimed to collect contextual information on the researcher's position, research interests, and training in archival/library research. Participants were also

invited to give an overview of the sources they use for their research and where those sources are preserved.

Seeking, collecting, and managing information

This section aimed primarily to understand:

- How researchers identify and locate documents, either primary sources or secondary literature, and which digital resources they use for this purpose.
- How researchers access, collect, and organise their sources in their own digital ecosystem.

Furthermore, researchers were invited to give a large overview of the digital tools they use to support their research. This helped providing more general information on scholarly digital practices and gave an idea of the participants' digital literacy.

Digital services offered by the partner institutions

This section aimed to understand how researchers currently use the digital services of the three institutions and how satisfied they are with them. Participants were invited to give good and bad examples of digital resources offered by libraries and archives in general. This question was particularly useful to understand what makes a good resource in their perspective.

Virtual research environments and future developments

This section investigated whether researchers were familiar with the concept of virtual research environment and how far these environments would fit their needs. Participants were also invited to evaluate suggestions of new services that could be added to the catalogues or collaborative platforms. Some final general questions concerned how the digital turn impacted their research.

2. Findings

Fifteen face-to-face interviews were conducted over a period of two months, from November 2016 to January 2017. Interviews lasted on average one hour and a half. They were recorded with the consent of the participants, then fully transcribed. Significant excerpts were extracted from the transcriptions and coded in Dedoose, a qualitative data analysis software.[26] Codes were then revised, systematised and organised in wider themes.

Figure 4: Themes addressed in the survey illustrates the various topics (themes and codes) that were addressed in the interviews. This report will focus on themes that are most closely related to the purpose of the survey, such as:

1. Information behaviour and access to historical sources in a digital era.

[26] http://www.dedoose.com

2. Research practices in a digital era: time management, work environment, and digital turn.
3. Researchers as users of digital services offered by the State Archives, the CegeSoma, and the Royal Library.

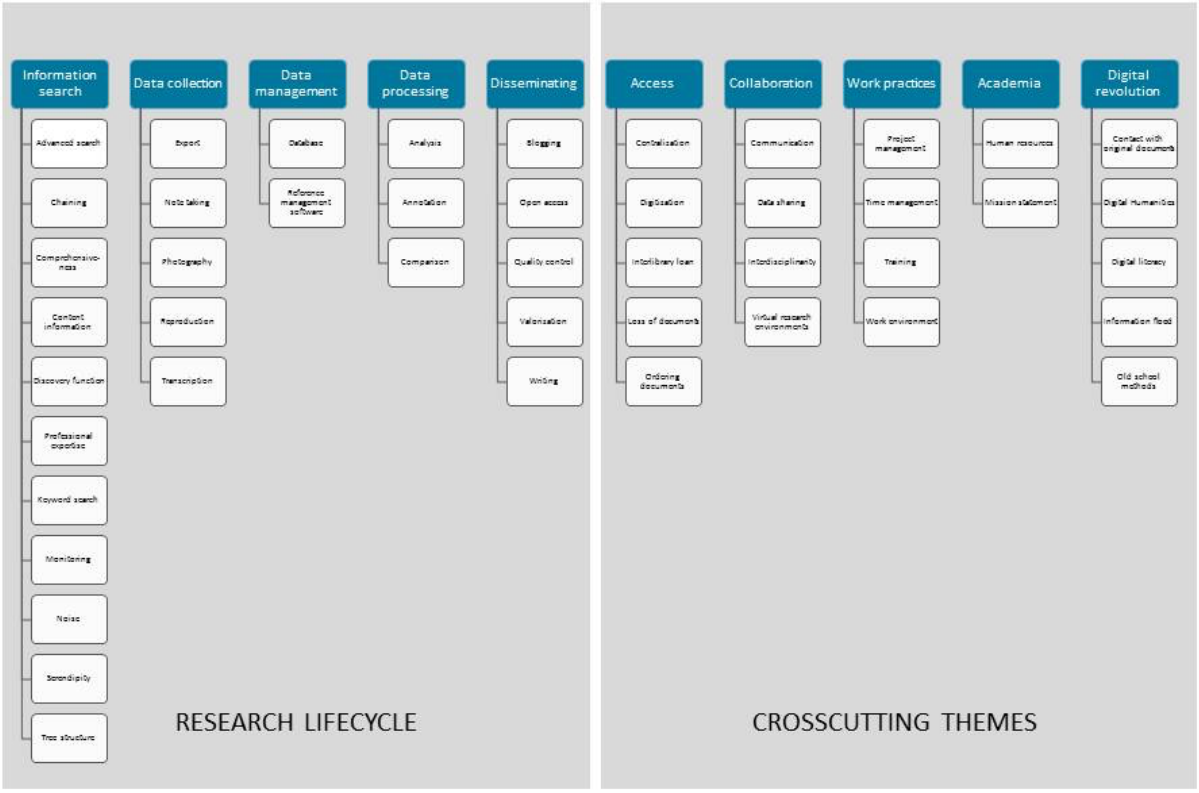


Figure 4: Themes addressed in the survey

2.1. Information seeking behaviour

Researchers use multiple strategies to search for information, whether primary sources or secondary literature (Figure 5: Information seeking strategies).

Figure 5: Information seeking strategies

Keyword search as a starting point

To initiate the search process, researchers regularly use keyword search in a selection of search engines, including Google, catalogues, and web-based publishing platforms. Once they have identified a few reference works or important sources, they proceed with more traditional techniques, such as browsing or citation chaining. Chaining consists in using one main information source to track relevant references in the bibliography or the footnotes, and so on. This technique is systematically used by researchers, who find it particularly effective, albeit old-fashioned.

> *"Q : Au niveau de la recherche bibliographique, est-ce que tu utilises ce mode de recherche par mot-clé ? A : Oui, mais dans une phase initiale. C'est-à-dire, vraiment quand j'aborde un point que je ne connais pas, je vais utiliser des mots-clés sans trop de difficultés, parce que je vais tomber assez rapidement sur des ouvrages généraux, mais une fois que j'ai des ouvrages généraux, je vais laisser le moteur de recherche de côté pour suivre ce que les ouvrages peuvent m'apporter comme références bibliographiques." (2.4)*
>
> *"Et puis effectivement – la méthode est difficile à expliquer – on cherche un peu dans les catalogues en ligne, les grands portails d'archives, APEx, etc. On regarde un peu, on tape des mots-clés, on regarde un peu ce qui existe. Et puis c'est en lisant les documents qu'on a aussi, si ça fait allusion, je ne sais pas, à telle commission, alors on va essayer de rechercher aussi s'il y a des archives de cette commission. Je compare toujours un peu l'histoire à un jeu de piste, puisqu'on lit quelque chose et puis ça nous amène à poser des questions, ça nous ouvre des portes, c'est un peu tentaculaire comme méthode." (2.5)*
>
> *"Quand c'est vraiment le tout début, je commence comme on le ferait tous par une sorte de collecte bibliographique, à partir de mots-clés, à partir d'auteurs qui reviennent déjà ou en tout cas dont j'ai entendu parler, et avec ce corpus bibliographique, les lectures commencent. Je regarde aussi à partir de quelle référence, et notamment sources, archives de première main, ces auteurs ont travaillé si c'est le cas. Ce qui fait qu'à côté et parallèlement à la recherche bibliographique, donc littérature secondaire, il y a une recherche plutôt de type archives, qui là se fait dans un premier temps aussi de manière quasiment aveugle et spontanée." (2.7)*
>
> *"La littérature scientifique, je vais sur Unicat, et puis de proche en proche quand j'ai trouvé un bon livre, je regarde qui il cite, etc. d'abord par mot-clé, et puis quand je vais trouver un auteur, je vais me focaliser sur cet auteur-là, je vais chercher ce qu'il a publié sur Cairn, etc. " (2.9)*
>
> *"Ce que j'utilise beaucoup pour ce qui est de la littérature, d'abord, ce sont les outils genre Cairn ou Persée. D'office, je sais que ça sera mon premier réflexe quand je cherche une référence sur un sujet, je pars par-là, essentiellement. Beaucoup de recherches mot-clé." (2.10)*

However, keyword search can raise some difficulties. When researchers use search engines or catalogues for discovery purposes, they are face-to-face with a blank page and need to use their imagination to translate their research topic into relevant queries.

"Ce qui me fait très peur dans les outils numériques et de recherche, c'est une barre à la Google, où on n'a pas beaucoup d'information, on n'a pas de recherche précise et où il faut se battre avec des mots-clés." (2.4)

"Pour ce qui est recherche de manuscrits, j'ai consulté les fichiers papier de la salle des manuscrits et qui procèdent aussi par mots-clés : il y a des tiroirs thématiques et il faut avoir beaucoup d'imagination et bien connaître son sujet pour entrer vraiment dans la matière." (2.4)

The exercise is all the more difficult that Google-like search is not very well suited to how library and archival systems work. Interviewed researchers are well aware of this fact, but this can be an issue with less experienced researchers or students, since Google has become the most intuitive way of searching for information on the internet.[27] In this respect, it is essential to teach students how to search for historical documents and secondary literature.

"If they [students] study, let us say, criminality, then most of them start searching 'criminality' and they can't find anything. That is why we teach a lot of these courses. If you want to study criminality, that means that you need sources created by institutions, judges and so on, so therefore trying to find sources by looking at institutions, what they have created, and not to search by topic." (2.13)

"I think that, especially when you look at students who are starting now, they are so accustomed to the Google-type search that they don't have the notion that you really should look further than what is suggested in your first… They don't have that anymore, I think they really treat everything as a type of Google, even when it doesn't really work that way. For instance, if I look at [topic], then you get the list of books, and then you can filter by article, newspaper, and so forth. And here it says 'sort by relevance', but it doesn't work like Google, that this is the most… you really have to go through everything to see it all, but most students don't realize this and they think that when they look at the first three, that they found the most relevant…" (2.14)

Among other issues associated with keyword search, researchers mention the "noise", i.e. irrelevant data cluttering the search results and requiring a lot of time to be sorted out. This is the reason why researchers still value the advanced search.

"Quand j'en ai l'occasion, quand j'utilise des moteurs de recherche des centres d'archives, j'essaie systématiquement d'aller chercher les recherches avancées pour calibrer ma recherche au maximum, pour ne pas perdre de temps dans le dépouillement des résultats." (2.4)

"Ce que j'aurais en tête évidemment, c'est de voir quand on fait des recherches un peu générales et de voir tous les résultats par centaines arriver, débouler et où on se dirait que là il faudrait plutôt affiner, affiner, affiner, mais bon, ça fait partie du jeu dans un catalogue de recherches." (2.7)

[27] Kemman, Kleppe and Scagliola 2014.

Ultimately, researchers distrust search engines because results of queries depend on many variables over which they have no control, such as the quality of the digitisation (OCR), the quantity and quality of metadata and the relevance ranking.

