



## DELIVERABLE

**Project Acronym:** Europeana Cloud  
**Grant Agreement number:** 325091  
**Project Title:** Europeana Cloud: Unlocking Europe's Research via The Cloud

---

**Deliverable D1.2 – State of the art report on digital research practices, tools and scholarly content use**

**Revision: Final**

---

**Authors:**

**Agiatis Benardou (Athena RC)**  
**Costis Dallas (Athena RC)**  
**Eliza Papaki (Athena RC)**  
**Panos Constantopoulos (Athena RC)**  
**Stavros Angelis (Athena RC)**  
**Thomas Baldwin (CERL)**  
**Marian Lefferts (CERL)**  
**Lorna Hughes (NLW)**  
**Owain Roberts (NLW)**  
**Stefan Ekman (UGOT/SND)**  
**Björn Sjögren (UGOT/SND)**

<b>Project co-funded by the European Commission within the ICT Policy Support Programme</b>		
<b>Dissemination Level</b>		
<b>PU</b>	<b>Public</b>	

## Revision History

Revision	Date	Author	Organisation	Description
<b>Scholarly Content Use state of the art</b>				
1	24.05.13	Eliza Papaki, Agiatis Benardou	Athena RC	Initial draft
2	07.06.13	Eliza Papaki	Athena RC	Update draft
3	27.06.13	Stefan Ekman	UGOT	Editing
4	04.07.13	Björn Sjögren	UGOT	Editing
5	10.07.13	Thomas Baldwin, Marian Lefferts	CERL	Editing
FINAL	10.07.13	Eliza Papaki, Agiatis Benardou	Athena RC	Processed, Editing
<b>Digital Research Practices state of the art</b>				
1	10.04.2013	Eliza Papaki, Agiatis Benardou, Costis Dallas, Panos Constantopoulos, Stavros Angelis	Athena RC	Initial draft
2	16.04.2013	Eliza Papaki	Athena RC	Update draft
3	24.05.2013	Stefan Ekman, Björn Sjögren	UGOT	Editing
FINAL	17.06.2013	Eliza Papaki, Agiatis Benardou	Athena RC	Processed, Editing
<b>Digital Tools state of the art</b>				
1	28.06.2013	Owain Roberts, Lorna Hughes	NLW	Initial draft
FINAL	31.07.2013	Owain Roberts, Lorna Hughes	NLW	Update draft

### Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

## **Deliverable D1.2 – State of the art report on digital research practices, tools and scholarly content use**

### **Executive summary**

This document constitutes desk research to analyse the current situation related to digital research practices, tools and content for the humanities and social sciences research communities and it comprises of three Milestones submitted separately in the course of the first nine months of the Project, each corresponding to digital research practices, tools and content respectively. Moreover, this report will serve as an exploratory stage which will provide the basis for the confirmation stage, that is the web survey undertaken within Task 1.3, relying heavily on work previously presented in Deliverable 1.1 (Research Communities Identification and Definition Report), in which the humanities and social sciences research communities were largely identified. The Research Communities Web Survey (T.1.3.4), which has been completed and is currently being processed, encompasses - amongst others - questions pertaining to the activities undertaken by Humanities and Social Sciences researchers, as well as questions regarding the use of content and metadata. This current document provides the background work which guided the design of the Web Survey and is meant to complement the related Web Survey findings which will comprise the User Requirements and Case Studies Analysis report, D1.3. Moreover, this work draws from and complements Deliverable 1.5 (Case Studies Expert Forum Report. The present desk research seeks to analyse the current situation related to digital research practices, tools and content for the humanities and social sciences research community, and will serve as an exploratory stage which will provide the basis for the confirmatory stage, that is the Content Strategy and User Requirements reports, due later on in the Project.



## Table of Contents

<b>1. Introduction .....</b>	<b>3</b>
<b>2. Desk Research: Scholarly Content Use State of the art.....</b>	<b>6</b>
2.1 State of the art: A literature review on scholarly content use.....	6
2.2 Documenting scholarly content use.....	12
2.2.1 Images.....	12
2.2.2 Text.....	13
2.2.3 Video and Sound.....	16
2.3 Conclusion.....	17
<b>3. Desk Research: Digital Research Practices State of the art.....</b>	<b>19</b>
3.1 State of the art: An overview of user information behaviour studies.....	19
3.2 Documenting digital research practices.....	24
3.2.1 Searching.....	24
3.2.2 Collecting.....	31
3.2.3 Reading and Writing.....	33
3.2.4 Collaborating.....	35
<b>4. Desk Research: Digital Tools State of the art.....</b>	<b>38</b>
4.1 What is a digital tool?.....	38
4.2 Early digital tools.....	39
4.3 The role of tools in the scholarly lifecycle.....	40
4.4 Digital tools in the context of scholarly primitives.....	41
4.5 State of the art tools.....	43
4.6 Tools in relation to Europeana Cloud.....	44
4.6.1 Text.....	45
4.6.2 Images.....	48
4.6.3 Video and Sound.....	49
4.6.4 Spatial Technologies (Includes 3D).....	50
4.7 Visualisation and Europeana Content Analysis.....	50
4.8 Conclusions on research into digital tools.....	50
<b>5. Conclusion.....</b>	<b>53</b>
<b>6. References.....</b>	<b>54</b>

