



## MADDLAIN

Identifying Needs to Modernize Access to Digital Data in Libraries and Archives

Florence Gillet, AGR-CegeSoma - Jill Hungenaert, AGR-CegeSoma -  
Melissa Hodza, KBR - Stéphanie Paul, AGR-DO1 - Anne Chardonens, ULB -  
Martin Vanbrabant and Gerald Haesendonck, iMinds

Axis 5: Major societal challenges



## NETWORK PROJECT

### MADDLAIN

Identifying Needs to Modernize Access to Digital Data in Libraries and Archives

Contract - BR/153/A5/MADDLAIN

### FINAL REPORT

<b>PROMOTORS:</b>	Florence Gillet, AGR-CegeSoma Sophie Vandepontseele, KBR Rolande Depoortere, AGR-DO1 Seth Van Hooland, ULB Erik Mannens, iMinds
<b>AUTHORS:</b>	Florence Gillet, AGR-CegeSoma Jill Hungenaert, AGR-CegeSoma Melissa Hodza, KBR Stéphanie Paul, AGR-DO1 Anne Chardonens, ULB Martin Vanbrabant and Gerald Haesendonck, iMinds



Published in 2018 by the Belgian Science Policy  
Avenue Louise 231  
Louizalaan 231  
B-1050 Brussels  
Belgium  
Tel: +32 (0)2 238 34 11 - Fax: +32 (0)2 230 59 12  
<http://www.belspo.be>  
<http://www.belspo.be/brain-be>

Contact person: Christine MATHIEU  
Tel: +32 (0)2 238 34 93

Neither the Belgian Science Policy nor any person acting on behalf of the Belgian Science Policy is responsible for the use which might be made of the following information. The authors are responsible for the content.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without indicating the reference :

Florence Gillet, Jill Hungenaert, Melissa Hodza, Stéphanie Paul, Anne Chardonnes, Martin Vanbrabant and Gerald Haesendonck. **Identifying Needs to Modernize Access to Digital Data in Libraries and Archives**. Final Report. Brussels : Belgian Science Policy 2018 – 48 p. - BRAIN-be - (Belgian Research Action through Interdisciplinary Networks)

## TABLE OF CONTENTS

### CONTENTS

<b>ABSTRACT</b>	<b>6</b>
CONTEXT	6
OBJECTIVES	6
CONCLUSIONS	7
KEYWORDS	8
<b>1. INTRODUCTION</b>	<b>9</b>
1.1. THE DIGITAL REVOLUTION	9
1.2. THE NEED TO ENGAGE WITH THE AUDIENCES	9
1.3. A PROTEAN STUDY	9
<b>2. STATE OF THE ART AND OBJECTIVES</b>	<b>10</b>
2.1. GENERAL APPROACH	11
2.1.1. What are Digital Libraries and Archives?	11
2.1.2. Digitalization and Lack of Accessibility	11
2.1.3. Lessons Learned from Case Studies	12
2.2. SPECIFIC AUDIENCES	13
2.2.1. E-learning	13
2.2.2. Virtual Research Environments	14
2.3. ASSESSING GENERAL AND SPECIFIC USER NEEDS	15
2.3.1. Quantitative Approaches	16
2.3.2. Qualitative Approaches	18
2.4. REFERENCES	19
<b>3. METHODOLOGY</b>	<b>23</b>
3.1. OVERALL METHODOLOGY	23
3.2. FACE-TO-FACE INTERVIEWS.	24
3.3. THE WEB ANALYTICS	25
3.4. E-LEARNING	26
3.4.1. Current practices in the participating FSI	26
3.4.2. Identifying user needs with regard to digital mediation and describing user behaviour	27
3.4.3. Good practices in similar institutions	27
3.5. VIRTUAL RESEARCH ENVIRONMENTS	28
<b>4. SCIENTIFIC RESULTS AND RECOMMENDATIONS</b>	<b>30</b>
4.1. GENERAL RESULTS	30
4.1.1. Access to the collections	30
4.1.2. Communication	33
4.1.3. The valorisation of the collections	34
4.1.4. Studying User's Digital Practices and Needs in Archives and Libraries	34
4.2. THE MEDIATION TOOLS	36
4.2.1. Current practices in the participating FSI	36

4.2.2.	Identifying user needs with regard to digital mediation and describing user behaviour	38
4.2.3.	Good practices in similar institutions	40
4.3.	THE VIRTUAL RESEARCH ENVIRONMENT	41
4.3.1.	Providing multiple gateways to access collections	42
4.3.2.	Providing online tools to prepare on-site research	42
4.3.3.	Virtual research environments and digital research tools	43
<b>5.</b>	<b>DISSEMINATION AND VALORISATION</b>	<b>45</b>
<b>6.</b>	<b>PUBLICATIONS</b>	<b>46</b>
<b>7.</b>	<b>ACKNOWLEDGEMENTS</b>	<b>47</b>
<b>8.</b>	<b>ANNEXES</b>	<b>47</b>

## **ABSTRACT**

### **Context**

In 2015, the Belgian State Archives, the Centre for Historical Research and Documentation on War and Contemporary Society (CegeSoma) and the Royal Library have engaged in a partnership with the aim of carrying out a totally new project. The three institutions hold an important part of the Belgian federal heritage collections: more than 7 million documents in the Royal Library and nearly 300 kilometers of archives in the State Archives and CegeSoma. Over the last decades, the public visiting these institutions has changed considerably, both in its composition as in its practices. Apart from academic researchers and amateur historians, genealogists, notaries and government officials now visit the Archives. More and more students visit the Royal Library and often use the reading rooms as places for studying as well as a means to access the collections. CegeSoma finally has received more media attention, as well as an increased interest from editors and exhibition venues within the context of the numerous commemorations organized over the last years around the two world wars. In each of these cases, digital technology has made the collections of these institutions accessible to a larger and connected public, possibly located further away, and certainly more demanding. A public also that has, in this era of new technologies, undergone a profound change in its way of life, of thinking, producing, distributing, consuming, discussing, working or travelling. With the aim of defining a plan of action and strategy that will meet as well as possible the ongoing changes in the practices and expectations of their readers with regard to digital access to information, the State Archives, CegeSoma and the Royal Library have, for the first time, given the users the opportunity to express their views.

### **Objectives**

The survey organised by the State Archives, CegeSoma and the Royal Library was carried out within the framework of a research project financed by the Brain Programme of the Federal Science Policy between 2015 and 2017: the MADDLAIN project. This project was coordinated by CegeSoma, which since 1 January 2016 is the fourth operational direction of the State Archives. The MADDLAIN project has made it possible to gather a panel of experts and researchers with an expertise in IT, archive management, librarianship, data management, historical research and mediation tools. The goal of the study has not only been to provide data on the practices and needs of the users in terms of access to digital information, but also to move the institutions forward in the management of methods and tools that will help them to acquire a more in depth knowledge of their audiences. Apart from the general survey, two lines of research have been analysed: on the one hand the question of the mediation tools to facilitate the access to the collections and on the other the expectations of the university researchers who constitute an important target group of the three institutions. Two types of data have been used for the overall project: the navigation data of the users in the catalogues and websites of the institutions, and the non-structured or semi-structured data gathered via survey questionnaires and interviews. The users have collaborated extensively to our research. Nearly 2.300 of them have participated in our online survey in the spring of 2016. That the project has produced useful and relevant results is due in large part to their contribution.

## Conclusions

The survey carried out by the State Archives, CegeSoma and the Royal Library questions only the present public of the three institutions. This is a largely Belgian public, in the reading rooms as well as online. The majority among them consult the State Archives, CegeSoma and the Royal Library for personal reasons (studies, family research, administrative procedures, etc.), but their profile varies from one institution to another. At the Royal Library and CegeSoma, researchers (professional or amateur) and students constitute the largest group. They are between 18 and 55 years old and hold a university degree (or are still studying). Then there are the journalists, editors, teachers, or persons conducting research on their family history. At the State Archives, students and researchers come second to genealogists (of whom a large part are French) who generally have lower levels of education and who are mostly more than 55 years old.

« Access » and « Communication » are the two key words that emerge from the results of the survey carried out for the MADDLAIN project. The tools and means of communication used by the readers of the three institutions may have progressed considerably in the digital era, but the users have not basically adapted their needs in relation to the fundamental missions of the institution. Access to the collections is undoubtedly the first reason why users come to the State Archives, CegeSoma and the Royal Library. Activities such as workshops, publications, exhibitions or audiovisual productions only come second in the priorities of the persons interviewed.

In the survey, the question of « Access » has been considered for the paper collections as well as for digitized documents. In both cases, it would seem that the users combine several methods to access documents, by searching the digital catalogues as well as the paper inventories. The professional researchers specify that they apply several strategies in their search for information: key words, hierarchical tree, quotes, cascading research, information updating, contacts with experts, etc. The majority of the users profess satisfaction with the digital catalogues. The search engines are indeed used extensively: 20.000 single visitors each month for the State Archives, 10.000 for the Royal Library and 2.000 for CegeSoma.

Nevertheless, the results of the survey underline several improvements that could be made. The principal points of dissatisfaction concern the unclear ergonomics, the too complex general structure, the difficulties for inexperienced users, the inadequate or incomplete description of documents and the lack of communication with regard to the documents that are entered, or not, in the catalogue. The survey has revealed another important element concerning the access to the collections: the wish of the researchers to study the original source. Their contact with the reading room has nevertheless changed over the last years, and some of them refer to the changes in their working conditions (time-consuming administrative tasks, precarious contracts) that result in a need for an efficient management of their working time. Thus, they try to make their visit to the institutions as efficient as possible by making a detailed preparation based on the information available online but also by taking photographs of documents or entire archive funds that they will then process later.

The third lesson we learn from the survey concerns the problem of access to the documents: the majority of the users understand that not everything can be digitized and be available online. They do however ask explanations with regard to the underlying technical procedures, the choices made in terms of digitization and also to the potential restrictions

with regard to copyright or privacy protection. In general, the users feel that once the collections are digitized, they should be openly online available. Only a minority has called on the digital reproduction services and only few are prepared to pay for the use of them. However, a larger part of the users would be prepared to make a financial contribution in the context of a crowdfunding campaign if this would allow the digitization of a fund which has hitherto not been digitized. The expectations also concern specific types of documents: civil registers, notarial archives, the press and the photographs. A majority of users would also participate in crowdsourcing projects with the aim of improving the quality of the available metadata.

The question of communication is also a key question in the results of the survey. In general, the users ask for more transparency and more explanations with regard to the collections, their availability online or in the reading room, their level of description in the catalogues or paper inventories as well as the digitization projects. The users often claim to be poorly informed and the answers to the questions in the survey reveal that a number of complaints originate in reality in a manifest lack of information or in a lack of awareness of the available facilities. In spite of the tutorials and explanations available on the websites of the institutions, the navigation data have revealed that less than 5 % of the users actually go to these pages. The excessive fragmentation of the information, its complexity and the lack of well-adapted communication tools has a discouraging effect on the users.

Several suggestions have been made to address this problem, such as information sessions for specific groups, frequent email alerts with regard to recently entered data, clear and comprehensive mediation tools available on the website of the institution (videos, images, brief explanations), specialized guides to sources to help researchers or the possibility to contact the personnel of the institutions (chat, forum, video conference, etc.).