> *"Il y a des phrases qui sont illisibles, parce que la qualité de la numérisation est épouvantable. (…) Alors quand on ne veut pas se déplacer pour soulever un volume, on reprend le texte et on passe son temps à essayer de le déchiffrer, mais il y a des nuances qui manquent, il y a un moment dans un débat, on ne sait pas si la phrase est positive ou négative, ou il y a deux mots qui sont ambigus et qui peuvent signifier une chose ou son contraire." (2.2)*

> *"Il y a une quantité de métadonnées pour chaque document qui est assez incroyable, pour chaque document qui a été numérisé, (…) dans la note descriptive il y avait tous les lieux qui étaient cités dans le document, tous les noms qui étaient cités dans le document, plus ou moins la thématique et évidemment la date ou en tout cas la période, si on ne savait pas, à laquelle ça avait rédigé. Donc là ça avait été assez impressionnant en un sens, parce que on pouvait vraiment faire une recherche très pointue, très précise." (2.3)*

> *"J'ai l'impression que c'est un miroir aux alouettes, qu'on donne le sentiment au chercheur, à l'utilisateur d'accéder à une masse d'information prodigieuse alors que peut-être il n'en voit que la surface. Et j'ai le même problème quand j'emploie Google pour des recherches qui sont complètement sur le côté, je m'aperçois que Google trie une série de choses et que ce qu'il me propose en tête de liste ne correspond pas nécessairement à mes besoins, et que c'est peut-être à la page 7 ou 8 que je vais trouver l'information précise qui m'intéresse. Et il y a d'autres moteurs de recherche qui sont concurrents et qui m'offrent des résultats qui sont complètement différents et je m'aperçois quand je les utilise que Google me présente ce qu'il veut bien me présenter." (2.4)*

> *"Ce qui manque, ce sont des bons mots-clés. (…) Si je sais ce que je cherche, je peux le trouver, mais je ne peux pas trouver ce que je ne connais pas encore." (2.14)*

The quest for comprehensiveness

> *"Quand la recherche se spécifie, quand elle va rechercher du détail, il y a une autre logique qui se met en place, qui est la logique panoptique, où il faut tout voir pour essayer de trouver ce qui nous intéresse. Et ces moteurs de recherche constituent un handicap, là où ils sont un avantage au début de la recherche pour préparer le terrain, quand on passe dans une recherche plus avancée, ça devient un handicap, il faudrait un autre outil." (2.4)*

In their information search, researchers aim to gain an overview of relevant sources that is as comprehensive as possible. While keyword search is a fantastic tool for discovery purposes, it is insufficient in its own to achieve this goal. Keyword search can make them feel as if they don't have full control over the material, since results highly depend on the quality of the search engine and on the relevance of the queries that are put in it. In addition, search engines do not provide clearly defined boundaries to the material at hand and researchers can never be sure that they have covered everything. Institutions do not always provide a clear information on the content of their digital catalogues, on which parts of the collections have been catalogued. To sum it up, the main questions

are: Will the search engine generate all the relevant documents that are preserved in the institution? And will researchers think of all possible queries associated to a certain topic?

> *"Ce qui est vraiment problématique avec internet, c'est que c'est tellement fluide, et si on fait des recherches par exemple exhaustives, il faut un livre, un CD-Rom, une base de données, fermée, où on voit les limites." (2.6)*

> *"Je ne sais pas si c'est toujours pertinent d'aller dans le fichier dans la salle de lecture." (2.3)*

> *"Quand je ne trouve pas quelque chose, je vais envoyer un mail pour vérifier que je ne me suis pas trompée et qu'il n'y a rien. (…) En tout cas, l'absence de données ne nous donne pas de certitudes sur le fait que les archives n'ont pas été conservées. Parce que je n'oserais jamais dire que les archives [...] ne sont pas disponibles pour telle date à telle date si je n'ai pas demandé à un archiviste de me le confirmer." (2.9)*

> *"[On] m'a dit qu'il y a quand même beaucoup de fiches qui n'ont pas été mises dans le système, donc ça c'est quelque chose que je ne savais pas auparavant et qui oblige quand même à aller faire des recherches dans les anciennes fiches." (2.11)*

> *"Je suis mal à l'aise avec le système des mots-clés, parce que je sais que choisir, c'est laisser de côté une série de choses, et que potentiellement dans ce que j'ai laissé de côté, dans les mots-clés que je n'ai pas pensé à utiliser, il y a probablement des choses qui sont vraiment intéressantes." (2.4)*

In this respect, one particular issue comes from variant spellings of ancient words and sources in different languages.

> *"J'ai de gros problèmes d'orthographe avec les noms anciens. Donc pour faire ma recherche de façon efficace, il faudrait que je puisse répertorier toutes les formes anciennes d'un nom pour pouvoir retomber sur tous les documents. Or, avant d'avoir vu les documents, je ne peux pas en établir la liste des différentes formes des noms, et les moteurs de recherche me posent problème à ce niveau-là, parce que je ne suis jamais sûr d'avoir parcouru tout ce qu'il y avait à parcourir." (2.4)*

> *"Je combine les choses, et surtout je combine les langues, je trouve ça aussi important. Parce que si on ne cherche que dans une langue, on rate aussi beaucoup de choses." (2.14)*

To compensate the haphazard character of the keyword search, researchers often combine it with a systematic browsing of the collections. For instance, they may skim the inventories and other finding aids that they consider relevant for their research. Therefore, the fact that more and more of these finding aids are now published online has been praised as particularly useful.

> *"Lorsque je suis en centre d'archives et que j'ai consulté les sources que je voulais voir, ma deuxième façon d'aborder la matière, c'est de prendre un catalogue très large et de le parcourir entièrement. Il y a énormément de choses que je laisse de côté, et puis parfois il*

y a des fonds ou des portefeuilles qui me semblent plus intéressants que d'autres, et là je vais les découvrir." (2.4)

"J'ai trouvé beaucoup de choses qui avaient échappé aux autres systèmes, en utilisant les fichiers, mais d'une manière systématique : des heures et des heures de filtrage." (2.6)

Such a method also presents the advantage of leaving some room for serendipity, which is an information seeking technique defined as the unexpected discovery of information. For instance, when researchers find the reference of an article, they will browse through the journal's table of contents to see if they can find other interesting contents.

"Je suis déjà allé sur des sites de revues et je croyais juste aller chercher un article et au final j'en trouve cinq." (2.3)

This method can be translated online by representing the multi-level hierarchical structure. Such a resource has been cited multiple times as an example of good practice, as it allows researchers to cover all the preserved collections and to visualise one item in its context.

"Et donc ça [l'arborescence] c'est utile, parce que j'ai une vue très globale de l'ensemble de la matière." (2.4)

"Il faudrait une arborescence où on peut développer. (...) on voit en un seul coup tout ce qui est accessible sur le sujet sans devoir taper comme mot-clé '[...]' dans le moteur de recherche, et ça va sortir de nouveau 500 résultats et on ne va pas savoir trop où chercher, donc je pense que ça gagnerait beaucoup en visibilité." (2.5)

Information professionals as a resource

Most researchers regularly ask the guidance of archivists and curators, whose expertise they highly value. It allows them to open new research avenues, resolve issues, discover collections which had not (yet) been inventoried, or simply compensate the shortcomings of online resources. However, they usually refer to information professionals only when looking for primary sources, and not secondary literature, which they feel that they can retrieve more easily through online resources.

"En discutant avec le préposé dans la salle des manuscrits, il m'a donné beaucoup plus d'information que ce que les cartons ont pu m'apporter. Il m'a dit qu'il y avait des fonds qui concernent les collègues Jésuites qui ne sont pas connus, mais qui existent. Et c'est parce que lui les connaissait que j'ai pu en avoir connaissance." (2.4)

"Je fais ça dans les bibliothèques, et je sais que ça peut complètement débloquer un dossier. J'estime beaucoup les professionnels dans les institutions." (2.6)

"On a pu mettre la main sur beaucoup d'informations grâce à des pistes qui avaient été, au départ, esquissées notamment par les archivistes." (2.7)

"Il est évident qu'il faut reconnaître l'expertise des gardiens des collections, et je devrais peut-être l'utiliser un peu plus." (2.8)

> *"Je sais que, quand je me dis quels ont été les déclics, c'est [...] qui m'a donné les bons filons et quand je cherche quelque chose [...], j'envoie un petit mail à [...], qui est toujours très éclairant, et très rapide et très efficace. (...) Je suis très admirative de leur promptitude à rendre service." (2.9)*

> *"I think that library staff and information professionals have a very big role to play in the research process as well. (...) they could really help me through their expertise, because they know everything that comes in, they know more or less what it is in, they are really familiar with the material." (2.15)*

Researchers may contact professionals at various stages of the research process. At the initial phase of collecting sources, it will help them to get a first orientation in the collections and to prepare a visit to the reading room. At a more advanced stage, it will help them complete previously collected information. Or they will contact them as a last resort, when they feel they are in a dead-end. To get in touch, they use most often the email, but may also meet them directly in the reading room, during informal encounters or through collaborative research projects. Participants emphasise the personal relationship that develops and needs to be fostered to bear its fruits.

> *"C'est vrai qu'aux étudiants, je leur dis de ne jamais négliger l'aspect humain et qu'il y a une relation de confiance à créer, c'est un investissement à faire au début." (2.10)*

Conversely, researchers sometimes encounter difficulties when seeking professional guidance. The reasons behind those difficulties are worth reviewing:

- They find difficult to identify and reach the right contact person;
- They do not know the curators or the archivists personally;
- Their requests remain unanswered;
- They are afraid of bothering them or distracting them from their work;
- They think that their research interests are too narrow;
- There is no expert on their period or research topics among the staff.

Monitoring

Researchers use monitoring as an information-seeking technique to keep up to date with what is going on in their field. In this case, new technologies provided by the internet are particularly useful. While they still browse through newly published journals and read book reviews to discover new publications, they are increasingly doing it online. Many of the participants have subscribed to newsletters from institutions or publishers and to mailing lists in their field. They also frequently use social networks, particularly scholarly networks such as academia.edu or Research Gate, but they also use Twitter and even Facebook as professional tools. However, this monitoring is not often organised in a systematic way, for instance through RSS feeds, as researchers emphasise how difficult it is to stay on top of the abundant flow of publications and to find the time to read everything.

2.2. Access to information

Access to information is one of the main issues addressed during the discussion with researchers, and turns out to be a key priority for researchers in terms of improving digital services.

> *"Avoir un accès aux collections qui soit le plus optimal possible, pouvoir brasser des volumes relativement importants sans que ça prenne trop de temps." (2.1)*
>
> *"Il n'y a pas de bon ou de mauvais système, le bon système est celui où on trouve l'information qu'on cherche, et un mauvais système c'est celui où on râle en se demandant pourquoi ils n'ont pas." (2.2)*
>
> *"L'accessibilité facile à des sources institutionnelles importantes est essentielle." (2.8)*

It is worth examining in more details the experience, both good and bad, shared by researchers when dealing with access in heritage institutions in general. Their feedback is summarised in the table below and organised according to various subjects: access to the reading room, service to the public, ordering items, copying documents, findings aids, and online access.

Subject	Positive feedback	Negative feedback
Reading room	○ Online registration form ○ Online detailed information about terms of access and rules	○ Complicated and time-consuming registration procedure ○ Limited opening hours ○ Low availability of seats ○ Access linked to privilege or credentials ○ Researchers unwelcome ○ Geographic distance
Service to the public	○ Helpful staff	○ Non-qualified staff ○ Difficult to identify the right contact person ○ No reply to requests
Ordering documents	○ Option to request items online and in advance ○ Online help for ordering procedure ○ Direct access to microfilms without registration	○ Long waiting time between ordering and receiving documents ○ Complicated ordering procedure ○ Limitation in the number of document requests per day ○ Lost/destroyed documents
Copying documents	○ Authorization to take pictures ○ Digitisation on demand	○ Reproduction restrictions ○ Complex reproduction rules ○ Reproduction costs
Finding aids	○ Finding aids available online ○ Online help for the use of finding aids	○ Non-digitised finding aids ○ Restrictions in access to finding aids ○ Non-inventoried collections ○ Outdated finding aids

Online access	○ High-quality digitisation ○ Overview of collections ○ Online access to finding aids and detailed descriptions ○ Unified access to digitised documents ○ Online (full or partial) access to digitised documents	○ Poor interface design and lack of usability of online resources ○ Lack of information on digital content ○ Cost of proprietary documentary databases ○ Siloisation of online resources ○ Access to digitised documents restricted to the reading room ○ Lack of digitised documents

Those comments give a good insight into the features recommended by researchers as good practices and into the issues that need to be improved. In particular, they often wish to have more information on the digital content, highly value the possibility to access finding aids online, and all kinds of practical information that can help them prepare their visit to a reading room.