### **Keywords**

digital data, libraries, archives, e-learning, virtual research environment , web analytics



## **1. INTRODUCTION**

### **1.1. The Digital Revolution**

For several years now, the cultural institutions have definitely entered the digital era. From the outset, digital technology has created huge expectations. Technological progress has not only brought new opportunities for managing and valorizing the collections, but has forced the institutions to rethink their work processes, to gain new competences, to secure supplementary budgets and to reconsider some of their missions. Today, thousands of documents have been scanned and new technologies form an integral part of each stage of the production of knowledge: accessibility of sources, content analysis, conservation and valorization. Furthermore, the transition to digital meets a global mutation towards a more connected society, where access to knowledge becomes easier and where tools and sources are increasingly virtual. The huge changes that have come about through the digital era have been compared with those brought about by the invention of writing and printing.

### **1.2. The Need to Engage with the Audiences**

In 2015, the Belgian State Archives, the Centre for Historical Research and Documentation on War and Contemporary Society (CegeSoma) and the Royal Library have engaged in a partnership with the aim of carrying out a totally new project. The three institutions hold an important part of the Belgian federal heritage collections: more than 7 million documents in the Royal Library and nearly 300 kilometer of archives in the State Archives and CegeSoma. Over the last decades, the public visiting these institutions has changed considerably, both in its composition as in its practices. Apart from academic researchers and amateur historians, genealogists, notaries and government officials now visit the Archives. More and more students visit the Royal Library and often use the reading rooms as places for studying as well as a means to access the collections. CegeSoma finally has received more media attention, as well as an increased interest from editors and exhibition venues within the context of the numerous commemorations organized over the last years around the two world wars. In each of these cases, digital technology has made the collections of these institutions accessible to a larger and connected public, possibly located further away, and certainly more demanding. A public also that has, in this era of new technologies, undergone a profound change in its way of life, of thinking, producing, distributing, consuming, discussing, working or travelling. With the aim of defining a plan of action and strategy that will meet as well as possible the ongoing changes in the practices and expectations of their readers with regard to digital access to information, the State Archives, CegeSoma and the Royal Library have, for the first time, given the users the opportunity to express their views.

### **1.3. A Protean study**

The study organized by the State Archives, CegeSoma and the Royal Library was carried out within the framework of a two-years research project financed by the Brain Programme of the Federal Science Policy between 2015 and 2017: the MADDLAIN project. This project

was coordinated by CegeSoma, which since 1 January 2016 is the fourth operational direction of the State Archives. Two university partners were also associated : the department of Science and Technologies of Information and Communication at the ULB and the Ghent University through imec. The MADDLAIN project has made it possible to gather a panel of experts and researchers with an expertise in IT, archive management, librarianship, data management, historical research and mediation tools. The goal of the study has not only been to provide data on the practices and needs of the users in terms of access to digital information, but also to move the institutions forward in the management of methods and tools that will help them to acquire a more in depth knowledge of their audiences. Apart from the general study, two lines of research have been analysed: on the one hand the question of the mediation tools to facilitate the access to the collections and on the other the expectations of the university researchers who constitute an important target group of the three institutions. The popular and pseudoscientific discourse on big data and web analytics gives the impression that online tracking tools and shiny dashboards now make it extremely easy to gather objective facts concerning how people interact with online content and that trends emerge magically by themselves through the use of simple statistical methods. But one of the main lessons of the MADDLAIN project is the advantage of combining both approaches : quantitative and qualitative methods. This is why two types of data have been used for the overall project: the navigation data of the users in the catalogues and websites of the institutions, and the non-structured or semi-structured data gathered via survey questionnaires and interviews. The users have collaborated extensively to our research. Nearly 2.300 of them have participated in our online survey in the spring of 2016. That the project has produced useful and relevant results is due in large part to their contribution. The aim of the study was to blend all this information to get a perfect feedback cocktail giving to the institutions the opportunity to better meet the needs of their users.

## **2. STATE OF THE ART AND OBJECTIVES**

Humanities scholars make use of a wide variety of source materials and when cultural institutions started digitising their collections in the 1990s a plethora of different digital resources was created. Unfortunately, these resources do not form a humanities canon since they lack connectivity between them. This entails that potential users of the resources are unlikely to find them. Consequently, there exists a need to develop tools and resources that can facilitate the discovery of e-content as well as its integration into research and teaching practices (Ell and Hughes, 2013).

Many existing digital collections also lack features that would allow researchers to perform deeper quantitative and qualitative analysis of the available contents. Additionally, the presence of tools and services does not necessarily entail that they are usable or that users find them useful. It is therefore important to engage users in the design and development of digital humanities technologies so that their expectations and needs can be met. Since different types of users have varying degrees of knowledge with regard to the collections, they will also require varying levels of support(Steiner et al., 2014).

## **2.1. General Approach**

### **2.1.1. What are Digital Libraries and Archives?**

'Digital library' remains until present an ambiguous term. Matusiak (2012) emphasises that the concept is still evolving and therefore cannot be captured by a simple definition. Brahaj et al. (2013) stress the fact that digital libraries are driven by user groups or communities and that all of them focus on their specific usage scenarios. The definition proposed by Lynch and Garcia-Molina (1996) also underlines the role played by the users: a digital library is a "system providing a community of users with coherent access to a large, organised repository of information and knowledge" (p. 85). Matusiak (2012) singles out university students, teaching faculty, and researchers as the intended primary audiences of digital collections created by academic libraries.

However, rather than simply providing content to user communities, it would be beneficial to include in the digital library concept the activities of those communities (Bishop et al., 2003). Researchers, for example, regard digital libraries as networked information systems, while librarians see digital libraries as institutions or services. Both of these views regarding digital libraries will be assumed in this article.

### **2.1.2. Digitalization and Lack of Accessibility**

The definitions presented in the previous section unfortunately describe an ideal which is out of touch with reality. One reason why technologies and digitisation often go hand in hand with more complexity and less usability is that the relationships between processes and tools are becoming less visible (Bishop et al., 2003). Matusiak (2012) highlights that students predominately use online resources and search engines for their academic research instead of digital libraries and archives resources such as online catalogues or other services. According to the author, 89% of students use web search engines at the beginning of an information search — they believe that "search engines fit their information seeking behaviour better than physical or online libraries" (Matusiak, 2012, p. 136).

A study carried out by Fast and Campbell in 2004 leads to the same conclusion: libraries' OPACs — this acronym stand for Online Public Access Catalog, namely libraries catalogues reachable via the web — are more and more neglected in favour of new web services. According to the authors, users are discouraged by the lack of logic, intuitiveness and the coldness of these online catalogues. Again, these difficulties arise from the implementation of new technologies: "the migration to Web-based OPACs has posed new problems as old and sometimes problematic technologies are transferred to a new environment without rethinking the problem or taking full advantage of the medium" (Fastand Campbell, 2004, p. 139). These issues, which are related to the digitisation of libraries, brings us back to the notion of usability:

The technology has moved much more rapidly than has our understanding of the nature of the tasks for which we use it (...). We have had many calls for more “user friendly” systems, but we don’t understand human-computers compatibility well enough even to agree on what “user friendly” means. Thus we are left with several distinct challenges: 1) we need to determine what factors make computers difficult to learn and use; 2) we need to define a set of characteristics for friendly systems; and 3) we need to apply the research to design (Borgman, 1984, pp. 33–34).

### 2.1.3. Lessons Learned from Case Studies

By analysing data collected by observing the evolution of popular and scholarly internet services, Sadeh (2008) has identified four main expectations, and other studies’ outcomes point in the same direction.

Accessible Hayden et al. (2005) conducted a user survey at the Waterford Institute of Technology and found that over 51 per cent of students and 68 per cent of staff used library services remotely. According to Sadeh (2008), users expect digital libraries to be integrated into their usual environment. Libraries and archives therefore need to develop strategies and digital tools that can be made available in spaces where college students search and interact with information (course content management sites, search engines, Wikipedia, and social networking sites). Libraries and archives should increase their efforts in metadata harvesting and search engine optimization to expose their unique digital collections on the web. Search engine optimization and other technical tools are currently available to ensure that digital collections are indexed by Google and other search engines. Institutions need to address the usability of their websites and provide more integrated and seamless resource discovery tools that allow users to search across multiple online components (Matusiak, 2012).

Easy-to-learn Matusiak (2012) outlines that ease of use is an important expectation of digital libraries and archives users. A concrete example is that users expect the system to overcome incorrect spellings (Sadeh, 2008). However, usability studies of digital libraries need to extend beyond a single system and must consider ease of use in the context of other components of online information systems (Matusiak, 2012).

Rapid Many scientific institutions force the user to repeat a search query in multiple information resources (Sadeh, 2008). Boland et al. (2012) point out that publications are usually not stored together with the research data upon which they are based. The two types of resources are normally stored in different information systems that have no links between them. This makes the search for previous research more challenging and time-consuming.

Gratifying One of the most important user needs pinpointed by Perley et al. (2007) is the provision of “just in time” information services; the authors go on to say that librarians could design and deliver sophisticated information services, even aimed at specific audiences of an organisation. According to Sadeh (2008), most of the internet services reassure users by

almost always presenting items after a request; library systems do not yet offer a similar experience.

## **2.2. Specific Audiences**

### **2.2.1. E-learning**

Sangrà et al. (2012) identified the need for an inclusive definition of e-learning and enlisted the help of international e-learning experts in order to create a valuable conceptual framework on the subject. 31 of the 33 participants evaluated the following definition favorably:

E-learning is an approach to teaching and learning, representing all or part of the educational model applied, that is based on the use of electronic media and devices as tools for improving access to training, communication and interaction and that facilitates the adoption of new ways of understanding and developing learning. (p. 152)

E-learning courses are free of space and time constraints, which entails they offer a high level of flexibility to their users. Learners can consult course contents and resources at a location and time of day that is most convenient for them and they can study at their own pace (Sun et al., 2008). Students can also select material that corresponds to their own knowledge level and interests to obtain the information they require to effectively perform a task or activity. Full engagement of students in the learning process is therefore essential for successful e-learning (Kigundu, 2014).

According to Jurado and Pettersson (2011), e-learning tools can be subdivided into four categories: distribution, communication, interaction and administration. Their research indicates that lecturers make far more use of tools for distribution than of those that facilitate communication or interaction. Many tools offered by e-learning systems also go unused by lecturers. Jurado and Petterson concluded that, in general, lecturers choose to only make use of tools that facilitate their teaching process and largely neglect to use tools that can enhance the students' learning experience.

Libraries can play a role in e-learning by implementing an e-learning strategy that allows staff to locate sources and library services that can be integrated into the courses they teach. Digital Libraries can be employed to offer networked resources in different formats and they can also include a range of services. Helping students acquire information skills can be seen as an integral part of successful e-learning. Ideally, information literacy would be integrated into the university curriculum since students tend to perform better when they are taught information skills while they are searching information for an assignment or project (Nfila, 2009).

The objectives of the part of the study devoted to digital mediation and online user education are, first and foremost, to establish current practices in the participating FSI. The next objective is twofold: to identify user needs (explicit demand) and to describe user behaviour (implicit). A third objective consists of exploring good practices and their components in similar institutions. Lastly, guidelines concerning digital mediation and online user education in heritage institutions will be formulated.

### 2.2.2. Virtual Research Environments

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, heritage institutions undertook major digitisation initiatives which made available a large quantity of sources in a digital format. Based on this material, new tools and methods developed in the wake of the Digital Humanities allow researchers to experiment with computationally-based techniques, such as quantitative analysis, text mining, named entity recognition. These provoke new research questions, and thereby promote scientific innovation.

To address these changes, national and international initiatives were launched to reflect on the development of online research infrastructures, such as the European Strategy Forum on Research Infrastructure (ESFRI) or programmes lead by the Joint Information Systems Committee (JISC) in the UK. For the Humanities, the building of consortiums and networks around the development of large-scale research infrastructures aims to bring greater standardization, interoperability, and mutualization of skills and resources in a field where digital projects are often compartmentalized. The European Commission launched several such projects as part of ESFRI, which are federated around the network DARIAH (Digital Research Infrastructure for the Arts and the Humanities).

As components of the online research infrastructure, virtual research environments aim to provide a team of researchers with access to a dataset and a number of tools to process it, thus promoting collaborative and innovative research. Ideally, VREs would be user-driven, i.e. flexible and customizable in order to be adapted to the research objectives, and interoperable (Buddenbohm et al., 2015; Candela et al., 2013; Carusi and Reimer, 2010; Voss and Procter, 2009). However, the development of VREs faces some challenges, the most important being a slow uptake by researchers, either because the environment is not user-friendly, or because it does not fit their needs. From this point of view, it is important to adopt a bottom-up, rather than top-down, approach, engaging users in every step of the development process. Other issues regarding sustainability, rights and copyright limitations also come to the fore.

Another question is whether the concept of virtual research environments, and more widely, large-scale online research infrastructures, is adapted to the specificities of Humanities research. Since the Humanities are based on a diversity of documents, contexts, and

research methods, all-encompassing research infrastructures might miss the point of most actual needs. Furthermore, the Humanities still have a strong conservative culture, where there is some resistance against new technology and where collaboration is not the norm. Finally, there is some skepticism as to the relevance of “big data” and as to the innovative character of research questions that can be applied to heritage material (ACLS, 2006; Anderson and Blanke, 2012; van Zundert, 2012).

Should those challenges be overcome, research infrastructures 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 such infrastructures not only as provider of content, but also through their valuable knowledge and experience in information sciences and data curation (ACLS, 2006; Speck and Links, 2013).

As researchers in the Humanities form a significant part of the audience visiting archives and libraries, the MADDLAIN project included a section to reflect on the opportunity for these institutions to develop new infrastructures to support scientific research. In concrete terms, this section revolves around two main questions:

1. 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?
2. 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?

Through a thorough reflection on this user group’s research practices, this study contributes to define possible avenues towards the future of scientific federal institutions’ digital infrastructures.

### **2.3. Assessing General and Specific User Needs**

According to Isfandyari-Moghaddam (2011), several factors can affect the evaluation of digital collections, such as a lack of resources or difficulties to make use of holistic approaches. This is due to the multiple facets of the concept: digital libraries and archives are composed of a wide range of phenomena embedded in a variety of external contexts. Nonetheless, the authors outline that, even though this task needs a broader amount of effort during the preparation phase, a well-planned assessment strategy can return more useful information for the same volume of resources. Evaluation methods are usually grouped in two principal categories: quantitative and qualitative approaches, the following two sections will explore both of them.

### 2.3.1. Quantitative Approaches

According to Matthews (2007), “quantitative research is usually used to estimate or predict a future outcome or to diagnose the existing or current state of a subject” (p. 61). Quantitative data can be collected with a variety of techniques, including statistical infrastructure, surveys, transaction log analysis, experiments etc. The main goal of this approach is to quantify users’ behaviours, expectations and other defined variables. Because the methods are more structured than qualitative tools, the outcomes of these approaches are easier to generalize.

Surveys are the best known quantitative approach to collect data (Matthews, 2007). They form part of academic library life and several standard instruments are available to librarians, such as LibQual+, SCONUL, the commercial product Libra and Rodski. Although user surveys aim at improving the services provided and high satisfaction rates are good publicity, it is critical to go beyond these indicators and to deepen all the facets of the outcomes with a further qualitative analysis (Creaser, 2006). One of the difficulties with surveys is the audience segmentation: the public is composed of students, researchers and other categories such as journalists, genealogists, enthusiasts and many other groups that can be clustered in a broad “general audience” category. Hayden et al. (2005) sketch a methodology based on two different groups: undergraduate students and academic staff, and outlines the benefits and weaknesses of several survey types such as e-mail, online and hard copy surveys. The authors also provide a sampling method based on these survey types.

Web analytics methods are also increasingly used in cultural contexts. This consists of statistical numbers that enable a better understanding of user’s behaviour.

Four concepts have offered an increasingly complex manner of interpreting web analytics: hits, page views, visits and outcomes (S. van Hooland et al., 2016). In the 1990s, the only data that an institution could acquire through a web traffic tool was the number of hits for that specific day. A hit takes place each time a request is sent to a web server and a file is downloaded from that server. For example, viewing a homepage which is generated based on an HTML file and which contains two images, counts as three hits. Some years later, the appearance of the page views metric was one step closer to precision. It was possible for an institution to know that on a specific day a specific user visited the website, 121 web pages were viewed. However, the institution would have had no way of determining if the page views were the result of a single enthusiastic user or rather forty visitors who consulted a small number of pages individually. This gap was filled by the visits metric, that allows an institution to determine how many visits were responsible for the 121 page views. Additionally, this metric made it possible to count each person visiting a website individually, leading to a more subtle way of measuring success. The last stage has been reached with the development of more precise metrics, known as ‘outcomes’. These metrics emerged due to the growth of e-commerce and the rise of online marketing strategies (Kaushik, 2009). Institutions are now able to conclude from the data that a user consulted the opening hours



page and also subscribed to the newsletter. These new metrics constitute the core of the web analytics research area, but before getting to the heart of the matter, it should be noted that they raise serious questions in terms of technology and privacy.

Technically speaking, two ways exist to monitor the elements discussed above: web logs and page tagging. Over the years, log files have been increasingly dropped in favour of the page tagging method. Each method has, however, its advantages and disadvantages. On one hand, log files contain raw data recorded automatically by a web server and enclose information concerning when, how and from which IP address requests are sent to the server. An important aspect of this method is that the website's owner can maintain control over the data. The page tagging method, on the other hand, requires a third-party service such as Google Analytics. The principle consists of adding a piece of JavaScript code to the pages that should be tracked and placing a cookie on the computers of visitors the first time they visit the website so that they are recognised as returning visitors if they access the website again in the future. Every time a page that contains the JavaScript code is opened, the code in question is executed and data is sent to a remote server. Whenever a user accesses the website, the web analytics tool creates a session ID and links it to the ID associated with the user so that it becomes possible to determine how many times a specific user visits the website (Voorbij, 2010).

Privacy issues emerge as soon as one wishes to store large data sets that concern users' behaviour. The identification of users, which is limited to an ID number, is based on either their IP address or cookies stored on their computer. IP addresses are useful except in two situations: when people use a collective access through a proxy server or when they operate in an environment that uses dynamic IP addresses (van Hooland, 2016; Voorbij, 2010).

Cookies offer a way to overcome these two limitations, since they are small text files that are stored on a user's hard disk. Furthermore, they make it possible to identify a connexion between a browser and a server in order to visualise web pages. This aims to personalise the user's experience by, for example, remembering the user's personal preferences. Collecting user profiles is also increasingly being used for behavioural targeting and other marketing purposes, which raises questions concerning people's privacy. Thus, to comply with Belgian legislation, some precautions should be taken. The Belgian law on cookies has recently changed (The Telecommunications Law of 12 July 2012 transposed the requirements of European Directives 2009/136/EC and 2009/140 EC into Belgian law) and now website owners must inform users that they may make use of cookies and explain why they use them. They are also required to ask a user's permission to store a cookie on the device used to visit the website. While a small amount of information is stored in cookies, the raw data is predominantly stored on the server hosting the analytics software, and are usually owned by a third party. This entails that the institution does not fully control the collected data. However, alternatives exist such as Piwik, which is a free open source software that can be installed on an institution's own server. This implies that the institution owns the raw collected data and has permanent access to it.

If the analysis of online audiences and related concepts have spread rapidly in the commercial sector (Peterson, 2004 ; Kaushik, 2009), the development of strategies and methodologies specific to the cultural sector was slower. It is not until 2013 that the Library and Information Technology Association (LITA, a division of American Libraries Association) has published a guide on this subject (Farney and McHale, 2013).

In 2003, a review of literature identified 200 research papers, published between 1995 and 2003, that focus on the use of electronic library resources (Tenopir 2003). Findings relate mainly to the comparison between print and electronic resources usage and provide a broad overview of the adoption of these new type of resources. While some studies were based on interviews or ethnographic approaches, others include transaction log analysis and first experiments in terms of web mining.

In sum, research papers can be classified into two categories: those making use of transaction/web log files and those making use of data collected by script (e.g. via Google Analytics). Among the first authors who published on this topic, Ghaphery (2005) falls into the first category: in order to improve an academic library web site, he used transaction log files and web server logs to detect patterns in clicks on 'quick links'; while Fang (2007) falls into the second category: he used data collected via Google Analytics to improve library website design but also to establish a strategy to create new contents.

In a European context and a more transversal approach, Voorbij (2010) investigated the use of web statistics in cultural heritage institutions in The Netherlands, including libraries archives, and museums. His principal findings underline the fact that most institutions mention web statistics in their annual report, but do not explain them or lack background information, which comparisons between institutions difficult.

Notwithstanding the vulnerability of incomplete explanations in reports, cultural heritage institutions were recently inspired by the commercial sector and started focusing more on outcomes rather than isolated metrics. Fagan (2014) reviewed key performance indicators, traditionally used in the commercial sector, and evaluated their validity in the context of academic libraries. These Web metrics include conversions rates, indicating the percentage of the users who performed specific actions ("goals"). These specific actions were typically financial transactions. If cultural heritage institutions want to successfully make use of these "goals", they need to adapt this concept to the non-profit sector. Fagan enumerated some examples, as well as Prom (2011), who described in detail the iterative development of a methodology to implement these goals and analyse the resulting data.

### 2.3.2. Qualitative Approaches

Matthews (2007) explains that "Qualitative research methods are particularly helpful when attempting to better understand complex relationships among and between variables" (p. 47). Qualitative tools are based on smaller samples and it is more difficult to generalise the outcomes of a study that makes use of them. These methods are used to define the "why"

whereas quantitative approaches rather help to identify the “what” (Matthews,2007). The adoption of multiple instruments when collecting data during a qualitative study can enhance the credibility of the analysis by providing complementary perspectives (Perley et al., 2007).

This observation was made in the context of a medical library’s user needs assessment project in Kansas. During this study, the three most used instruments for data collection were adopted: namely self-reporting surveys, telephone interviews, and focus groups. Other interesting qualitative information can be collected indirectly by studying surveys results from other projects and by observing the evolutions of both scholarly and “general public” Web services. By doing so, Sadeh (2008) has collected a large amount of data on users’ expectations.