Digitisation of heritage collections

Digitisation policies of heritage institutions are a major challenge for enhancing access to their collections and the topic has been often addressed during the interviews. Online access to digitised sources is obviously of great interest to researchers in terms of convenience and time saving, but also as an opportunity to apply new methods of investigation to historical documents. They can easily search OCRed PDF files by keywords, not to mention performing advanced digital methods such as text mining and distant reading. For the study of ancient documents, a high quality digitisation and adjustment of display settings allow them to see details that would be otherwise invisible to the naked eye.

However, researchers also point out some issues associated with those digitisation policies. First, digitisation does not necessarily equate enhanced access to the sources. Copyright restrictions prevent the institutions to make all their digitised collections freely accessible online. In the case of the newspapers collection digitised by the Royal Library in Belgium, only newspapers which are in the public domain, i.e. published before 1918, can be perused online. The rest of the collection can only be accessed on site. Furthermore, digitisation demands important financial and human resources, and some heritage institutions chose to outsource the process to private firms, who charge high fees to access their databases, fees that only a small number of universities can afford to pay. The lack of coordination between institutions in the digitisation process may also result in a duplication of efforts and a siloisation of web-based platforms. In this respect, aggregated catalogues which provide a single point of contact to several digital collections are particularly useful for researchers, as it saves them from introducing the same queries in all individual search engines.

From a methodological point of view, digitisation change fundamentally the relation between researchers and historical documents. It also inevitably result in the loss of

information. For instance, the context of an archive is not always easy to replicate on screen where documents are viewed in isolation. Viewing the document in its context is important for researchers as it allows them to better understand its relation with the rest of the fonds, in addition to giving them the opportunity to find other documents they did not necessarily expect.

> *"C'est pour ça que je suis encore très papier, c'est que dans les archives du [...], ce sont souvent des gros dossiers, et si on a de la correspondance qui concerne une seule affaire, les lettres sont toutes enchâssées les unes dans les autres et on peut voir facilement que tel paquet enchâssé concerne une affaire, et généralement les lettres sont classées par ordre chronologique. Et je trouve que quand c'est numérisé, on perd cette cohérence physique des documents, on voit les documents les uns à la suite des autres effectivement, mais on ne voit plus qu'ils sont liés ensemble en fait. Je ne vois pas comment on pourrait faire, on est obligé de numériser papier par papier…" (2.5)*

> *"Je trouve qu'en termes de réflexion scientifique, d'avoir quelque chose en ligne, un corpus tout beau, tout propre en ligne, ou bien avoir le carton ou la liasse d'archives et réfléchir à la logique de sa création… Et au fond on se rend compte qu'on a demandé le dossier 4, mais qu'il y a six dossiers dans la boîte et que le dossier 6 est tout aussi intéressant, mais on n'y avait pas pensé pour telle ou telle raison, c'est intéressant." (2.10)*

Digitisation cannot fully account for all aspects related to the materiality of historical documents, such as the quality of the paper, the ink, watermarks, dimensions, colours, etc. Researchers focussing on this field therefore still need to study the originals.[28] Interestingly, the relation of researchers with ancient sources is not the same for all types of documents. In the case of newspapers, content takes systematically precedence over form, and the advantages of digitisation in terms of usability and searchability far exceeds the need to see the originals.

> *"Je ne pense pas que pour l'instant c'est important de tout numériser, parce que ça déforme l'aspect unique. Par exemple, on n'a pas vraiment la perception des dimensions quand on voit ça sur écran, ou du toucher. Pour des manuscrits du Moyen-Âge, oui, parce que si tout le monde les touche, c'est fichu, mais ça crée une distance, on n'est plus en contact [avec l'objet]. Aussi la façon de manipuler un livre, c'est différent si on le fait manuellement ou si on fait ça sur l'écran, déjà les mouvements sont différents." (2.14)*

Another methodological implication of digitisation is that it may introduce a distortion in the representation of collections. Digitised collections will inevitably benefit from more exposure and, consequently, attract more research projects. On the other hand, non-digitised collections might be easily overlooked. The digitisation policies of heritage institutions thus have an important role to play in the matter, since they are not always research-driven, but are based on other criteria such as frequency of consultation or trends.

[28] On the link between digitisation and materiality of historical documents, see Roustan 2016.

"Il y a aussi un risque que les projets ne soient écrits que pour les collections qui seront le plus facile d'accès, et où on va plus ou moins déconseiller aux chercheurs d'investir leur temps dans le travail heuristique là-bas, comme on a déjà ça, ça, ça et ça. (…) Et évidemment, si on concentre l'effort sur un corpus donné, ça va créer de la masse dans les publications, dans la visibilité dans les conférences et ce genre de recherches est évidemment privilégiée par les avantages qualitatifs et les avantages d'échelle qu'on a maintenant avec la digitalisation. Donc c'est inévitable, je crois que ça va distraire, tordre un peu la représentativité de ce qu'on a dans la littérature. On voit déjà que certains jeunes historiens ne consultent pas ce qui a été écrit il y a 20 ou 30 ans, qui était généralement basé sur les grandes archives institutionnelles, il y a un danger que ça continue dans cette voie-là." (2.8)

"Donc de nouveau, que faut-il mettre dans les bases, que faut-il numériser, est-ce qu'on ne tronque pas, avec les priorités, des parties de l'histoire, qui doit décider des priorités de numérisation, des priorités d'encodage, des priorités… C'est toutes les questions qu'on pose." (2.10)

However, historians have the particularity of enjoying the quest for rare, unpublished or overlooked pieces of evidence that could bring a significant contribution to their field. In this respect, some of them could never be content with the sources that are easily available online.[29]

"Et c'est ça qui me fait très peur avec les archives, j'ai très peur qu'on en arrive à une consultation de documents qui ont déjà été consultés, et qu'il y ait une idée folle d'inscrire un facteur de popularité des sources, parce que là on va labourer toujours le même champ, alors qu'il y a peut-être des choses à côté qui sont tout aussi riches, inconnues et qui permettraient de faire avancer la recherche. En fait, ce qui me serait vraiment utile en matière de recherche, ce serait de savoir ce qui a été inventorié, quels sont les terrains qui sont déjà défrichés, et ce qui ne l'a pas été. Et de me dire, est-ce que je vais aller voir dans ce qui a déjà été défriché et donc employer mes inventaires pour gagner du temps, ou bien est-ce qu'il y a du potentiel dans ce qui n'a pas été exploré – je sais que par exemple il y aurait un fonds du Conseil d'État qui couvrirait 25 ans, qui couvrirait une surface de deux mètres linéaires, qui n'aurait pas été inventorié, je ne suis pas sûr que je n'irais pas faire un coup de sonde dans ce qui n'a pas été inventorié, parce que je sais que ça existe, pour essayer de trouver quelque chose qui n'a pas encore été trouvé par d'autres. C'est cette part d'ombre qui m'intéresse, et j'ai l'impression qu'avec un moteur de recherche qui serait pensé pour faciliter l'accès, on perdrait l'accès à cette part d'ombre-là." (2.4)

Aside from these methodological issues, participants overall support the digitisation policies of the State Archives, the CegeSoma, and the Royal Library. They are aware that, given their financial situation, these institutions will never be able to publish all their collections online. They also understand that the sources they study in their research are not necessarily those who interest the major part of the public. The digitisation of the most frequently used collections, such as genealogical sources and newspapers, as well

[29] Lemercier 2014.

as of the most useful tools, such as finding aids, should remain the priority of the institutions in their opinion.

> *"I think it is an impossible task to fulfil all the dreams of the researchers. As I told you, my research is quite topical, and I have my own needs, but I guess there are other people who would have another view on this problem." (2.12)*

Suggestions have been made to improve online access to the collections, based on good practices observed elsewhere, which is partial digitisation. One participant speaks of another institution which offers previews of collections on its website, by digitising small samples of the collections. These previews give users a better idea of the content of a specific fonds and acts as an incentive for them to make a trip to the reading room to access the rest. Another idea to enhance digital access to the collections, particularly for international researchers, is to implement digitisation-on-demand according to a system where the first person to request the scan pays the fees and then the document is made available online for all.

> *"Ils ont entamé des programmes de numérisation partielles de documents, et bien en fait ça nous a donné vraiment envie de dire qu'il fallait poursuivre l'enquête, poursuivre la recherche en archives sur place. Donc je crois que l'idée, ce n'est pas de faire de l'exhaustif, c'est juste dire qu'on va faire quelques numérisations parce qu'on ne peut pas tout faire, et là le chercheur va se rendre compte si oui ou non c'est quelque chose qui lui parle et alors dans ce cas-là il faut soit envoyer une demande à un archiviste, ce qui se fait dans pas mal de cas, soit même déjà à partir d'un inventaire qui est en ligne, faire une commande, une réservation de documents sélectionnés. (...) et là aussi ils ont sélectionné plusieurs pièces qu'ils ont numérisées, notamment les pièces les plus consultées d'habitude, des choses comme ça. (...) ça m'a donné, pas l'envie, mais plutôt l'opportunité de me dire qu'il fallait poursuivre la recherche sur place, parce que ce qui était disponible en version numérique n'était qu'une partie et que la suite, je devais y aller." (2.7)*

> *"La numérisation sur demande me paraît dans un monde de recherche très international où tout le monde peut travailler sur n'importe quel sujet très important. (...) Le premier utilisateur paie le coût de la numérisation, et après la numérisation reste en ligne pour tout le monde. Ce qui me semble une pratique meilleure, ce qui sous-entend évidemment que les tarifs ne soient pas prohibitifs et c'est un service assez généreux et assez efficace parce qu'on voit les parties de la collection qui sont consultées, ça vient progressivement sur le net et ça bénéficie à tous. L'idée d'avoir un accès ouvert aux archives est incorporé dedans également et aussi l'idée que les prestations seront rémunérées et qu'on n'ait pas à faire un projet colossal pour tout numériser d'abord, et pour en récolter les fruits après." (2.8)*

2.3. Gathering information

Once researchers have identified information that is relevant to their research, they collect it by downloading it, photographing it or asking for a copy. Alternatively, they may take notes summarizing the content, or transcribe interesting passages.

> *"Rassembler l'information, information que je photographie si c'est un support papier, là où c'est, en archives quand c'est permis, ou que je rassemble sous forme de fichiers, donc généralement c'est soit en imprimant, puisqu'on peut imprimer certains articles, c'est payant généralement, ou alors je prends note, ou je prends mon ordinateur et j'écris ma notice directement en fonction des informations dont je dispose." (2.2)*

These strategies are not equal. If the document has been digitised and is available online, the preferred option will be to download it. This is, for instance, systematically the case for scientific papers collected from web-based publishing platforms. Even if the document is available online in open access, researchers will usually keep a copy in their own digital ecosystem, on their computer's hard drive or in their cloud. In this way, they centralise all their documents in one single place and are able to easily retrieve one specific file, regardless of where it came from. In some cases, storing the document on their hard drive, especially if it is voluminous, allows them to browse through it quickly, independently of the speed of an external server. Finally, researchers may be concerned regarding future access to the document, which highly depends on the sustainability of the website, but also on their institutional affiliation, if the access is restricted.

> *"Je pars d'un principe, c'est que pour le moment, beaucoup de choses sont gratuites, et que ce n'est pas sûr que demain, ça le sera toujours. Deuxième principe, c'est que je trouve une information intéressante, elle est accessible aujourd'hui, il n'est pas sûr que demain le site sera toujours là." (2.2)*

If the document is only accessible in the reading room, researchers will preferably take a picture, if the institution permits it. The democratization of digital photography undoubtedly brought one of the major changes in the historian's research workflow. Formerly, visiting an archives reading room meant reading the material, assessing its relevance, synthetizing the content and transcribing interesting excerpts. Now it is very often aimed at collecting photographic reproductions of documents, thus postponing the analysis of the material to a later stage.[30] One participant speaks of a "holdup in the archives" (2.8), while another admits taking picture "almost in a frenzied way" (2.7). Researchers highly value the time saving and comfort that this technique of collecting sources provides. It spares them from having to constantly go back to the reading room to consult the same documents. It is somewhat reassuring to keep a digital copy of the sources at hand if they need to check them again later.

> *"Je me souviens que pendant mes études, je prenais des notes au tout début, et il y a toujours un élément qu'on oublie de noter, alors après on doit retourner dans le document et c'est un peu une perte de temps." (2.5)*

> *"Il y a un côté rassurant aussi de se dire qu'on peut télécharger tout le document et ensuite ramener ça chez soi et faire la recherche à son aise, donc il y a un côté qui permet de transférer ce temps de traitement." (2.7)*

[30] Artières 2016: 19-21.

This practice requires excellent organizational skills in order to be able to easily retrieve and identify photos afterward. Researchers usually take notes to record the references or a summary when they photograph a set of documents, to help them sort out the files later. Mostly they open a blank document in a word processing software, or occasionally use a more elaborate system such as a virtual notebook, a spreadsheet or a database. It is also essential for them to organize their files as soon as possible after the visit, while their memory is still fresh.

> *"Pour les sources primaires, je préfère travailler sur place, parce que j'ai toujours l'impression que si je ne prends que des photos et si je ne lis pas la source, et je ne prends pas de note, je vais oublier, je ne vais pas le faire au bureau, j'ai autre chose à faire au bureau, et je pense que c'est important, au moment où je consulte la source, le dossier, que je le lise et que je prenne des notes." (2.11)*

> *"Et faire attention dans cette stratégie de tout de suite reclasser les photos. Sinon c'est l'enfer après." (2.10)*

> *"Je sais que c'est impératif que je mette très vite de l'ordre alors, après la journée de dépouillement, que je réorganise, que je vérifie si tout est dans le bon ordre." (2.9)*

> *"Je suis passé au numérique pour les archives, très clairement, et en prenant bien soin aussi – ça c'est un danger, notamment la première fois que j'avais systématiquement fait tout ça, il y avait un danger de la référence. Je m'étais rendu compte que je n'avais pas référencé systématiquement toutes les archives que je photographiais, donc c'était au final une perte de temps, un peu comme quand on est étudiant, on prend plein d'infos et on se demande c'était quelle page, on a oublié, et donc il faut retourner au document et vérifier la page." (2.7)*

However, reproducing archival documents through photography may create a false sense of expediency. With digital photography, researchers have virtually no limitations as to the number of pictures they can take. While it allows them to quickly collect a vast amount of documents, it does not prevent them from having to process the data afterwards. On the contrary, sorting out so many photos can require a lot of time.

> *"Le traitement peut être fastidieux, il l'est, puisqu'on a photographié beaucoup trop, par rapport à ce dont on avait besoin" (2.7)*

> *"La numérisation parfois je trouve ça une fausse sensation de facilité. C'est bien d'avoir accès, mais il faut avoir le temps de l'exploiter, et les horaires sont ce qu'ils sont, les agendas sont ce qu'ils sont. Au moins, l'avantage, c'est que parfois quand on va aux archives même, on est obligé de travailler. Ce n'est pas tant le processus de petite souris, d'accumuler, au moins on est dedans et on lit des documents." (2.10)*

Therefore, some researchers still enjoy working in the archives reading room "in the old way", that is reading and processing documents on site (see also 2.5). Often, strategies vary according to their circumstances or to the type of documents. When on a research trip abroad, researchers are subject to a bigger time pressure, and have to collect as much material as possible in a limited period.

"Maintenant je fais des photos, ce que j'ai mis du temps à faire, parce que j'aimais bien d'abord être dans les archives et les traiter déjà." (2.9)

"Parce que d'ailleurs dans ma stratégie, je ne numérise pas systématiquement quand je suis aux archives. Parce que parfois je trouve que c'est plus long en fait de prendre des photos, parce qu'il faut les retravailler après et parfois j'ai l'impression de faire deux fois le travail et qu'au final ça va tout aussi vite de retranscrire ou de synthétiser. Après voilà, je note, s'il y a un document qui fait 40 pages, c'est vrai je ne vais pas retranscrire les 40 pages, alors là je numérise [de manière] très limitée, très ciblée. Voilà, après je suis très flexible dans mes stratégies : à l'étranger je prendrai beaucoup plus de photos s'il y a moyen. Parce que ça dépend. Pendant ma thèse, quand je travaillais sur les Pays-Bas, c'est vrai que là j'allais pour des plus courtes périodes, donc quand il y avait moyen, je photographiais. En France, j'avais des plus longues périodes, donc je ne me refusais pas non plus de travailler à l'ancienne, ordinateur portable." (2.10)

"I take a lot of pictures of course, to do this at home, but mostly I try to read them while using the document itself. I still prefer to go to the archives and be there two or three days. Of course, if it is too much, I take pictures, or when I go [abroad] for instance, then I take pictures. But I prefer still to go to the reading room. But again, I think I am one of the last generations doing that." (2.13)

"Étonnamment, en fonction de la nature du document, j'ai une pratique différente. C'est maintenant en parlant avec vous que je m'en rends compte. Littérature secondaire de type travaux d'historiens, politistes ou autres, j'essaie de les télécharger pour les avoir en mémoire dans ma liste bibliographique, dans mon ordinateur, les documents sources (…), là je vais avec mon ordinateur ou alors de quoi noter – parce que ça m'arrive aussi – et je prends note et j'essaie de déjà faire le traitement sur place, ce qui n'est pas le cas dans la troisième partie qui sont les archives où là j'essaie de faire le plus de photographies possible et je transfère le temps de traitement au bureau. C'est un saucissonnage, finalement, en fonction de la nature des documents." (2.7)

"On a qu'une semaine pour dépouiller une centaine de liasses d'archives, on ne va pas passer son temps à lire la page, on se dit que là il y a quelque chose d'intéressant, il faut la reproduction." (2.2)

2.4. Managing research data

Data management refers here to how researchers organise and process the information collected throughout the research workflow, including references, notes, transcriptions, as well as digitized documents. In other words, this section explores which systems researchers use to organise their documents in their digital ecosystem and what is the role of computational tools in the way they use and share their research data. Two types of data are considered: scholarship, that is bibliographical references, electronic articles, e-books, etc., and primary sources.

Scholarship

There are two categories of researchers: those who use a reference management software to organise their bibliographical references and those who do things, in their own words, "the old way".[31]

A reference management software is designed to record and organise bibliographic references and electronic versions of documents. It also allows users to import references from online catalogues or publishing platforms, annotate their documents, insert citations into word processing documents and automatically format them in various bibliographic styles. The open-source software Zotero is the most often cited as reference, but some use Endnote or Mendeley, depending on which software is promoted by their institution.

Scholars who do not use this system typically keep a list of references in a word processing document which they regularly update. To insert references in their work, they copy-paste them from this document. Additionally, they organise the PDF files of books, papers, and other documents in folders on their computer's hard drive, naming them by "author" and "date".

Almost all researchers praise the many advantages of the reference management software in terms of time saving and efficiency, even those who do not use one.

> *"J'essaie d'emmagasiner un maximum, et là Zotero m'est très utile dans ce travail qui s'apparenterait à de la veille scientifique." (2.1)*
>
> *"Ça facilite quand même le changement de références quand on envoie un article et que l'éditeur demande un autre style que celui qu'on a employé, ça facilite grandement le travail. (…) L'avantage, c'est qu'on peut vraiment créer son classement personnel avec des fichiers et des sous-dossiers, et donc ça fonctionne pas mal." (2.4)*
>
> *"À chaque fois, avec le centre de recherche, on en discute et on sent bien que les jeunes sont déjà aguerris à la pratique Zotero ou autres, qui permettent en effet avec un ou deux clics d'avoir une note de bas de page déjà bien préparée, d'organiser finalement l'appareil critique." (2.7)*
>
> *"J'ai des collègues qui utilisaient Zotero, ils joignaient à chaque fois des fiches de lecture, enfin ils utilisaient vraiment toutes les potentialités, c'était assez stimulant." (2.10)*
>
> *"J'utilise aussi beaucoup les tags, je trouve que c'est très pratique, on a des mots-clés et c'est facile de retrouver les textes qui vont avec ce mot-clé. On peut aussi surligner." (2.11)*
>
> *"I discovered Zotero and now basically I use it as much as possible because it is so useful, especially for storing PDFs and scans can also be inserted, I mean you can do everything with it." (2.15)*

[31] Among 15 participants, 5 frequently use a reference management software, 3 use one occasionally, 6 heard of such a software but do not use one, and 1 never heard of it.

In this respect, there is a paradox in the discourse of some researchers who are very much aware of being on the other side of a scholarly digital divide. Although they feel at odds with what they acknowledge as a good practice, they are still reluctant to change their method, sometimes without any solid argument to make:

> *"Rationnellement, je trouve que ça serait intéressant, mais je ne sais pas, c'est peut-être la force d'inertie, ou je n'en ressens pas un besoin énorme." (2.9)*

Some issues with respect to digital literacy or lack of compatibility may come to the fore:

> *"Ça fonctionne avec des étudiants qui sont à l'aise avec l'outil informatique et toute notre population estudiantine n'est pas à l'aise avec l'outil informatique." (2.4)*

> *"Ça me semble beaucoup plus à ma portée que le système d'organisation Zotero." (2.7)*

> *"À l'époque, il n'y avait pas encore beaucoup de sites qui étaient compatibles, donc il fallait quand même encoder directement. Je n'en voyais pas trop l'intérêt." (2.5)*

More often, researchers face the challenge of having to convert a considerable amount of data collected over several years if they change their system. This would require a significant investment in time that they are not always ready to make, especially if they are on short-term contracts.

> *"J'avais accumulé une bibliographie sur un simple fichier Word, et je suis tombé sur Zotero trop tard, et je n'ai pas eu le courage d'encoder 60 pages de bibliographie en un coup." (2.3)*

> *"La difficulté étant que j'ai découvert ça dans la seconde moitié de ma thèse, et donc c'était de nouveau un moment charnière, est-ce que ça vaut la peine d'investir complètement, oui ou non. Je ne suis pas revenu à 100 % en arrière, et je n'ai pas continué à 100 % non plus, donc j'ai un Zotero un peu boiteux, un peu incomplet." (2.10)*

Other issues of concern are sustainability and fear of losing control over their data, particularly if it is linked to a proprietary software to which long-term access is not guaranteed.