## 2.4. References

- *Bibliothèque Nationale de France. Enquête Auprès Des Usagers de La Bibliothèque Numérique Gallica*, BNF, avril 2017.
- ‘Business Planning for Digital Libraries: International Approaches’ in *Online Information Review*, 35, no. 2 (2011): 316–17.
- ‘Collection Development in the Digital Age’ in *Program: Electronic Library and Information Systems*, 47, no. 4 (2013): 450–52.
- ‘Introduction to Digital Heritage’ in *Museum Management and Curatorship*, 23, no. 4 (2008): 307–8.
- ACLS, *Our Cultural Commonwealth. The Report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*, 2006.
- Amar, Muriel, Véronique Delannay, and Véronique Mesguich., *Le Web 2.0 en bibliothèques : quels services ? quels usages ?*, Paris: Electre-Ed. du Cercle de la Librairie, 2009.
- Anderson, S. and T. Blanke (2012) ‘Taking the Long View: From E-Science Humanities to Humanities Digital Ecosystems’ in *Historical Social Research/Historische Sozialforschung*, 37(3): 147–64.
- Anderson, Sheila. ‘What Are Research Infrastructures?’ in *International Journal of Humanities and Arts Computing*, 7, no. 1–2 (2013): 4–23.
- *La bibliothèque à l’heure du web 2.0.*, Paris: Archimag-Guide Pratique, 2008.
- Bats, Raphaëlle, *Construire Des Pratiques Participatives Dans Les Bibliothèques*, Presses de l’ENSSIB, n.d.
- Berende, Hans, Karin van der Heiden, Theo Thomassen, Charles Jeurgens, Christian van der Ven, and Harriet de Man, *Schetsboek: Digitale Onderzoekomgeving En Dienstverlening. Van Vraag Naar Experiment*, ’s-Gravenhage, 2016.
- *Bibliothèques : Les Nouveaux Modèles*, Paris: Archimag-Guide Pratique, n.d.
- Bishop, Ann P., Nancy A. Van House, and Barbara Pfeil Buttenfield, *Digital Library Use: Social Practice in Design and Evaluation*, Whole. Cambridge, Mass: MIT Press, 2003.

- Blanke, Tobias, Leonardo Candela, Mark Hedges, Mike Priddy, and Fabio Simeoni. 'Deploying General-Purpose Virtual Research Environments for Humanities Research' in *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 368, no. 1925 (2010).
- Bonnamour, Pascale, and Ludivine Garnaud, *Monétiser l'information sur le Web: les modèles économiques gagnants*, Paris: CFPJ, 2011.
- Buddenbohm, S. et al., 'Success Criteria for the Development and Sustainable Operation of Virtual Research Environments' in *D-Lib Magazine* 21(9), 2009.
- Calhoun, Karen, *Exploring Digital Libraries: Foundations, Practice, Prospects*, Chicago: Neal-Schuman, 2014.
- Candela, L. et al., 'Virtual Research Environments: An Overview and a Research Agenda' in *Data Science Journal*, 12, 2013.
- Caraco, Benjamin. 'Les Digital Humanities et Les Bibliothèques: Un Partenariat Naturel' in *Bulletin Des Bibliothèques de France*, 57, no. 2 (2012): 69–73.
- Carusi, A. and T. Reimer, *Virtual Research Environment. Collaborative Landscape Study. A JISC funded project*, 2010.
- Chambers, Sally, *Catalogue 2.0: The Future of the Library Catalogue*, Chicago: ALA Neal Schuman, 2013.
- Chaumier, Jacques, *Document et Numérisation: Enjeux Techniques, Économiques, Culturels et Sociaux*, Paris: ADBS, 2006.
- Covey, Denise Troll, *Usage and Usability Assessment: Library Practices and Concerns*, Washington: Digital Library Federation, Council on Library and Information Resources, 2002.
- Creaser, Claire. 'User Surveys in academic libraries' in *New Review of Academic Librarianship*, 12, no. 1 (2006): 1–15.
- Dadzie, Perpetua S. 'E-Learning and E-Library Services at the University of Ghana: Prospects and Challenges' in *Information Development*, 25, no. 3 (2009): 207–17.
- Devauchelle, Bruno, *Comment Le Numérique Transforme Les Lieux de Savoir: Le Numérique Au Service Du Bien Commun et de L'accès Au Savoir Pour Tous*, Limoges: Fyp éditions, 2012.
- Dufrêne, Bernadette, Madjid Ihadjadene, Denis Bruckmann, Benjamin Barbier, and Bruno Racine, *Numérisation Du Patrimoine: Quelles Médiations? Quels Accès? Quelles Cultures?*, Paris: Hermann, 2013.
- Ell, Paul S., and Lorna M. Hughes, 'E-Infrastructure in the Humanities' in *International Journal of Humanities and Arts Computing*, 7, no. 1–2 (2013): 24–40.
- Fagan, J., 'The Suitability of Web Analytics Key Performance Indicators in the Academic Library Environment' in *Journal of Academic Librarianship*, 40(1) (2014) :25–34.
- Farney, T., and N. McHale, *Web Analytics Strategies for Information Professionals*, Ala TechSource, 2013.
- Fast, Karl V., and D. Grant Campbell. "I Still like Google': University Student Perceptions of Searching OPACs and the Web' in *Proceedings of the American Society for Information Science and Technology*, 41, no. 1 (2004): 138–46.

- Flyvbjerg, Bent. 'Five Misunderstandings About Case-Study Research' in *Qualitative Inquiry* 12, no. 2 (2013 2006): 219–45.
- Bibliothèque nationale de France, *BnF - Numériser Le Patrimoine Pour Le Centenaire de La Grande Guerre*, Accessed March 5, 2015.
- Ghaphery, J., 'Too quick? Log analysis of Quick Links from an academic library website' in *OCLC Systems & Services: International digital library perspectives*, 21(3) (2005):148–155.
- Given, Lisa M., and Lianne McTavish, 'What's Old Is New Again: The Reconvergence of Libraries, Archives, and Museums in the Digital Age' in *Library Quarterly*, 80, no. 1 (2010): 7.
- Guigueno, Brigitte, *Qui Sont Les Publics Des Archives ? Enquêtes Sur Les Lecteurs, Les Internauts et Le Public Des Activités Culturelles Dans Les Services Publics D'archives ( 2013-2014 )*, Service interministériel des Archives de France, 2015.
- Hayden, Helen, Terry O'Brien, and Maoilíosa Ó Rathaille. 'User Survey at Waterford Institute of Technology Libraries: How a Traditional Approach to Surveys Can Inform Library Service Delivery' in *New Library World*, 106, no. 1/2 (2005): 43–57.
- Isfandyari-Moghaddam, Alireza. 'Evaluation of Digital Libraries: An Insight into Useful Applications and Methods' in *The Electronic Library*, 29, no. 4 (2011): 555–56.
- Kaushik, A., *Web Analytics 2.0: The Art of Online Accountability and Science of Customer Centricity*, Alameda, 2009.
- Lynch, Clifford, and Don E. Carleton, 'Lecture: Impact of Digital Scholarship on Research Libraries' in *Journal of Library Administration*, 52, no. 6–7 (2012): 456–73.
- Lynch, Clifford, 'Digital Library Opportunities' in *The Journal of Academic Librarianship*, 29, no. 5 (2003): 286–89.
- Maisonneuve, Marc, *Le catalogue de la bibliothèque à l'heure du Web 2.0: étude des opacs de nouvelles génération*, Paris: ADBS Editions, 2008.
- Manovich, Lev., *Le Langage Des Nouveaux Médias*, Dijon: Les presses du réel, 2010.
- Matthews, Jacob T., and Philippe Bouquillion, *Le Web collaboratif: mutations des industries de la culture et de la communication*, Saint-Martin-d'Hères (Isère): PUG, 2010.
- Matthews, Joseph R., *The Evaluation and Measurement of Library Services*, Westport, Conn: Libraries Unlimited, 2007.
- Muet, Pierre-Alain, 'Impacts Économiques de La Révolution Numérique' in *Revue Économique*, 57, no. 3 (2006): 347–75.
- Palmer, C.L. et al., *Scholarly Information Practices in the Online Environment. Themes from the Literature and Implications for Library Service Development (Report commissioned by OCLC Research)*, 2009.
- Perley, Cathy M., Camillia A. Gentry, A. Sue Fleming, and Kristin M. Sen, 'Conducting a User-Centered Information Needs Assessment: The Via Christi Libraries' Experience' in *Journal of the Medical Library Association : JMLA* 95, no. 2 (2007): 173–e55.
- Peterson, E., *Web analytics demystified*, CafePress, 2004.
- Prat, Marie, *Réussir votre projet e-learning: méthodologie et outils web 2.0.*, Nantes: ENI, 2014.

- Prom, C., 'Using Web Analytics to Improve Online Access to Archival Resources' in *The American Archivist*, 74(1) (2011): 158–184.
- Rebillard, Franck, *Le Web 2.0 en perspective: une analyse socio-économique de l'Internet*, Paris: L'Harmattan, 2007.
- Speck, R. and P. Links, 'The Missing Voice: Archivists and Infrastructures for Humanities Research' in *International Journal of Humanities and Arts Computing* 7(1–2) (2013): 128–46.
- Tenopir, C., 'Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies' in *Council on Library and Information Resources*, August, 72, 2003.
- Tsakonas, Giannis, and Christos Papatheodorou, 'Exploring Usefulness and Usability in the Evaluation of Open Access Digital Libraries' in *Information Processing and Management*, 44, no. 3 (2008): 1234–50.
- Tsakonas, Giannis, and Christos, Ph. D Papatheodorou, *Evaluation of Digital Libraries: An Insight to Useful Applications and Methods*, Oxford: Chandos, 2009.
- Unsworth, J., 'Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?' in *Symposium on Humanities Computing: Formal Methods, Experimental Practice*, 2000.
- van Hooland, S., M. De Wilde, S. Hengchen and F. Gillet, *Introduction aux humanités numériques : méthodes et pratiques*, Bruxelles : De Boeck, 2016.
- van Zundert, J., 'If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities' in *Historical Social Research/Historische Sozialforschung*, (2012)
- Voorbij, H., 'The use of web statistics in cultural heritage institutions' in *Performance Measurement and Metrics*, 11(3) (2010): 268–269.
- Voss, A. and R. Procter, 'Virtual Research Environments in Scholarly Work and Communications' in *Library Hi Tech* 27(2) (2009): 174–90.
- Walt, Thomas Van der, and Perpetua S. Dadzie, 'Digitising University Libraries in Ghana: How Technology Is Facilitating to Digital Content and Service Content and Services' in *Mousaion*, 33, no. 3 (2015): 95–114.
- Warren, Jacques, Nicolas Malo, and Marc Lolivier, *Web analytics: mesurer le succès et maximiser les profits de votre site Web*, Paris: Ed. d'Organisation, 2009.
- Wusteman, Judith, 'Editorial. Virtual Research Environments: Issues and Opportunities for Librarians' in *Library Hi Tech*, 27, no. 2 (2009): 169–73.
- Xie, Iris, and Krystyna K. Matusiak, *Discover Digital Libraries: Theory and Practice*, Elsevier, 2016.