> *"C'est plutôt encore à l'ancienne je dirais, tout simplement, accumulation d'une série d'articles, de documents, dont je note les références scrupuleusement - là c'est une perte de temps totale, je sais bien, mais je ne sais pas pourquoi, je me sens plus à l'aise avec cette façon d'avoir un contrôle direct sur la matière que j'organise plutôt que d'avoir ces listes de références qui assez rapidement jouent l'inflation. Mais ça c'est très psychologique. Donc c'est un déclic que je n'ai pas eu à un moment ou l'autre de mon parcours." (2.7)*

> *"Le problème, c'est la dépendance, donc il faut toujours avoir accès à l'université qui a une licence pour le logiciel en question. Et là, j'ai des collègues qui sont allés vers Zotero, alternative open source, et je procéderai probablement de même si je trouve un moment pour organiser ça. Parce qu'évidemment, si on a ça pendant des années, c'est un très gros fichier et on ne veut surtout pas le perdre, c'est la colonne vertébrale de la recherche." (2.8)*

Primary sources

Primary sources include electronic versions of sources, which were downloaded, photographed or scanned, as well as related metadata, such as notes, references, summaries, transcriptions, etc. For the former category, researchers typically organise the files in folders on their hard drive, according to the same hierarchical structure that is used within the library or archive where the source is preserved. The considerable amount of files that they need to handle – between 30,000 and 35,000 according to one participant, about 250,000 for another – requires a meticulous organisational system. Ideally, these are OCRed PDF files allowing annotations and keyword search. Some researchers also mention the use of an application to convert photos into OCRed PDF files.

> *"J'organise d'une manière très claire sur mon PC, par centre d'archives, par fonds, par numéro, et puis je mets la photo. Parce si on met les photos comme ça, c'est impossible de s'y retrouver, il faut vraiment garder une méthode de classement." (2.5)*
>
> *"Je fais une bibliothèque électronique sur un disque dur que j'organise par dossier, bibliothèque, et puis par cote." (2.6)*
>
> *"Reprendre la logique archivistique : dépôt d'archives, dedans il y a un dossier pour telle sous-série ou tel fonds." (2.10)*
>
> *"Je les nomme d'après leur numéro d'inventorisation pour les retrouver plus facilement." (2.11)*
>
> *"Pour les travailler, je collecte les photos dans un PDF et je travaille le PDF avec Good Reader, une app sur IPad, où je mets des commentaires, qui me permettent très vite, quand je transfère le fichier vers mon ordinateur de voir où j'ai fait des notes et donc d'avoir un arsenal de citation de lettres à portée de main." (2.8)*

As for notes and metadata, various systems are used. Some keep word processing documents containing notes on the documents that they have consulted. Others use a spreadsheet or a relational database to keep track of the sources useful to their research, which also allows them to conduct searches and compare datasets. In this respect, the way researchers organise their data highly depends on their research questions.

> *"Je travaille beaucoup avec des bases de données que j'ai créées moi-même, surtout quand je prévois que ça va durer un certain temps et que je serai obligé de consulter beaucoup de choses." (2.6)*
>
> *"J'utilise Endnote surtout comme une base de données, comme recueil de tout ce que j'ai pu voir ou tout ce qui m'a été signalé." (2.8)*
>
> *"Je me suis créé en Access un outil d'encodage, plus qu'une base de données, qui me permet d'encoder rapidement les références de la pièce et au choix, la retranscription ou un résumé, et de classifier la pièce selon des problématiques de recherche." (2.10)*
>
> *"I used [Excel] just to make a database of every title and author I had, just to make an index of my corpus." (2.15)*

"Une fois que j'ai fini ce travail sur plusieurs portefeuilles et que j'ai mes fichiers Word à disposition, je ré-encode le tout dans une base de données, que j'ai faite moi-même, qui me permet de faire des recherches par date et de rapprocher des fonds qui sont peut-être très différents autour d'un événement ou d'une date, et donc de faire des recherches croisées." (2.4)

Building a relational database is not necessarily self-evident for researchers who are not technically minded, if they did not complete a training course to learn. Alternatively, the added value of a database is not worth the time investment.

"Je crois aussi qu'avec FileMaker, il faut être très précis et avoir bien pensé sa base de données, parce qu'à la moindre erreur, il y a des résultats de recherche qui n'apparaissent pas." (2.3)

"Being a professor, I don't have the time to make big databases. I only have the time to write smaller publications, articles, therefore I don't need datasets which are enormous, only very specific ones. I can handle the information found in the archives just by putting that on a Word file, for instance." (2.13)

Aside from a few researchers who annotate PDF versions of their sources, it appears overall that digital sources and related metadata are compartmentalised within their digital research ecosystem. Notes and transcriptions are not directly linked to the digital reproduction of the sources, and it is not possible to seamlessly navigate between them. Some researchers deplore the lack of a tailor-made system to manage historical documents, for which a reference management software is not well suited. Those thought of alternative systems with varying degrees of satisfaction. One participant uses a qualitative data analysis software, originally designed for social sciences research, to code and analyse transcriptions of archival records. Others contemplated to tag their photos, but the large amount of files makes it impracticable.

"Il manque un outil pour les historiens, une base de données qui sera à destination des historiens. Il existe un programme qui s'appelle Zotero, qui permet de faire un peu de classement et de prendre des notes, mais on sent bien que l'outil est orienté essentiellement références bibliographiques et pas catalogage ni recherche, donc là il manquerait un module dans Zotero pour faire du classement et du traitement de sources." (2.4)

"Ce qui serait idéal, ce serait de taguer les photos d'archives. J'ai essayé ça pendant un petit temps pour ma thèse, mais on peut faire ça avec 500 photos, mais c'est impossible pour 35.000, c'est juste impossible, ça sous-entend qu'on ait un projet où des gens travaillent pour soi pour réaliser ça." (2.8)

"Je me suis déjà posé la question de savoir si je ne devrais pas utiliser, trouver un outil qui permettrait d'aller plus loin, de taguer les photos, etc. Le problème, c'est que je n'ai jamais le courage, parce que je me dis que je ne vais savoir aller à rebours de ce que j'ai accumulé, il faudrait quasiment que je puisse me prendre trois ou quatre mois pour faire la bascule, chose pour laquelle je n'ai pas le temps. Donc j'ai déjà réfléchi, mais il n'y a rien qui m'a convaincu à 100 % non plus." (2.10)

Whereas the management of bibliographic references becomes increasingly standardised, interviewed scholars use highly diverse customised systems to manage their research data. This seems counterintuitive with notions of portability and potential reuse of data in the wake of open science. History is a traditionally seen as an individual discipline full of "lone scholars". Although a few interviewees do collaborative work, sharing digital sources on a cloud with students or research teams, for most of them, collaboration is not a common practice. Furthermore, the question of open data, which is now strongly promoted by funding agencies, remains a matter of debate among interviewees. Concerns are expressed as regards fear of errors, improper reuse of data that was collected with great effort over a long period of time, and copyright issues.[32]

As for the analysis, traditional methods such as cross-searching through collections and comparing series of sources still prevail. Computational methods and tools developed within the Digital Humanities are not much used (yet) among interviewees, although a few researchers are experimenting with more innovative methods, such as network analysis, data visualisation and geolocation, and computational image analysis.

2.5. Temporalities, work environments, and digital turn

This section explores three themes linked to researchers' work practices: time management, the reading room as a work environment, and the impact of new technologies on the research workflow.

The challenge of time management

Over the course of the interviews, time management frequently appeared as one of the important challenges for researchers. This problem is closely associated to the context of the academy and to the new project-based funding model that now also pervades research in the arts and the humanities. Faculty members undertake an increasing amount of administrative tasks on top of their teaching duties, which leaves them with limited time for actual research. On the other hand, the increasing number of fixed-term contracts against the decreasing number of permanent positions has led to a casualization of academic employment. As a result, early career researchers are under pressure to publish as much and as quickly as possible, a phenomenon epitomised in the phrase "publish or perish".[33]

> *"Si on regarde aussi tout ce qui pèse sur les professeurs, sur les chargés de cours, on a tendance à établir des priorités et à choisir des schémas qui sont faisables et qui donnent un rendement." (2.8)*

> *"Plus on vieillit, plus on enseigne, moins on a le temps de faire de la recherche juste personnelle." (2.9)*

[32] On these subjects, see Rygiel 2011; Dacos 2013.
[33] On this topic, see Bouffartigue and Lanciano-Morandat 2013; Cauchard and Vilardell 2013.

"To be honest, it is a luxury for me to find a day to be able to work in archives. For the rest of the time, and it is not to complain, but I am mostly working at the university, doing my work, and I am doing my research during the weekends and in the evenings." (2.12)

"Ça tient sans doute aussi au statut de l'incertitude de la recherche, je suis encore sur des postdocs, contrats temporaires, donc il faut toujours être le plus efficace possible, produire le plus rapidement possible." (2.10)

Consequently, the time slot that researchers can allocate to visit an archives or library reading room is often very limited. This comes in addition to external constraints such as opening hours, geographic distance or financial constraints, especially in the case of research trips abroad. It is therefore essential for researchers to maximise the time spent there in collecting as many sources as possible in a minimum time.

"La contrainte de collections qui ne sont pas numérisées m'oblige à me rendre sur place, et donc dans le cas des sources [étrangères], ça prend du temps, il faut dégager du temps et trouver des institutions pour m'accueillir sur place." (2.4)

"Une fois par semaine, je vais à la bibliothèque et j'organise ça de façon à consulter un maximum de sources en une journée." (2.6)

"Je ne vais pas pour une notice à un endroit, quand j'ai une masse critique qui justifie le déplacement, je passe quatre heures à faire ce qu'il faut, à rassembler l'information." (2.2)

To make their research trip as most efficient as possible, researchers dedicate some time ahead to carry out a careful preparation work. They familiarise themselves with the way in which the institution works, with its collections. They get in touch with archivists or curators. First and foremost, they explore online finding aids to make a first selection of sources that they need to check first.

The lesson to be learned in terms of improving online services from the perspective of heritage institutions is that all kinds of resources that facilitate access to the reading room and help researchers to get straight to the point, that is studying the sources, are particularly useful. For instance, shortening the registration procedure, helping them in their orientation within the reading room, allowing them to get acquainted with the collections, or requesting items online in advance.

"Les archives [...] sont ouvertes pendant une longue matinée, de 9h à 14h, donc il s'agit d'être vraiment rentable. Donc ce que je faisais, c'était que je préparais au maximum ma visite dans le centre d'archives avec les instruments dont je disposais, et puis une fois que j'avais épuise ce que j'avais préparé chez moi en archives, je passais à une seconde phase où j'ouvrais les inventaires disponibles sur place et où je pointais l'un ou l'autre fonds qui pouvait s'avérer intéressant." (2.4)

"La plupart du temps, maintenant on a les inventaires en ligne, en un clic on a le PDF qui arrive. C'est quand même là vraiment très intéressant et c'est un gain de temps aussi pour la préparation des recherches dans les sites en tant que tels, surtout quand c'est à l'étranger, d'ailleurs." (2.7)

"I think a lot of inventories for instance, these could be digitized. (…) Some of them are. But sometimes, handwritten things, inventories or files are available within the reading room. The problem is that I do not have the time to go there, say, every month, unfortunately, so that could be very useful, I think, these entries to be available online." (2.13)

"On a les inventaires qui sont en ligne, on sait exactement ce qu'on veut, on clique, on dit que c'est là à telle date, on arrive à 9h et on commence à travailler, on a tout à sa disposition." (2.2)

"Si je sais qu'en arrivant j'ai les sources qui sont prêtes à être consultées, je gagne deux heures de temps, ça pourrait être pas mal." (2.4)

"On envoie nos demandes, ils préparent la commande et puis on arrive et c'est beaucoup plus rapide. Donc c'est vraiment une façon assez efficace." (2.7)

"Je fais souvent mes recherches préalables, après je prends contact pour confirmer, pour préparer la visite, pour rentabiliser la visite. Il y a l'aspect humain qui reste important, beaucoup parce que j'ai pas mal dépouillé à l'étranger. Il faut être rentable quand on part et qu'on a qu'une semaine sur place. C'est bien d'avoir des cartons qui soient prêts dès qu'on arrive." (2.10)

Work environment

The globalisation of scientific research, the increased mobility of researchers, and the large digitisation initiatives from heritage institutions redefine the role of an archives or library reading room in the academic work practices. Undeniably, digitised documents offer a significant value to researchers in terms of convenience, but also opportunities to develop new methods of research, as discussed above (2.2).

"Le rêve pour moi, ce serait vraiment de pouvoir manipuler les catalogues, les inventaires en ligne et obtenir directement l'information de manière numérisée. Je suis un historien en chambre, donc j'aime bien mettre mes pantoufles quand je suis chez moi, boire une tasse de café et travailler sur mon ordinateur." (2.4)

At the same time, research trips to the archives or library reading rooms remain unavoidable for the interviewees, who primarily work on non-digitised sources. In fact, many of them even confess enjoying these work sessions.[34]

"J'aime beaucoup aller aux archives, donc je ne voudrais jamais que tout soit numérisé et mis en ligne." (2.10)

First of all, an archives or library reading room is perceived as a calm environment where researchers can focus and disconnect from daily distractions. One researcher even praised the poor quality of the Wi-Fi signal at the Royal Library (which has been resolved since then). Visiting the reading room also gives them the opportunity to get in touch with curators or archivists who work there.

[34] On the relationship between researchers and their work environment in a digital era, see Rimmer *et al.* 2008.

"Idéalement, tout de même, j'aime bien travailler sur place, être en contact avec les archives, dans un cadre surtout propice à la recherche." (2.7)

"Au moins, l'avantage, c'est que parfois, quand on va aux archives, on est obligé de travailler." (2.10)

"Généralement, je me rends en bibliothèque pour écrire. Comme le WIFI est très mauvais à la KBR, je m'y installe pour écrire (…) c'est un grand avantage, c'est qu'on est tranquille." (2.8)

"J'aime bien aller dans les archives (…) je trouve qu'il y a à chaque fois aussi un contact personnel qui s'établit avec l'archiviste." (2.9)

Beyond that, handling original documents also offers a significant added value, one that strikes an emotional chord particularly among historians.

"Juste pour le plaisir, en tant qu'historien, ce serait triste de devoir tout faire devant son bureau, devant un PC, devant un écran, on perd le contact." (2.5)

"Then we have the document in hand, and it is an emotional feeling of course." (2.13)

For historians working on texts, photos or scans of archives is simply a way to reproduce the content faster than a transcription. However, there are aspects for which the digitised version cannot entirely replace the original document. Scholars studying the materiality of historical documents use photos as a support for visual analysis of images, layout, and so on, but the digitised version comes in addition, but can never be a substitute for the autopsy of the original (see 2.2).

Digital turn

Another subject that came to the fore during interviews is the impact of the digital turn on researchers' work practices. Many interviewees are aware of being on the other side of a digital divide as regards the use of digital tools and methods especially for data management and analysis. Most of them got their degree before or at the same time of the digital revolution. Training in the use of these tools was not a part of their university curriculum, and they had to learn by doing. Furthermore, few of them work on digitised documents apart from photos that they took in archives and libraries. Researchers often sound apologetic for their so-called "traditional" or "old-fashioned" practices. A significant example is the use of a reference management software, which some researchers resist using, despite acknowledging its benefits.

"Donc là on sent bien qu'on est au milieu du gué, je suis un peu entre-deux." (2.7)

There are many reasons which could provoke resistance on behalf of researchers in adopting a new tool:

- Time-consuming training, important learning curve
- Tools that do not answer needs
- Sustainability issues

- Non-exportable data
- Loss of control over the data
- Tool determining the research instead of the opposite
- Lack of interoperability
- Too many tools
- Lack of information

Researchers highlight the need for training to these new tools during the university curriculum, in association with an epistemological reflection on the consequences of their use and the necessity to keep methodological safeguard.

> *"C'est en perpétuelle recherche. Je veux dire que c'est un questionnement permanent, parce que la révolution est arrivée en même temps – j'ai commencé mes études en 2000, donc « je l'ai vécue » au fur et à mesure de mes étapes. C'est vrai que quand j'y repense, en première candi on nous expliquait pour la bibliographie d'acheter des fiches papier de couleur différente, et selon la thématique, fiche jaune, fiche rose, et d'avoir une boîte. Et je sais que je suis allé à la papeterie ici acheter des fiches de couleur et j'ai commencé ma formation là-dessus, donc en soi c'est surréaliste de se dire qu'on nous a appris ça – moi j'avais 18 ans en l'an 2000 et on m'a appris ça, mais en même temps, je trouve que ça forme à la méthode, à l'esprit, c'est intéressant. Et puis voilà, l'ordinateur est arrivé petit à petit." (2.10)*
>
> *"Disons que j'ai à peu près grandi avec l'avancée des technologies, donc l'iPad est venu quand je faisais ma thèse, si je regarde en arrière ça m'a beaucoup aidé, ça m'a permis de travailler constamment sur des sources d'archives, et de débrancher et de me brancher quand j'en avais envie, ça a débloqué énormément la situation, je crois que j'ai fait 30.000-35.000 photos d'archives [...] pour ma thèse. Sans cela, ça aurait été assez difficile." (2.08)*
>
> *"Ma crainte, c'est que ce florilège d'outils fasse que ce soient les outils eux-mêmes qui déterminent la recherche. Et ça, c'est vraiment un problème. Ça c'est un risque. Le risque en matière de déluge d'information, il est là, on l'a évoqué, il me semble moins problématique que l'effet inverse, c'est-à-dire qu'une recherche déterminée par la technologie plutôt que l'inverse." (2.7)*
>
> *"De mon expérience personnelle, je pense que le problème, c'est qu'on n'a pas – j'ai été formée avant la révolution numérique, donc ce sont beaucoup des choses que j'ai apprises sur le tas – il serait vraiment essentiel de former les chercheurs de demain à ces choses-là. Je pense que même ici dans notre centre de recherche, on a une petite équipe de jeunes, de doctorants qui nous initient aussi à toute une série de choses qui sont hyper intéressantes, mais par rapport aux grands centres d'archives et aux grandes bibliothèques, la manière dont circule l'information sur les outils qui sont disponibles serait importante à améliorer, mais je ne pense pas que la multiplication des outils soit vraiment un frein. Il faudrait déjà qu'on apprenne à mieux les connaître en général, à mieux connaître aussi ce qui est disponible." (2.1)*

2.6. Researchers as users of archives and libraries' digital services

Current resources

A section of the questionnaire focused on how researchers use the online resources provided by the State Archives, the CegeSoma, and the Royal Library and how satisfied they are with them. This question generated a lot of very useful comments on issues they frequently experience with those resources and on their requirements for future improvements. Their remarks were integrated with the findings of the general survey of the MADDLAIN project. The key points concerned:

- Information on the content, as regards descriptions of items, but also on the collections preserved in the institutions and on what is catalogued or digitized.
- User-friendliness and simplicity of user interface.
- Lack of understanding as regards access restrictions due to copyright.

In addition, interviewees were invited to cite resources developed by other archives and libraries that they found particularly useful for their research, and conversely, resources particularly ill-conceived. Responses were very informative as to what researchers consider a good practice. Three main criteria stand out:

- Access to digitised documents and quality of digitisation.
- Searchability either through keyword search or browsing.
- Ergonomic design and usability.

Features concerning visualization and analysis of sources do not come to the fore in this context. It is noteworthy that some researchers appreciate platforms which offer a single point of contact to several collections. For instance, many researchers use Unicat, the Union Catalogue of Belgian Libraries instead of their institution's library individual catalogue, because its interface is simple and it allows them to directly locate an item even if their usual library does not have it. However, library catalogues such as the "Explore" or "Discovery" model, which allows to search both within the library's collection and external content available on the internet, is controversial.

> *J'apprécie surtout les catalogues qui ont la possibilité d'éliminer tout le matériel en ligne : généralement, c'est conçu pour les étudiants et pour le personnel de tout trouver, mais évidemment, comme visiteur externe, quand on consulte le catalogue d'une autre institution, c'est forcément pour quelque chose qu'on n'a pas chez soi, donc pour la collection sur papier. (2.8)*

> *Ensuite on a accès à la fois à des ressources numériques, à des références d'ouvrages papier, à des comptes rendus, à des revues payantes, et à des références de revues payantes que nous n'avons pas. Et donc on a accès à une masse d'information qui semble prometteuse dès le départ, et puis qui s'avère très vite décevante parce qu'on n'a pas accès au texte définitif, au final. Cet outil-là qui est mammouth et qui veut répondre à toutes les attentes n'aide plus, on se perd dedans. (2.4)*

That is my primary tool to find both databases and secondary literature. I don't really use bibliographic database or anything all that much anymore, because we now have this tool that promises to give you everything at once. (2.15)

Future developments

A list of suggestions for new features that could be implemented to current services was proposed to interviewees, who were invited to tell whether they would be prone to use them.

User account to save work documents, references, searches

Most researchers think that this feature would be useful to save work documents or searches in the tools they most often use, and especially if it is linked to an option to order items online and in advance.

Pros: Access data from anywhere and managing requests of documents.

Cons: Decentralisation of data and managing multiple user accounts. Researchers work with collections from many institutions, and usually prefer to download and store their documents within their own digital ecosystem, which they organise as they see fit. Storing items on the website of the archives or libraries also means that they need to remember where the document comes from before they can recover it.

Annotation of digitised documents or item descriptions

The same issue as for the user account arises: researchers usually keep their notes within their personal organisational system. In order for this tool to be useful, it would therefore need an export feature for notes.

Comparing digitised documents

Most interviewees are interested in the content of archives and fail to see the benefit of a comparison tool, which they rather associate with visual or philological studies (comparing images or various editions of a same work). Others already compare documents "manually", for instance with two monitors or by copy-pasting images in a word processing document, and do not see the added value of such a tool.

Transcribing documents

Many researchers who work on handwritten documents or on audio-visual archives are used to transcribing them, at least partly. To do so, they normally open a word processing document and do not always see the added value of a specific online tool for this purpose. Furthermore, in order to exploit their transcriptions, they would need to centralise all of them at some point, and would therefore require an export feature.

Analysing documents

Most interviewees did not spontaneously have specific ideas as to which analytic tools could apply to their research, especially since they often work on non-digitised documents. However, many of them showed some interest for tools allowing to perform

lexical analysis, to study word frequency, or to integrate reference works or dictionaries, for instance.

Exporting references to a reference management software

Interest in this feature obviously depends on whether researchers use a reference management software (see 3.3), but most interviewees strongly support it.