### 3. METHODOLOGY

#### 3.1. Overall Methodology

The MADDLAIN project was created with the aim of analysing the behaviour and the needs of different audiences with regard to the digital access to the collections offered by the aforementioned institutions. The research corpus on which this report is based was gathered between July 2015 and June 2016. It contains quantitative as well as qualitative data.

The qualitative side of the project (interviews with the staff members of the institutions and a user survey) was conducted by one researcher at the CegeSoma for the part about the general public and by two other researchers at the State Archives and the Royal Library for the issues about virtual research environment and the mediation tools. This qualitative side spread over the entire duration of the project. The interview phase took around eight months to complete (1 month for the preparation stage, 1 month to conduct the interviews, 3 months to create transcriptions and 3 months to perform the analysis) and the user survey phase around five months (2 months for the preparation stage, 1 month for the post-processing of the data and 2 months for the analysis). Two specific surveys were also realised in the second year of the project on a smaller sample of people to deepen the issue of virtual research environments as well as the one of mediation tools.

For the quantitative side, the corpus consisted of web logs acquired by making use of Piwik during the whole period. Goals were added in Piwik from January 2016 onwards. The researches of imec were responsible for the integration in the websites analysed by the project of Piwik, a free open source software that allows to extract navigation data<sup>1</sup>. The researcher associated with the Université libre de Bruxelles made use of web analytics to analyse the collected data. The web analytics phase of the project was completed in eight months (1,5 months to develop the global methodology and to define the goals, 2,5 months to set the goals, 1 month to collect data and to perform quality control and 3 months for the analysis).

The main research question can be divided into four sub questions, which can themselves be subdivided into many smaller ones:

- a) **Analysis of the existing services:** More precisely, we are targeting here the services related to the access to library collections and archival holdings and their valorisation. We will try to determine what is currently on offer by the three institutions and if there is a part of the content that is not available — e.g. because it is uncatalogued or for confidentiality reasons — or accessible only by paying a fee.
- b) **Current user behaviours:** The goal here is to identify which content users need and if this content is provided in an adequate way by the institutions and found by the

---

<sup>1</sup> More information can be found on <http://piwik.org> (accessed 28 August 2016).

users. Another objective will be to determine the characteristics of the most popular content in order, for example, to suggest ranking systems in which these characteristics are used to “push” particular content to the top of the search results. This information could also be used to improve the existing interfaces in order to reduce the number of clicks needed to acquire the desired document, and more generally, bring these interfaces more in line with the user behaviours.

- c) **User expectations:** This sub question is concerned with identifying which kind of features audiences would like to have at their disposal and how they expect systems to function.
- d) **Potential new services:** Once the existing services and the user behaviours and expectations have been identified, we will reflect upon solutions to adapt the services to the user behaviours. They will also propose new features that would help the institutions to fully meet the expectations of their users. It will be an opportunity to create and foster a spirit of interactivity, in which new services would be provided, allowing the valorisation of both library collections and archival holdings, through the use of e-learning structures and scientific collaboration tools such as Virtual Research Environment.

### 3.2. Face-To-Face Interviews.

Because of the complex ways in which content is made accessible, we decided to begin by conducting an unstructured or semi-structured interview, as defined by Myers and Newman (2007) and whose methodology will be explained in this subsection. These interviews will serve several purposes:

- identify **new interviewees** in an iterative process ;
- **collect information** about the research question: the ways in which content is made available, but also user behaviours and expectations;
- identify **new information sources** of access related data — e.g. email requests to access content, reading room documentation, etc.;

**Planning** of the interviews consisted of several steps. (1) At first, we had to determine the role and responsibilities of the stakeholders. Two researchers performed the interviews, and the repartition of the interviewees was based on their mother tongue (Dutch or French). The protocol and the interview guides were created by the researchers in collaboration with the coordinator of the research project. The other promoters of the project were also consulted. (2) The second step consisted in identifying the interviewees: they were selected from the staff of the three federal institutions. The selected group included library collection managers, archival holdings managers, educational services members, communication services members and also digitisation services members. (3) Finally, we developed tools: namely an interview protocol — which indicates how to commence and conclude the interview and which behaviour the interviewers should adopt — and an interview guide — which contains all the questions, but also definitions of crucial concepts. The interview guide



contains about 15 questions, as recommended by Boyce and Neale (2006). Both Boyce and Neale (2006) and Turner III (2010) emphasise that it is best to test both the guide and the protocol before starting the actual interviews. Because of the complexity of the research questions and the context of the three institutions in a bilingual environment — we decided to not only perform a test interview in both Dutch and French but to also arrange **introductory meetings** with all the interviewees in order to present the MADDLAIN project to them. We also provided them with instructions on how to prepare for the interview and offer them the opportunity to ask questions and request clarifications.

### 3.3. The web analytics

The adopted methodology for the whole ‘Web Analytics’ process consists of four main steps: the literature review, the data collection, the creation of goals, and the data analysis. First, an extensive literature review on Web Analytics was conducted in order to choose an appropriated tool to capture data and to define Web metrics (Key Performance Indicators). Publications and projects carried out in other institutions were reviewed, leading to the selection of a “hybrid” tool: Piwik. This free open source software brings together the ease of use of a graphical user interface (with dashboards similar to those of Google Analytics) and the completeness of searchable raw data (permanently owned by the institution). Reviewing the literature also helped to select Web metrics adapted to cultural institutions websites, and, more broadly, suitable for the Maddlain project. They include isolated metrics such as the number of pages viewed per visit as well as more innovative metrics based on “outcomes”. These new metrics require to define goals and calculate conversion rates: these concepts will be detailed in the third step of the methodology.

Secondly, the data collection was implemented and documented during the first year of the project (summer 2015). Concretely, two researchers were responsible for the integration of Piwik in the 12 websites analysed by the project (official websites, online catalogues and virtual exhibitions from the three involved institutions: the Royal Library of Belgium, the State Archives of Belgium and the Centre for Historical Research and Documentation on War and Society). The three main challenges encountered during this step were the management of ambiguous URLs, the recovery of raw data and the presence of performances issues.

Thirdly, advanced parameters have been set to work with more elaborate metrics. Concretely, a dozen “goals” were added in Piwik from January 2016 onwards. For each of these goals, a “conversion rate” has been registered, i.e., the percentage of visits during which the goal was accomplished (for example: “accessing digitised content”). These conversion rates can be analysed to gain in-depth insight. In the context of the Maddlain project, a pragmatic method to implement goals in a cultural heritage context has been developed. The method consists of six major steps: Defining Goals; Identifying Relevant Data; Creating Filters; Setting Goals; Conducting Quality Control and Performing Segmentation (details can be found in Hungenaert and Chardonnens, 2017). It should be noted that while goals share common definitions (for example: “click to access tutorial or online help”), they have to be adapted to the specific URLs of each website.

Lastly, the collected data were examined in two ways: firstly by using the Piwik Graphic User Interface (GUI), secondly by exploiting the Piwik raw data. The first part of the analysis aimed at isolating quantitative data related the Web metrics chosen during the literature review. Some numbers about online visits have been accessed via the basic functionalities offered by the Piwik GUI, while other percentages, such as conversion rates, have required customizable goals mentioned in the previous step. All these statistical data have been compiled in Excel files. The second part of the analysis was conducted to overcome the limits imposed by the GUI (which only displays statistics based on aggregated data). The raw data collected by Piwik were exploited by means of tailor-made solutions, which require more time and are technically more demanding. The tools used to extract and process the raw data include MySQL, Pandas (a Python library) and OpenRefine. This in-depth analysis mainly focuses on online catalogues: user preferences regarding the search engine advanced facets were extracted as well as user queries (details can be found in Chardonnens and Hengchen, 2017).

The whole corpus covers a period of six months (From 1 January 2016 to 1 July 2016), with an exception for the conversion rates, which vary according to the websites (due to the time required to implement the goals and to perform the quality control), but cover at least a period of 3 months (distributed from June 2016 to November 2016).

### **3.4. E-learning**

#### **3.4.1. Current practices in the participating FSI**

The current state of digital mediation and user education at the AGR, CEGESOMA and KBR was mapped through content analysis of the websites of the three institutions. Content, related to digital mediation of the collections and search tools, was added to an Excel sheet entitled 'Mediation\_(FSI)' describing the following properties:

- URL
- Content
- Content type
- Instruction / Media
- # Clicks
- Goal (PIWIK)
- Description

The properties describe: the location of the content (URL), the subject of the digital mediation or user education practice (such as research (topics), parts of the collection, search engines, ...), the type of content (tutorial, descriptive, FAQ, database or catalogue), the structure or form of the content (PDF, plain text, images, lists ...), the lowest number of clicks required to get from the homepage to the concerned webpage (to establish the 'depth' of the concerned information). Lastly, when available: the corresponding PIWIK goal in order to easily retrieve web statistics if necessary.

Additionally, an online questionnaire amongst the personnel who are in contact with the audience served as a means to gain insight into the current accompaniment for the use of the main digital catalogues in the participating FSI. The questionnaire sought out to shed a light on the following topics:

- Perceived ease of use of the digital catalogues;
- Frequently asked questions concerning the use of the digital catalogues;
- Services, concerning the use of the digital catalogues, that are currently offered.

#### 3.4.2. Identifying user needs with regard to digital mediation and describing user behaviour

The responses of users of the AGR, CEGESOMA and KBR resulting from the online survey have been analysed in order to identify user needs with regard to digital mediation. Using the raw data from the user survey, a qualitative analysis was done, following these steps:

- Selection of those particular questions that shed a light on (a) the necessity for improved online user education and digital mediation, (b) current user behaviour that indicates how users navigate the website and search information, (c) ways to improve online user education and digital mediation in the concerned FSI
- A close reading of the responses, resulting in a brief analysis per question.

#### 3.4.3. Good practices in similar institutions

Current good practices with regards to digital mediation and online user education at institutions similar to the FSI were identified in order to gain a somewhat representative overview of key components, digital services and their functionalities.

- Selection of a sample of 7 similar institutions per participating FSI (archives, libraries and centers for research/archives on war and/or aspects of contemporary society) for a total of 21 similar institutions.
- Identification of digital services that are offered by said institutions through analysis of their website.
- Visual presentation of the digital services in a PowerPoint document.
- Adding the digital services into an Excel sheet and identifying the following components:
  - Identification
    - Institution;
    - URL: Homepage;
    - URL: Catalogue;
    - Catalogue name.
  - Key components
    - Definition: archive (AI);
    - Glossary, terms (AI);
    - Definition: finding aid (AI);
    - How to: finding aid (AI);
    - How to: catalogue (AI);

- Rules (AI);
- What's accessible (AI);
- What's digitally accessible;
- Central research page;
- First visit aid (AI);
- Video tutorials (AI);
- IM, chat (AI);
- FAQs;
- Personal account.

The abbreviation '(AI)' – Archival Intelligence – was added in a second phase. These “markers” originate from Benjamin Bromley's 2010 thesis on user education resources and archival concepts. Archival Intelligence is a concept first described by researchers Yakel and Torres in a 2003 paper.

- Questionnaire sent to the total of 21 institutions. The questionnaire consisted of open-ended questions regarding digital mediation in order to research the following aspects:
  - Motivation
  - Target audience
  - Processes, workflow
  - Outcome, value
  - Unforeseen consequences or challenges

The responses are added to the project documents as raw data, seeing as there wasn't enough time left for a thorough analysis.