Further suggestions

Researchers were invited to suggest other ideas for new features or tools that could be implemented to the current services:

- Improving the searchability and adding an export option for images, in particular illustrations from newspapers.
- Linking items with related collections and documents.
- Enriching item descriptions through user contribution (collaborative tagging).
- Implementing a timeline creation tool.

Collaborative platforms and virtual research environments

A section of the interview focused on how familiar researchers were with virtual research environments (VRE) and whether they would be interested in using one if it fitted their research needs. VREs were defined for the purpose of the interviews as collaborative web platforms supporting the scientific use of data – in this case, collections of documents – by integrating tools to analyse them. It quickly appeared that VRE was a rather abstract concept for interviewees, and it was decided after the third interview to provide a visual support with a few examples to help them better comprehend it. Two thirds of the interviewees never heard of this kind of environment, and none already used one for their research, but most of them showed some interest in their potential to share data and work collaboratively on the same documents, without necessarily a precise idea on how to apply it to their research. Only four interviewees thought that VREs would not be useful. Among the issues that were raised was the fact that most researchers are conducting research on their own, without being part of a large team. They also pointed out a lack of awareness of the existing possibilities.

At a second stage, interviewees were presented with the scenario of a collaborative platform developed by the State Archives, the Royal Library, or the CegeSoma, with the purpose of valorising their collections, for instance through collaboratively annotating and transcribing documents, publishing blog posts, creating collaborative bibliographies on specific topics, etc. Researchers were asked whether they would be willing to participate to such a platform and to what extent. Their responses are summarised here:

Pros:

- Offers outreach potential for their research, which most interviewees view as an essential part of their job, and an opportunity to valorise the fundamental role of heritage institutions.

- Directly benefits their research.
- Initiates future collaborations.
- Offers potential in terms of projects for university students.
- Allows amateur researchers to share their work.

Cons:

- Lack of time for peripheral activities which are not much valued in terms of career advancement. This is especially important given the employment insecurity and fierce competition in the academy.
- Reluctance to share work products such as annotations and transcriptions in open access, either because they view them as preparatory work or by fear of being plagiarised.
- Competition with similar actions initiated by their own institutions.

Furthermore, researchers emphasise the need for a strong editorial work to ensure both quality control of the contributions if the platform is open to all and the platform's sustainability over time.

Another scenario consisted in implementing an online environment such as a forum to enhance collaboration between researchers and professionals from the institutions. Interviewees who already had regular contacts with curators or archivists were more dubious as to the interest of this platform, since other means of communication already exist. Others showed more interest in this idea.

3. Summary

A user requirements survey was conducted from November 2016 to January 2017 via semi-structured interviews with 15 researchers, mainly historians, from Belgian universities and research institutions. The main purpose of this survey was to get insight into how the State Archives, the CegeSoma, and the Royal Library can improve current services and implement new tools in support to scientific research, by understanding their users' research practices, particularly in the light of the digital revolution. The key themes that emerged in their discourse concerned information search, access to information, ways to collect and manage data, time management, work environment and digital turn.

Researchers implement various strategies to search for information. Keyword search is often used as a starting point or in an exploratory way, but needs to be complemented with other strategies such as citation chaining, browsing, and monitoring, in order to acquire the most comprehensive overview possible. The expertise of professionals in the heritage sector is also perceived as an important resource to find primary sources.

Access to information turned out to be a key priority for researchers. Although they particularly value having online access to digitised sources, those form just a small part

of their documentation. Visiting an archives or library reading room to study original sources is still an absolute requirement for all interviewees, in particular those conducting research on the materiality of historical documents.

There are various ways in which researchers collect documents relevant to their research. Preferably, they will get a digital copy by downloading or taking pictures and store it in their own digital ecosystem, thus postponing the thorough reading and analysis of the documents to a later stage. Only occasionally and when time allows will they read the documents on the spot and carefully sort out which documents need to be photographed.

The use of a reference management software to organise their bibliographic material becomes increasingly common among researchers. Concerning primary sources, researchers struggle with the lack of an adequate system that could link digital versions of documents to their annotations. Sharing data and using computational tools to analyse historical documents are far from standard practices among interviewees.

Although researchers still need, and mostly enjoy, visiting an archives or library reading room, external constraints mainly related to their working conditions do not leave them much time to do so. As a consequence, they need to make the most of their visit, by conducting careful preparatory work ahead, for which online access to digitised finding aids and the possibility to order items in advance are particularly useful.

Finally, many interviewees are aware of being a victim of a digital divide as regards the use of digital tools to manage their data. In addition, most of them do not work with digitised documents (other than pictures they took themselves) and few are using the new methods for analysing historical documents that are being developed in the field of digital humanities. “The old-fashioned way”, “old-school methods”, were recurring phrases in their discourse. It seems that we are now in a period of transition, where most established researchers had to train themselves to these new methods and tools and learn by doing, while a new generation of researchers graduating from university are more aware of the possibilities offered by digital technologies and use them in a more intuitive and spontaneous way. Five years from now, the findings from a similar survey may very well be significantly different.

CONCLUSIONS AND RECOMMENDATIONS

1. Access to the collections

Among the findings from the user requirements survey, we may highlight five key points that have implications for institutions wishing to improve their online services.

1. Researchers use multiple strategies to retrieve information, both online and in traditional ways, in a quest for comprehensiveness.
2. Researchers gather and manage their data within their own digital ecosystem.
3. Optimal access to collections is researchers' top priority. In comparison, they are less interested in tools or features to process sources online.
4. On-site research is still a requirement, but the time researchers can allocate to visit archives and libraries is limited.
5. The use of online resources to search for information is now firmly anchored in research practices. However, most researchers are aware of being on the other side of a digital divide as regards the use of digital tools and methods for data management and analysis.

Based on these observations, we can make suggestions to improve researchers' user experience with online services.

1.1. Providing multiple gateways to access collections

Keyword search is the typical entry point provided by heritage institutions to access their collections online. However, findings from the survey show that researchers are often dissatisfied with online search engines, because they often generate a lot of noise and result in multiple blind spots. Furthermore, researchers need to have a prior idea of what they are searching in order to formulate a relevant query. Since one of the main principles underlying researchers' information behaviour is comprehensiveness, keyword search is therefore always combined with other strategies, such as browsing, citation chaining, consulting professionals, and monitoring.

As a result, institutions should work towards optimising their catalogues following those lines:

- Improving indexation and searchability.
- Providing more information on the content through detailed descriptions, making clear which parts of the collection have been indexed and which have not.
- Creating user-friendly and modern user interfaces.

In addition, institutions should think about ways to offer multiple gateways to their collections. For archives, presenting the archival fonds in a hierarchical structure seems like an optimal way to allow researchers to get an overview of the collections and to seamlessly browse through them. As regards digitised collections, the concept of generous interfaces offers creative and innovative ways to present them: "Generous interfaces provide rich, navigable representations of large digital collections; they invite exploration and support browsing, using overviews to establish context and maintain orientation while revealing detail at multiple scales."[35] Citation chaining is also an information retrieval technique that is highly valued by researchers, albeit rarely translated online. Improving the linking between collections or items descriptions could be a first step in that direction. Figure 6: Suggestions to improve access in regards with scholarly information behaviour synthesises suggestions to optimise online access to heritage collections.

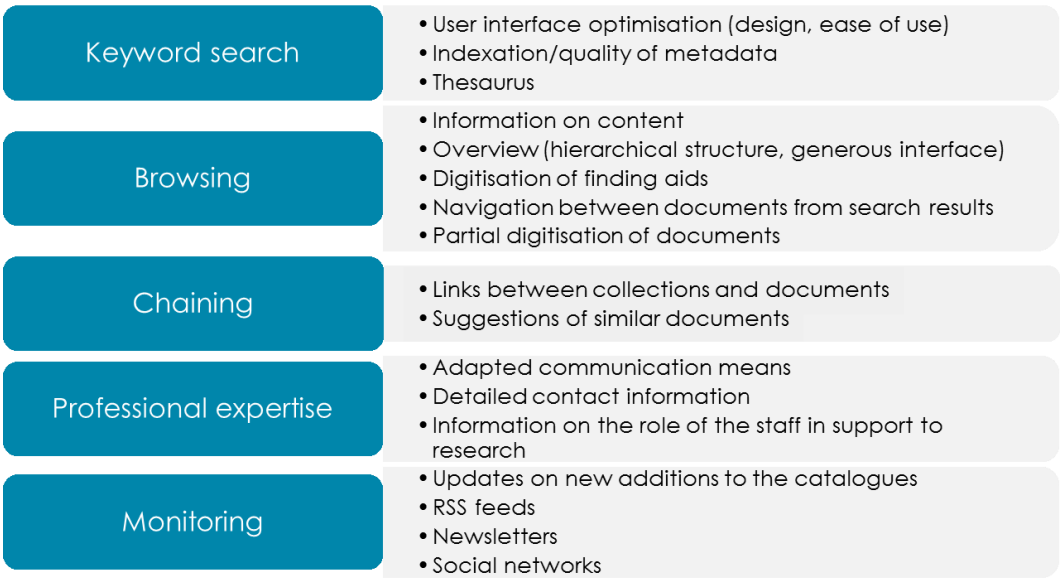


Figure 6: Suggestions to improve access in regards with scholarly information behaviour

1.2. Providing online tools to prepare on-site research

Time management turned out to be a major challenge in researchers' work practices. Academics in full time positions face an increasing amount of administrative tasks and teaching duties, which leaves them limited hours for actual research. On the other hand, early career researchers in fixed-term contracts are under pressure to publish as much and as quickly as possible to secure future employment. While research in archives and libraries is still a requirement for researchers, who work predominantly with non-digitised sources, they wish to maximise the time spent on site collecting as many sources as they can. All kinds of tools and resources that help them prepare their visit in advance are therefore particularly welcome. Those can fill various purposes, such as:

[35] Whitelaw 2015.

- Becoming familiar with the institution and understand how the reading rooms work:
    - Practical information on access, opening hours, rules for use, etc.
    - Video tutorials on the theme “first visit”.
- Shortening administrative procedure:
    - Online registration form.
    - Option to request items online in advance.
- Getting acquainted with the collections:
    - Overview of the collections held by the institution.
    - Finding aids available online.

2. Virtual research environments and digital research tools

There are several research programmes ongoing at the national or European level aimed at reflecting on the development of large-scale research infrastructures and virtual research environments in the Humanities. Many of these infrastructures are still very much works in progress and few are fully operational. Some of these projects do not go beyond the prototype stage, others experience low uptake by users, either because they do not meet their needs, are too complicated, or suffer from poor communication. It is therefore of paramount importance not only to conduct a careful preliminary research to assess users’ requirements, but also to involve them at every stage of the development process, thus following a bottom-up approach.

The literature review on virtual research environments also shows that it is difficult to come up with one single definition which would encompass all scenarios, since they are designed to meet various purposes, either generic or specific. In the context of the MADDLAIN project, VREs were defined as collaborative platforms promoting the scientific processing of data, in this case collections of sources, through the integration of digital research tools.

The MADDLAIN institutions are currently involved in several research projects aimed at building platforms or digital research tools, but they mainly act in their capacity as content providers. For heritage institutions, developing VREs would present several advantages such as:

- Promote their collections in an innovative way, by supporting new methods of scientific research.
- Build bridges between scattered collections.
- Develop further collaborations between researchers and information professionals.
- Put their expertise in information sciences and data curation into action.

Furthermore, VREs seem to be the logical next step in terms of digital services offered by archives and libraries. A digital roadmap published recently by the National Library of

France (2016) announces the future development of a new service for researchers, to offer "a secure infrastructure that allows them, according to the data they are interested in and in particular the legal status of the collections they are studying, to build their research corpora and explore them, either with their own tools or with those of the BnF."[36] These kinds of services aim to respond to a tendency to do more and more things directly online. Consecutive reports on Gallica, the digital library of the National Library of France, exemplify this trend. According to a user survey conducted in 2011, 57% of respondents declared often downloading documents while only 31% of respondents declared often carefully read documents online. According to the user survey conducted in 2016, however, the percentage of respondents who often or always carefully read documents online (66%) is now more or less equal with the percentage of those who often or always download documents (65%). The report attributes this evolution to the fact that the viewer had been considerably improved, which makes reading documents online easier and more pleasant.[37]

However, the findings of the user research conducted in the context of the MADDLAIN project show that, as of yet, virtual research environments and other kinds of collaborative platforms are not necessarily a requirement of researchers with regard to digital services provided by archives and libraries. In this respect, the survey highlighted a number of possible issues:

1. Most interviewed researchers are unaware of what VREs are and how they could benefit their research.
2. If they are aware of it, they mainly consider VREs as useful to support collaborative research projects, and therefore as a customised environment designed for a specific purpose. However, the MADDLAIN institutions, and particularly the Royal Library and the State Archives, have a more generic focus. A "one-size-fits-all" approach to encompass all types of documents and research interests may actually miss the point of researchers' needs.
3. VREs to manage and process research data seem counterintuitive with the notion of personal digital ecosystem highlighted during the survey.
    - Researchers use sources from various institutions. They need to centralise their documents and annotations in order to be able to retrieve them easily at later stages and to conduct broad comparisons. It would be impractical for them to navigate between multiple online environments, accounts, passwords, and to depend on external systems. In addition, it implies that they would have to remember the provenance of the source in order to retrieve it.
    - The organisation of their research data management system is driven by their research questions. Their system therefore needs to be flexible and customisable.

[36] http://www.bnf.fr/documents/digital_roadmap.pdf#page=115 [accessed 2017.05.23].
[37] GMV Conseil 2012; TMO Régions 2017.

4. Few researchers work with digitised sources provided by the institutions, but rather on photos of documents that they took themselves.
5. Most researchers are reluctant to openly share work products or research data if it is not in published form.
6. Most researchers are reluctant to adopt new tools if there is a high learning curve, if it requires a major time investment, or if the benefit for their research is not immediately visible.

According to the survey, some researchers struggle to manage their research data effectively. This could be partly due to the lack of an adequate, user-friendly, tool to organise and annotate historical sources. If the MADDLAIN institutions were to provide a VRE enabling researchers to process historical sources, it would need to allow imports of items from various sources, which would potentially require a lot of server space, and be flexible enough so that researchers do not feel constrained in the organisation of their data in respect with their research projects (cf. no. 3 above). However, institutions should assess if providing such a data management system is within the scope of their missions as heritage institutions or if they should rather stick to their role of content provider. Incidentally, a system like Tropy, a research photo management system currently under development by the Roy Rosenzweig Center for History and New Media at George Mason University, may better fit researchers' needs as it would provide greater interoperability and more easily enable collaborations at an international level.[38]

The possibility to develop a platform aimed at expanding collections through collaborative annotations and transcriptions was also contemplated in the context of MADDLAIN project. Researchers' reluctance to share their work and their lack of time for peripheral activities (cf. no. 5 and 6 above) suggests that massive uptake would be unlikely. These crowdsourcing initiatives are rather viewed by researchers as primarily targeted for people with a lot of leisure time. However, they could be particularly adequate to host student projects promoting heritage collections.

To conclude, in light of the current budgetary situation of MADDLAIN institutions, who suffer from important lack of funding and staff, findings of this study do not lead towards recommending the implementation of virtual research environments as part of their digital services, but rather to focus on priority issues which are improving online access to the collections and providing tools to help users maximise their visit in reading rooms. Instead of an all-encompassing environment, another, perhaps more sustainable, approach would be to reflect on a the possibility to add a suite of individual tools which could be selected by researchers depending on their research purpose, as presented in section 2.6 of the user requirements chapter of this report. Furthermore, the institutions have a valuable role of content provider and expert in data curation that they should continue to offer through their participation in research projects developing VREs and digital research tools. At this time, there are still many uncertainties about the field of

[38] http://tropy.org.

research infrastructures and virtual research environments and it is not clear where this path will lead nor what will be the impact on research practices in the Humanities. Given the fast pace of technological changes and the probable evolution of research practices in the future, it would be advised to reassess this opportunity in a few years.

BIBLIOGRAPHY

ACLS 2006. Our Cultural Commonwealth. The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences.

Anderson, S. 2013. "I. Thinking Infrastructure. What are Research Infrastructures?," *International Journal of Humanities and Arts Computing* 7, 4–23.

Anderson, S. and T. Blanke 2012. "Taking the Long View: From e-Science Humanities to Humanities Digital Ecosystems," *Historical Social Research/Historische Sozialforschung* 37, 147–64.

Anderson, S., T. Blanke and S. Dunn 2010. "Methodological commons: arts and humanities e-Science fundamentals," *Philos Transact A Math Phys Eng Sci* 368, 3779.

Artières, P. 2016. "Du dépouillement au dévoilement. L'historien et les archives numérisées," in Roustan, M. and A. Monjaret (edd.), *La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques* (Papiers).

Atkins, D.E. *et al.* 2003. Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.

Atkins, D.E. *et al.* 2009. Report of the International Panel for the 2009 Review of the UK Research Councils e-Science Programme.

Benardou, A. *et al.* 2010. "A Conceptual Model for Scholarly Research Activity," in Unsworth, J., H. Rosenbaum and K.E. Fisher (edd.), *iConference 2010 Proceedings* (Urbana-Champaign (ILL)) 26–32.

Blanke, T. and M. Hedges 2013. "Scholarly Primitives: Building institutional infrastructure for humanities e-Science," *Future Generation Computer Systems* 29, 654–61.

Bouffartigue, P. and C. Lanciano-Morandat 2013. "Les temporalités de la recherche, introduction," *Temporalités. Revue de sciences sociales et humaines* .

Brunvand, S. and M. Duran 2010. "A Taxonomy of e-Research Collaboration Tools: Using Web 2.0 to Connect, Collaborate and Create with Research Partners," in Anandarajan, M. and M. Anandarajan (edd.), *e-Research Collaboration* (Berlin, Heidelberg) 105–16.

Buddenbohm, S. *et al.* 2015. "Success Criteria for the Development and Sustainable Operation of Virtual Research Environments," *D-Lib Magazine* 21.

Candela, L., D. Castelli and P. Pagano 2013. "Virtual Research Environments: An Overview and a Research Agenda," *Data Science Journal* 12, GDRI75-GDRI81.

Carusi, A. and T. Reimer 2010. Virtual Research Environment. Collaborative Landscape Study. A JISC funded project.

Cauchard, L. and I. Vilardell 2013. "L'histoire face aux financements sur projet : autonomie professionnelle, temporalités et organisation de la recherche," *Temporalités. Revue de sciences sociales et humaines* .

Dacos, M. 2013. "Cyberclio. Vers une Cyberinfrastructure au coeur de la discipline historique," in Clavert, F. and S. Noiret (edd.), *L'histoire contemporaine à l'ère numérique. Contemporary history in the digital age* 29–41.

Deschamps, S. 2010. Enquête sur le profil des lecteurs des Archives de l'État de l'année 2008.

Dombrowski, Q. 2014. "What Ever Happened to Project Bamboo?," *Lit Linguist Computing* 29, 326–39.

Dovey, M. 2015. "Chapter 6: International Collaboration and the Changing Digital World - Opportunities and Constraints," in Hall, M., M. Harrow and L. Estelle (edd.), *Digital Futures. Expert Briefings on Digital Technologies for Education and Research* (Burlington) 41–50.

GMV Conseil 2012. Evaluation de l'usage et de la satisfaction de la bibliothèque numérique Gallica et perspectives d'évolution. Rapport détaillé.

Kemman, M. and M. Kleppe 2015. "User Required? On the Value of User Research in the Digital Humanities," *Selected Papers from the CLARIN 2014 Conference, October 24-25, 2014, Soesterberg, The Netherlands* (Linköping Electronic Conference Proceedings).

Kemman, M., M. Kleppe and S. Scagliola 2014. "Just Google It - Digital Research Practices of Humanities Scholars," in Mills, C., M. Pidd and E. Ward (edd.), *Proceedings of the Digital Humanities Congress 2012* (Studies in the Digital Humanities).

Koureas, D. *et al.* 2016. "Community engagement: The 'last mile' challenge for European research e-infrastructures," *Research Ideas and Outcomes* 2, 2, e9933.

Lemercier, C. 2014. Les grands corpus en ligne changeront-ils la boîte à outils de l'historien.ne ?

Lipp, A. 2009. "An integrated digital research environment: DFG perspectives," *Library Hi Tech* 27, 496–504.

Llewellyn Smith, C. *et al.* 2011. Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century.

Palmer, C.L., L.C. Teffeau and C.M. Pirmann 2009. Scholarly Information Practices in the Online Environment. Themes from the Literature and Implications for Library Service Development (Dublin OH).

Pothen, P. 2007. Developing the UK's e-infrastructure for science and innovation.

Priestner, A. and M. Borg (edd.) 2016. *User Experience in Libraries. Applying Ethnography and Human-Centred Design* (London).

Rimmer, J. *et al.* 2008. "An examination of the physical and the digital qualities of humanities research," *Information Processing & Management* 44, 1374–92.

Roustan, M. 2016. "La vague numérique et le ressac du réel. Du rapport aux sources des chercheurs en sciences humaines et sociales," in Roustan, M. (ed.), *La recherche dans les institutions patrimoniales : sources matérielles et ressources numériques* (Papiers) 170–92.

Rygiel, P. 2011. "L'enquête historique à l'ère numérique," *Revue d'histoire moderne et contemporaine* 58, 30–40.

Speck, R. and P. Links 2013. "The Missing Voice: Archivists and Infrastructures for Humanities Research," *International Journal of Humanities and Arts Computing* 7, 128–46.

TMO Régions 2017. Bibliothèque nationale de France. Enquête auprès des usagers de la bibliothèque numérique Gallica.

Unsworth, J. 2000. "Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this?," *Symposium on Humanities Computing: Formal Methods, Experimental Practice* (King's College, London).

van Zundert, J. 2012. "If you build it, will we come? Large scale digital infrastructures as a dead end for Digital Humanities," *Historical Social Research/Historische Sozialforschung* , 165–86.

Verachten, L. 2011. *Het bezoekersprofiel van het Rijksarchief 2004-2008* (Brussel).

Voss, A. and R. Procter 2009. "Virtual research environments in scholarly work and communications," *Library Hi Tech* 27, 174–90.

Warwick, C. 2012. "Studying users in digital humanities," in Warwick, C., M. Terras and J. Nyhan (edd.), *Digital humanities in practice* 1–21.

Whitelaw, M. 2015. "Generous Interfaces for Digital Cultural Collections," *Digital Humanities Quarterly* 9.

Wilkins-Diehr, N. *et al.* 2008. "TeraGrid Science Gatways and Their Impact on Science," *Computer* 41, 32–41.

Yang, X. and R.J. Allan 2010. "Web-based Virtual Research Environments," in Yao, J. (ed.), *Web-based Support Systems* (London) 65–80.