### **3.5. Virtual research environments**

The adopted methodology consists in two main approaches: state of the art and user requirements analysis.

First, 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 projects 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. Secondly, the user research aimed to identify the target audience – i.e. draw the profile of researchers who use the collections of MADDLAIN institutions – and their needs. A user requirements survey was therefore conducted to evaluate how professional researchers use the online services of MADDLAIN institutions 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 their research.

Among the various ethnographic methods to conduct a user experience research, the qualitative survey via semi-structured interviews appeared to suit best the overall aims of the project. Interviews based on open-ended questions provided a more flexible approach than a written survey disseminated online, for instance. Although a written survey would have resulted in a higher response rate, thus providing a better representativeness, interviews allowed to cover larger themes, left enough room for researchers to reflect thoroughly on their practices, and gave way to a necessary interaction between interviewer and interviewee.

The concept of scholarly primitives/activities provided the conceptual framework for the design of the questionnaire and the interpretation of the findings. The concept of scholarly primitives was introduced by J. Unsworth, who defined them as “basic functions common to scholarly activity across disciplines, over time, and independent of theoretical orientation” (Unsworth, 2000). A similar model develop in a report commissioned by the *Online Computer Library Center* more specifically emphasised the role of information in research practices. It identifies five core scholarly activities, *searching, collecting, reading, writing, collaborating*, which encompass in turn several scholarly primitives, in addition to four cross-cutting primitives, *monitoring, notetaking, translating, data practices* (Palmer et al., 2009). For the purpose of MADDLAIN, the survey focused on those activities which concern the researchers’ information behaviour. In other words, the aim was 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.

The survey took place from November 2016 to January 2017. 15 interviews were conducted with researchers, both Dutch- and French-speaking, from Belgian universities and research institutes. The participants were mainly historians (13), including a book historian and an art historian. Two other researchers specialised in communication sciences and literature. This overrepresentation can be explained by the fact that historians represent the major part of the academic audience of archives, but is also due to the fact that the response rate from historians was considerably higher than for all the other fields included. Interviews were recorded with the consent of the participants, then fully transcribed. Significant excerpts were extracted from the transcriptions and coded in a qualitative analysis software. Codes were then revised, systematised and structured around wither themes.

## 4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

The results of the MADLAIN project have been the subject of several scientific and non-scientific publications and research reports (see section 5). The results presented here are intended to be intentionally synthetic.

### 4.1. General results

« Access » and « Communication » are the two key words that emerge from the results of the study carried out by the MADDLAIN project. The tools and means of communication used by the readers of the three institutions may have progressed considerably in the digital era, but the users have not basically adapted their needs in relation to the fundamental missions of the institution. Access to the collections is undoubtedly the first reason why users come to the State Archives, CegeSoma and the Royal Library. Activities such as workshops, publications, exhibitions or audiovisual productions only come second in the priorities of the persons interviewed.

#### 4.1.1. Access to the collections

In the survey, the question of « Access » has been considered for the paper collections as well as for digitized documents. In both cases, it would seem that the users combine several methods to access documents, by searching the digital catalogues as well as the paper inventories. The professional researchers specify that they apply several strategies in their search for information: key words, hierarchical tree, quotes, cascading research, information updating, contacts with experts, etc. The majority of the users profess satisfaction with the digital catalogues. The search engines are indeed used extensively: 20.000 single visitors each month for the State Archives, 10.000 for the Royal Library and 2.000 for CegeSoma.

Web analytics analysis and tests conducted by the MADDLAIN research team revealed that the access to the collections through both the institutional websites and search engines leaves to be desired. The website of the State Archives scored best with regard to the website functioning as a gateway to the different catalogues of the institution. However, the fact that there are four catalogues on offer (the main digital catalogue 'Search archives' and three more specialized catalogues: 'Search archives producers', 'Search Persons' and 'Search library') creates a confusing situation because it is not always clear to users in which of the catalogues they can locate the information they need. One suggestion is to remove the catalogue 'Search archives producers' from the main menus since an 'archives producer' is a term taken from archival science and largely meaningless to the general public. Users wonder what the difference is between the two catalogues with the word 'archives' in their names. The information in 'Search archives producers' contains information that is useful to archivists but not to the average user so the catalogue could be relocated to a specialised section of the website. The three remaining catalogues are less likely to confuse users since their names make clear which types of information they contain: 'archives', 'persons' and 'library'.

The website of the Royal Library scored abominable with regard to its function of serving as a gateway to the collections. The new website introduced in 2016, however, has tried to remedy this issue by placing a text field to search the main digital catalogue at the top of the homepage. It will be interesting to see to what extent this change will improve the Royal Library's score with regard to users finding their way from the homepage to the digital catalogue. The CegeSoma's website scored in between the State Archives and Royal Library with regard to guiding users to the digital catalogue. Until recently, users had to click three times in order to access the digital catalogue from the homepage of the institutional website. In March 2017 the institution changed the website so that the pathway to the digital catalogue Pallas (as well as other catalogues in the form of Excel lists) became more clearly indicated and could be accessed from two locations instead of one. However, users still need to click two times in order to access the digital catalogue Pallas. The same approach as the one taken at the Royal Library, i.e. a search field on the homepage, would greatly benefit the users of the CegeSoma. Unfortunately, this is at present technically not achievable with the current digital catalogue.

As discussed previously, the photo collection of the CegeSoma and the newspaper collection of the Royal Library are practically invisible in Google. Unfortunately, the same hold true for these two institutions metadata contained in their main catalogues. Ideally, it would be possible to find records present in the catalogues by performing a search in search engines such as Google. This is, for example, the case for commercial websites such as Amazon as well as the digital catalogue of the State Archives. A portion of the metadata relating to the collection of the CegeSoma can only be found in Excel lists and this is also not beneficial for their findability. Ideally, all metadata would be accessible through the digital catalogue Pallas.

The MADDLAIN project revealed another element concerning the acces to the collections : the descriptions concerning the collections leave a lot to be desired. Staff members explained that in the past retro-conversion projects have taken place that were not always completed satisfactorily as there was not enough personal to perform a quality control of the scanned index cards and to correct any errors. Participants of the user survey also indicated that they often come across mistakes in the digital catalogues and that when these mistakes are signaled they are not always corrected. Since the institutions still suffer from a lack of personnel, they should look into other options to improve the quality of their metadata. One way is to engage the general public through 'crowdsourcing'.

Crowdsourcing projects are projects in which institutions request the help of the public in order to improve the accessibility of their collections. Users can help online by adding keywords/comments to descriptions in a catalogue or by completing specific tasks set by the institution (e.g. comparing a digital description with a scanned index card). A third of the user survey participants indicated they were interested in both types of crowdsourcing and a further 15% found one of the types appealing, so almost half of those surveyed were interested in contributing their time in order to improve the institutions' collections. Almost

30% of participants claimed to have no opinion on the subject so if the institutions launch interesting projects it might be possible to also persuade those respondents to participate.

At present, only the State Archives have a crowdsourcing project, i.e. Demogen. This project focuses on genealogical sources and leads to the creation of an enormous database containing personal names and dates, taken from birth certificates, marriage certificates, death certificates etc. Participants need to download a software package in order to be able to contribute to the project. Demogen was launched in 2007 and since then a little over 400.000 certificates have been analysed, which accounts for around 2 million names in the database 'Searching persons'. In total, the database contains around 28,5 million names. This entails that around 7% of all the names in the database have been added by people participating in the crowdsourcing project. The large majority of names, however, was contributed by the volunteers of the various depots of the State Archives. Since the system has been running for 10 years, this entails that per year around 40.000 scans are treated by people working at home through the Demogen application.

Comparing Demogen with other initiatives can shed light on the success of the State Archives' application. A popular crowdsourcing platform in the Netherlands is [www.velehanden.nl](http://www.velehanden.nl). Its first project was launched in 2011 and a year and a half later the project was finalised, with more than 300.000 archive records treated and more than 1,5 million person names indexed. On the website, other successful projects can be seen and in some (e.g. 'Tot uw Dienst!', 'Volgende Patiënt!' and 'Glashelder!') more than 10.000 scans were treated per month. This entails that there are projects that process more than 120.000 scans per year, which is three times as much as Demogen. Most projects on 'Vele Handen' originate from city archives so as a national institution, the State Archives should have a much larger pool of possible participants.

What is interesting in the user survey results is that there is a difference between the three institutions: of the State Archives users around half wish to participate in crowdsourcing projects while at the CegeSoma and the Royal Library the amount of users who are claim to be open to this kind of projects is around 8% higher. When we look at the results for the State Archives split into three groups, genealogists, professional users and people who make use of the sources for personal reasons, we see that the last two groups have a percentage similar to the CegeSoma and the Royal Library, while the genealogists have a much lower percentage, i.e. only 43%. This is striking as the Demogen system was specifically created for genealogical sources. The user survey does, however, suggest that this group is the last likely to participate in such endeavours. A possible reason for this situation might be that genealogists make use of various other websites (e.g. Family Search) where a lot of quality information can be found without the users having to contribute in any way (no payment system and no crowdsourcing system).

Around one in three genealogists indicated they had no opinion on the matter, which seems to suggest that they are not familiar with what crowdsourcing entails. For the CegeSoma and the Royal Library only one in five participants indicated that they did not have an opinion and



for the users who use the State Archives for either personal or professional reasons, the percentage was even lower. It is therefore particularly notable that the group for which a crowdsourcing system is currently in place, does not have an opinion on the practice. This seems to suggest that the current Demogen system does not reach the intended audience. Analysis of the data collected through Web Analytics seems to confirm this hypothesis: in a six month period the webpages presenting Demogen was only consulted during 680 visits. In contrast, the webpage that explains the contents of the name database and provides a link to the database was consulted during more than 20.000 visits.

Last point but not least : the policy of digitization of the institutions. When cultural institutions and commercial companies such as Google started digitising collections it was thought that it was possible to eventually digitize all human knowledge that was ever produced. Nowadays, however, it is clear for information professionals that this goal cannot easily be attained. The digitisation department of the State Archives, for example, calculated that if the institution continues its digitisation efforts at the same rate as today, it will take 9000 years to digitize the current collection completely. It is therefore important that users realize that a very large portion of the collections of cultural heritage institutions cannot be accessed through digital technologies. The sources that have been digitised are also not all freely available online. There are various reasons for this: institutions might wish to charge for the consultation of the sources they digitised, the metadata of the sources might be insufficient and cause them to be impossible to find, there might be copyright issues that prevent the sources from being placed on the internet or the sources might contain information of a sensitive nature, which entails privacy laws prohibit their distribution. The average user of heritage collections, however, does not realize that many sources are only accessible inside the institutions that preserve them.

The institutions studied in the MADDLAIN project attempt to digitize substantial parts of their collections but unfortunately they suffer from a lack of structural funding. The institutions largely have to rely on subsidies from the Belgian Science Policy Office (currently within the context of DIGIT03) and other projects that contain a budget for digitisation, for example because their aim is to study a particular part of the collection and having digital access to the sources makes it easier to conduct research. This entails that the institutions can almost never work on long term projects and can only digitize their collections in small increments. Consequently, choices with regard to which sources will be digitised have to be made and the institutions have to accept that they will be able to make their full collections digitally available. Additionally, they are confronted with the fact that most projects only contain financial resources for the digitisation of sources and not for their storage valorisation or transfer to other digital storage media in order to guarantee durable storage.

#### 4.1.2. Communication

The question of communication is also a key question in the results of the survey. The institutions are used to communicate extensively through their websites, social media profiles and press releases about events (exhibitions, workshops, conferences...) but the

collections are much less frequently placed in the spotlights. When the institutions do communicate about their collections it is usually about the master pieces of their collections. It be beneficial to the users if the institutions would more clearly communicate about which metadata can be found in their digital catalogues and which items of their collections are also available in digital format. The way in which the National Archives of the United Kingdom communicate about their collections could serve as an example. On their website they make clear which sources they preserve and which they do not as well as what can be found online and what not. The institutions are continuously trying to add metadata and digitised documents to their catalogues as well as new digitised documents. The communication concerning these efforts could be greatly improved towards both staff members and external users. It would also be a good idea to explain clearly to users which steps have to be taken in order to digitise documents so that they will better understand why making collections available online is very time consuming. The user survey revealed that certain users are unaware that the institutions have to follow copyright and privacy legislation when they make information publicly available. A remarkable example is the fact that quite a few genealogists who participated in the survey were unaware that personal details can only be made available after 100 years have passed. The State Archives might consider communicating more clearly about the consequences of the applicable legislation since that might prevent users of getting frustrated because they cannot find certain information.

#### 4.1.3. The valorisation of the collections

The User survey revealed that the user public that consults the institutions' collections regards digital access to those collections as the top priority. For all three institutions the option of a catalogue that offers direct access to digitised original sources was by far more popular than all other options. It seems that the public that consults the collections essentially wishes a better access to the collections and regards other types of valorisation as nice additions. The institutions should, however, not stop the development of all other forms of valorisation. The User Survey mostly attracted answers from people who actively make use of the institutions and not the general public. Over the past few years the institutions have organised exhibitions, for example the one about king Albert and queen Elizabeth, that were well received by the wider public. Also publications aimed at a wider public, for example Knack Historia, remain popular.

#### 4.1.4. Studying User's Digital Practices and Needs in Archives and Libraries

In addition to the results presented above, the MADDLAIN project also allowed the institutions to question the most appropriate way of communicating with different types of audiences in order to improve the user experience. At the end of a 2-year process that involved 6 different researchers, attracted the participation of 2,300 external users, brought to dozens of hours of interviews and the analysis of millions of data, two essential questions arise: how to transform the results obtained into concrete actions and how to continue the dialogue started with the public outside a specific research project?

If it is essential to communicate with your audience, listen to their grievances, create a dialogue to establish a relationship of trust, it is even more fundamental to perform actions that meet their expectations and that shows that the institution holds really account for his needs. This proactive approach requires that operational teams can be aware of the improvements to be made to the various digital tools and services offered by the institution as well as about the opportunities for their future development. This means being able to get out of the scientific reports of the project, to draw the substantive marrow of the results, to identify the priority actions of others, those which can be easily realized from those which require more investments of time and money.

The first concern of the institutions was therefore to organize in-house information sessions on the results of the project both for the managers of the various departments and for the staff responsible for encoding, scanning or having relations with the visitors in the reading rooms. Collective workshops on the model of the World Café were also organized, offering everyone the opportunity to react on the results presented and to share their ideas. The final goal is to write a general policy note on digital development in each institution.

In addition to internal communication, the dissemination of the results to the general public also appeared as a key element, the aim being to satisfy his curiosity about the findings of the survey while showing him the importance of his participation. A poster and a brochure summarizing the results of the survey were produced.

In parallel with the communication work on the project, initial actions have been undertaken. The aim of the institutions was to respond quickly and effectively to grievances concerning easily-resolved issues in a relatively short time and with an investment of time and money that remained realistic. As examples, the State Archives in Belgium have produced a small explanatory video on the functioning of their search engine while CegeSoma has clarified the access to the collections on the home page of its website modifying certain terms or adding access points.

The second essential question that arises at the end of a project such as the MADDLAIN project is that of continuing the dialogue with the public outside the research work carried out for two long years and therefore with the teams in place in the institution. Conducting field surveys, collecting structured data and analyzing them requires indeed both time and skills that are sometimes lacking. The priority of the researchers was therefore to be able to propose simple and adapted solutions that could be easily followed by the staff concerned, particularly through an article detailing the methodology applied during the project.

On a methodological point of view, the web analytics research has resulted in seven main findings:

- The creation and documentation of “custom data” to manage ambiguous URLs for several websites involved in the Maddlain Project (details can be found in Article Symposium)

- The archiving and the development of a sustainable access to more than one year of Piwik data for the 12 involved websites via a Virtual Machine.
- The definition of new “goals” specific to the cultural heritage context. Each goal is related to the access to different types of content: practical information; digital/digitised heritage content; meta-information; news and agenda; help and tutorial; valorisation websites/documents; main catalogue; metadata related to the collection; ordering reproductions (details can be found in Hungenaert and Chardonnnens (2017)).
- A typology of the limitations encountered by applying the “goals” method (details can be found in Hungenaert and Chardonnnens (2017)):
- The development of Python scripts to extract user queries and perform text mining techniques on them (details can be found in Hungenaert and Chardonnnens (2017)).
- An exhaustive report presenting quantitative data per website and a comparative overview, based on a list of selected metrics:
  - Comparative view: visits and visitors numbers; visits by referrers; visits by device type; visits by country and by region; visits by the browser language; bounce rate; visitor loyalty; visit duration; visit depth;
  - Specific analysis per website: synthetic summary by website; traffic per hour; details about referrers; entry and exit pages; conversion rates; when relevant: user queries and user search behaviour.

## **4.2. The mediation tools**

### **4.2.1. Current practices in the participating FSI**

Researching current digital mediation practices resulted in an Excel sheet describing the content that is publicly available to users of the participating FSI and the way the content is structured. The Excel sheets can be used as a template with indicator components for evaluation for other institutions or future practices. The data can be used as a means to evaluate one's practices and reflect on topics such as (a) what exactly do we put online? Do users need this information or other? (b) how do we structure our information? Is the structure ideal for the type of information we are providing? (c) is our content centralized or scattered across the website? Do users quickly retrieve the information they need? (d) do we differentiate in order to approach different audiences? (e) how many clicks does it take to reach certain informative pages? Is the information too deeply buried in the website? etcetera.

The online staff survey helped identify perceived ease of use of the digital catalogues and services, concerning the use of the digital catalogues, that are currently offered. A report was written in order to cluster certain topics. The raw data can be consulted for additional insight.

**Notable findings**

- AGR and CEGESOMA: The State Archives provide separate, specific information for 6 different profiles in such a way that there is less (unnecessary) information overload. Other, general digital information on arch.be is scattered across at least 13 webpages. Almost all information is structured as text with hyperlinks. About half of these include either bulleted or numbered lists. One webpage refers to a video tutorial. Search.arch.be tends to have more information in a FAQ (answered questions) format and databases for extra, structured information. None of the information on the arch.be websites is deeply hidden in the website, mostly because of the drop down menu on the homepage.

31 staff members who are in contact with the public of the State Archives for a total of about 232 employees (0.13:1) answered the survey. 37% of the respondents affirm that there is a service system in place to accompany users in the use of the digital catalogue. Current user accompaniment consists of in-person explanation in the reading rooms, information service by telephone and by e-mail. Since August 2016, a video tutorial on the use of the digital catalogue is featured on the State Archives' YouTube channel in the Dutch and French language.

The Cegesoma provides information on their digital collections through at least 10 webpages. On almost all of these pages, the collections are described using text, often with hyperlinks. Occasionally, a static image is added to the webpage. The depth of the information ranges from 2 to 6 clicks, which is a considerable depth.

12 staff members who are in contact with the public of the Cegesoma for a total of about 44 employees (0.27:1) answered the survey about accompaniment of users. 33% of the respondents affirm that there is a service system in place to accompany users in the use of the digital catalogue. The responses to the survey indicate that services provided by the CegeSoma are based upon one-on-one contact between a member of the personnel that is present at the reading room and the user of the archive. It isn't clear to which extent external users can make use of information services, e.g. by e-mail or telephone.

- KBR: The Royal Library describes digital collections and related information across at least 13 webpages. These pages contain mostly textual information, bulleted lists and hyperlinks. Most information is buried no deeper than 3 clicks into the website kbr.be.

33 staff members who are in contact with the public of the Royal Library for a total of about 270 employees (0.12:1) answered the survey. 65% of the respondents affirm that there is a service system in place to accompany users in the use of the digital catalogue. Accompaniment is currently offered mainly via in-person guidance at the reading rooms: step by step guidance at the computer, guidance is provided if people

ask questions, and if people seem hesitant about using the OPAC, the staff will approach them discretely and try to solve their issues. During a guided visit or workshops, users get a demo to learn how the OPAC works. Explanation by e-mail and by telephone is also provided. If users find themselves at home, they can consult the online manual, but the personnel agrees that it isn't the most useful tool. There are webpages with information on library topics.

### ***Limits***

Researching components of a website is an objective tool. No information may be wilfully left out. However, due to websites' ephemeral nature, the current analysis may be outdated soon.

The responses of the staff at the FSI are valuable because they are in touch with the audience. Most of the staff members who are in contact with the public responded to the staff survey, but this wasn't a certainty. Moreover, staff members can only share information about users who communicate with them. Users who don't ask questions or file their complaints are left out of the equation.

#### **4.2.2. Identifying user needs with regard to digital mediation and describing user behaviour**

### ***Value***

An analysis of the user survey resulted in a PDF containing a summary of responses to relevant questions (i.e. relating in some way or another to the topic of digital mediation) and a brief analysis from the perspective of digital mediation and online user education. The survey and its data are an essential, direct means for identifying user needs and behaviour. Additionally, the online staff survey helped identify frequently asked questions concerning the use of the digital catalogues. A report was written in order to cluster certain topics. The raw data can be consulted for additional insight.

### ***Findings and recommendations***

- AGR and CegeSoma: Most respondents who do not consult physical sources at the State Archives point to time and travel constraints. More than 20% respondents find the sources in a digital format sufficient for their research. Archival description seems to be the most important to the respondents who use the search engines at a distance. Most people would appreciate a webpage explaining what can be found online, why some information is not (yet) available, etcetera. Note it is unclear whether or not users' dissatisfaction concerning the performance of the search engine stems from a poor know-how of archival structure. A minority of less than 10% consult the collections at the State Archives only. Respondents mainly travel to the State Archives for complementary research or to ask questions. Almost 80% of the respondents (French-speaking excluded) consult the search engine 'Search

Persons' at locations other than the State Archives. Service at distance is strongly recommended, as these respondents rarely visit the State Archives. Archival knowledge is absolutely necessary to comprehend the underlying structure of the search engine. Users do not necessarily possess this knowledge, especially new users. Some users provide feedback about the search engine, without reply. Users have to be kept up-to-date. All in all, most respondents demand more transparency. 17 out of 30 staff members who are in contact with the public at the State Archives receive questions regarding the digital catalogues from users. When asked to list the most frequently asked questions, an array of problems come to the surface. 9 of 36 common questions concern the use of the search engines. Almost all of these questions were very broad, except for one concerning finding acts of marital status. 8 of 36 common questions concern 'missing' pieces of information: ranging from incomplete inventories, inventories in only one language, missing descriptions, etcetera. 6 of 36 common questions concern the finding of digitized images. The State Archives require users to create an online account in order to consult digital content. Most people ask general questions: "How do I find digitized material?", "How do I log into my account?". One staff member notes that people also enquire about downloading and saving digital content. 5 of 36 common concern the location of the documents. Only 2 questions regard the consultation of material, not including the questions about the location of the documents.

People are likely to consult the collections of the CegeSoma more often if the number of online digitized documents is higher. It is also advised to clearly indicate on the website what can be found in the collections and how to conduct a research. Digital mediation could help clarify at least some of the topics: explaining how keywords are attributed, explaining why some inventories aren't found online, explaining how archives are structured, explaining that searching in different languages yields different results, etcetera. Most respondents, over 60%, use Pallas, the digital catalogue, both inside and outside the CegeSoma. More than 30% use it only outside of the CegeSoma. Users who prefer to consult these digital sources outside of the CegeSoma (70%) can be accommodated through digital means only. Currently, there are few digital mediation practices in place on the CegeSoma website. 60% of the respondents who don't consult digital sources through Pallas don't because they did not know it is possible to gain access to them. It is recommended to clearly indicate what is and what isn't online.

10 out of 12 staff members who are in contact with the public at the CegeSoma receive questions regarding the use of the digital catalogue Pallas. 8 of 16 questions regard the correct and efficient use of Pallas. 5 of 16 questions concern the interpretation of search results and following steps, such as how to reserve documents and view digital images. 2 of 16 questions are on the definition of a catalogue and where to find it.

- KBR: Respondents tend to consult the digital collections but abstain from going to the library itself, mainly because of time constraint. To accommodate users that prefer digital search instruments (over 80%), specific education resources regarding the use of the digital search instruments of the Royal Library could be helpful. Note one should not however forget about the usefulness of complementary paper inventories. Respondents want to be kept up-to-date of what resources can be found through the digital catalogue. They feel the manual isn't sufficient for searching in particular collections. A good idea would be to research if users respond well to information pertaining to a particular collection. It is advised, however, to keep the general manual, and provide additional support for inexperienced users. It isn't clear to users how they can access the digital sources that the Royal Library offers. Recommendations include straightforward, intuitive web design, an online guide or webpage about all digital sources (e.g. "What we have online" versus "what not"). 17 out of 32 staff members who are in contact with the public at the Royal Library receive questions regarding the use of the OPAC. 16 of 33 common questions relating to the correct use of the OPAC. "I found the document, what's next?" makes up 9 of 33 common questions. 7 of 33 common questions related to technical issues and/or accessing the OPAC.

### ***Limits***

The responses of the users were especially valuable when it came to qualitative data (open answers, comments) on digital mediation. However, a limited number of users take the time to write out their opinion, cutting down on representation.

The responses of the staff at the FSI are valuable because they are in touch with the audience. Most of the staff members who are in contact with the public responded to the staff survey, but this wasn't a certainty. Moreover, staff members can only share information about users who communicate with them. Users who don't ask questions or file their complaints are left out of the equation.

#### **4.2.3. Good practices in similar institutions**

### ***Value***

Good practices research resulted in a visual presentation of digital services in other institutions (in PDF format). Secondly, an Excel sheet in which key components of digital mediation are compared, provided information on the strengths and weaknesses in similar institutions. Finally, answers by members of the personnel of these other institutions were collected and reflect the process the institutions have gone through in order to be able to offer certain digital services. These answers were also collected (but not analysed due to lack of time) to avoid reinventing the wheel and learn from others' experiences.



## **Findings**

- AGR and CEGESOMA: The National Archives and other archives of United Kingdom, the Netherlands (including Brabants Historisch Informatie Centrum), Luxemburg, France (including the French High Alps) and Australia were researched and are all good examples to follow, with the National Archives of United Kingdom fulfilling 92% of above components. (They don't provide a definition of a finding aid.) The Dutch NIOD and French INA are decent examples to follow, with the NIOD fulfilling 58% of the above components.
- KBR: The National Library of France, United Kingdom, the Netherlands, Ireland, Australia and the libraries of Genève and New York are good examples to follow, with the National Library of France offering 83% of all components. (They don't provide video tutorials or help you plan your research through their website.)

## **Limits**

As with all web-based research, the collected data can be outdated quickly, due to the nature of the study object. Additional qualitative data, context, and background knowledge are needed to study these components in depth.

### **4.3. The virtual research environment**

The topics highlighted in the user requirements survey concerned information search, access to information, ways to collect and manage data, time management, work environment and digital turn. Five key points have implications for institutions wishing to improve their online services:

1. **Researchers use multiple strategies** to retrieve information, both online and in traditional ways. Keyword search is often used as a starting point or in an exploratory way, but researchers are often dissatisfied with online search engines, because they often generate a lot of noise and results from a query potentially have 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 main principle underlying researchers' information behaviour is comprehensiveness, keyword search needs to be combined with other strategies, such as systematic browsing of the collections, citation chaining, consulting information professionals, and monitoring.
2. **Researchers gather and manage their data within their own digital ecosystem.** Preferably, they will get a digital copy of documents by downloading them or taking pictures, thus postponing the thorough reading and analysis of the documents to a later stage. Whereas the use of a reference management software becomes increasingly common, researchers struggle with the lack of an adequate system to organise primary sources and link digital versions of documents to their annotations. They come up with highly diverse, customised, solutions to organise their research

data, which seems counterintuitive with notions of portability and potential reuse of data in the wake of open science.

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. 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.
5. **Whereas the use of online resources to search for information is now firmly anchored in research practices,** 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, suggestions can be made to improve researchers' user experience with online services.

#### 4.3.1. 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 keyword search is but one of the techniques used by researchers to retrieve information. As a result, institutions should reflect on 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 (Whitelaw, 2015). 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.

#### 4.3.2. Providing online tools to prepare on-site research

Time management turns out to be a major challenge in researchers' work practices. Since the time that they can allocate to visit an archive or a library is often limited, all kinds of tools and resources that help them prepare their visit in advance are particularly welcome. Those tools can help researchers to become familiar with the institutions and understand how the reading rooms work, to shorten administrative procedure, and to get acquainted with the collections. Digitised finding aids available online and the option to request items online in advance are particularly appreciated.

#### 4.3.3. 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 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. VREs seem to be the logical next step in terms of digital services offered by archives and libraries, in response to a tendency of users to do more and more things directly online. 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.

However, the findings of the user requirements survey 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.
- 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.

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. 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 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.

## 5. DISSEMINATION AND VALORISATION

- 27/11/2015: Presentation about *MADDLAIN*, given at the DARIAH-BE Kick-Off Meeting by Jill Hungenaert.
- 31/03/2016 : Presentation about *The MADDLAIN project: methodological aspects*, given at the Forum des Archivistes Français (Troyes) by Jill Hungenaert
- 28/10/2016: Presentation about *What does the MADDLAIN project on the public of libraries and archives tell us?*, given by Melissa Hodza, Jill Hungenaert and Stéphanie Paul at the “Lunch of the research” organized by the Royal Library of Belgium.
- 17/11/2016: Presentation about *Digital access to the collections: e-learning in the library*, given by Melissa Hodza at the “Day of the Librarian” organized by the Royal Library.
- 22/02/2017, Symposium MADDLAIN "Inside the User's Mind" organized at the Royal Library
- 07/03/2017: Presentation about *The MADDLAIN project*, given by Anne Chardonnes at the Internal workshop “Best practices exchange” organized by the Bibliothèque nationale de France (Paris)
- 13-15/03/2017: Presentation about *Text Mining for User Query Analysis. A 5-step Method for Cultural Heritage Institutions*, given by Anne Chardonnes at the 15th International Symposium of Information Science (Humboldt-Universität zu Berlin).
- 15/05/2017: Presentation of the results of the MADDLAIN project, given by Florence Gillet, Rolande Depoortere and Stéphanie Paul at the day “Printemps scientifique” organized by the State Archives. Organisation of a World café with the colleagues to get their reactions to the results.

## 6. PUBLICATIONS

### Peer review:

- Chardonens, Anne and Simon Hengchen, 'Text Mining for User Query Analysis: A 5-Step Method for Cultural Heritage Institutions' in *Proceedings of the 15th International Symposium on Information Science (ISI 2017)*, Berlin, Germany, 13th-15th March 2017.
- Hungenaert, Jill and Anne Chardonens, 'Combining Quantitative and Qualitative Methodologies to Analyse User Behaviour and Access Paths to Digital Content: Case Studies from Three Belgian Federal Scientific Institutions' in *Archives et Bibliothèques de Belgique*, mars 2017.

### Other publications :

- Chardonens, Anne, Melissa Hodza, Jill Hungenaert and Stéphanie Paul, 'MADDLAIN montre la voie: Une recherche sur les utilisateurs pour améliorer l'offre numérique des institutions patrimoniales' in *Science Connection* (53), march-april-may 2017.
- *Inside the user's mind. Studying User's digital practices and needs in Archives and Libraries*, Proceedings of the conference organised on 22 February 2017 as part of the BRAIN project MADDLAIN, Archives de l'Etat-Bibliothèque royale de Belgique, 143 p.

### Reports :

All these reports are online available at the following [place](#) :

- *Digital technology in the Royal Library, the State archives and the CegeSoma. The results of a study on the digital needs and practices of the users*, a brochure containing a summary of the main issues that need to be realized, 2017, 9 p.
- *Digital technology in the Royal Library, the State archives and the CegeSoma. The results of a study on the digital needs and practices of the users*, poster containing a summary of the main issues that need to be realized, 2017, 1 p.
- Gillet, Florence and Hungenaert, Jill, *Studying user's digital practices and needs in Archives and Libraries. Final Report of the MADDLAIN project.*, 2017, 48 p.
- Paul, Stéphanie, *Virtual Research Environments. MADDLAIN final report*, 2017, 61 p.
- Hodza, Melissa, *Digital mediation and online user education: meeting the needs of archive and library users*, 2017, 14 p.

- Statistics reports of the raw data concerning the general public MADDLAIN survey by category :
  - o all users 73p.
  - o genealogists 47p.
  - o those who use the collections of the State Archives for personal reasons 39p.
  - o those who use the collections of the State Archives for professional reasons 38p.
- Raw data MADDLAIN Web Analytics 198p.

## **7. ACKNOWLEDGEMENTS**

Word of thanks goes out to :

- Members of the personnel at the three studied institutions (Royal Library, State Archives and CegeSoma) who accepted to collaborate to the interviews
- The 2300 users having collaborated in our online survey. That the project has produced useful and relevant results is due in large part to their contribution.
- Members of the personnel at the Royal Library and the CegeSoma for the help in the organization of the MADDLAIN symposium on the 22 February 2017.
- Members of the personnel at other institutions for filling out the good practices questionnaire for the e-learning study: the ANLux, the INA, the KADOC, the NaNed, NLA, BHIC, BnF, the Archives Départementales des Hautes-Alpes and Bibliothèques de la Ville de Genève.

## **8. ANNEXES**

All the data and the full reports of the projects are online available at at the following [place](#).