---

## List of Tables

Table 1: Overall impact on quality of research of electronic research.....	22
Table 2: Essential access and discovery methods for percentage of survey.....	26
Table 3: Percentage rating sources ‘essential’ (Sample from U.K. RAE Census).....	29

## List of Figures

Figure 1: Essential access and discovery methods for percentage of surveys.....	27
Figure 2: Sources rated as ‘essential’ .....	30
Figure 3: Scholarly primitives associated with disciplinary approach.....	41

## 1. Introduction

Europeana Cloud will establish a cloud-based system for Europeana and its aggregators to store metadata and, eventually, content. It will provide new content, new metadata, a new linked storage system, new tools and services for researchers and a new platform – Europeana Research – by linking data, functionalities and people. Its users will stem from multiple backgrounds and disciplines, and they interact with Europeana following diverse goals, building on dissimilar data and methodologies, and employing various entry-points into Europeana. This platform is being developed by an iterative process throughout the construction of Europeana Research and beyond. Hence, while the key concepts are likely to remain stable, Europeana Research will grow and evolve over time.

This document constitutes desk research to analyse the current situation related to digital research practices, tools and content for the humanities and social sciences research communities and it comprises of three Milestones submitted separately in the course of the first nine months of the Project, each corresponding to digital research practices, tools and content respectively. Moreover, this report will serve as an exploratory stage which will provide the basis for the confirmation stage, that is the web survey undertaken within Task 1.3, relying heavily on work previously presented in Deliverable 1.1 (Research Communities Identification and Definition Report), in which the humanities and social sciences research communities were largely identified. The Research Communities Web Survey (T.1.3.4), which has been completed and is currently being processed, encompasses - amongst others - questions pertaining to the activities undertaken by Humanities and Social Sciences researchers, as well as questions regarding the use of content and metadata.

<b>Web Survey Activity-related Questions</b>	<b>Web Survey Content-related Questions</b>
Assess how important are the following activities as you seek relevant information for your research	Assess the importance of the following aspects of the content of a resource as you seek relevant information for your research
Please give an example of a common query you may ask as you seek relevant information for your research	Assess the importance of the following properties of a resource as you seek relevant information for your research
Assess the importance of the following properties of a resource as you seek relevant information for your research	

Assess how important is each of the following activities as you organize your research	
Assess how important is each of the following activities as you study and annotate information relevant to your research	
Assess how important is each of the following activities as you work with others for your research	

This current document provides the background work which guided the design of the Web Survey and is meant to complement the related Web Survey findings which will comprise the User Requirements and Case Studies Analysis report, D1.3.

Moreover, this work draws from and complements Deliverable 1.5 (Case Studies Expert Forum Report), in which observations and outcomes generated by the disciplinary-based groups were consolidated under the following broad subject headings: quality and user ranking of metadata, multilingual resources, easily recognizable date ranges, spatial and temporal mapping of results, and more transparent citation methods. In addition to that, according to the Case Studies Expert Forum Report, there was an overall consensus from researchers both from the Humanities and the Social Sciences that metadata sets should be able to be exported into specialized, often domain-specific tools. Therefore, it was strongly recommended that Europeana not develop specialized tools, but rather focus, in the context of the Europeana Cloud project, on developing tools which work with existing Europeana data and would be of specific benefit to researchers. For example:

a) generalised tools (as mentioned above) to aid in discovery tools that allow for (meta)data export into a variety of formats, b) tools that provide more sophisticated and targeted filtering of results than are currently available. To complement these empirical findings, this present document will seek, through extensive desk research and based on previous relevant experience from other projects and initiatives, to tackle issues such as the ways in digital tools should be categorised since categories of tools for the Digital Humanities are not very well defined, or c) tools that could not be applied to current Europeana content but could prove of use to Europeana Research in the future.

The present desk research seeks to analyse the current situation related to digital research practices, tools and content for the humanities and social sciences research community, and will

serve as an exploratory stage which will provide the basis for the confirmatory stage, that is the Content Strategy and User Requirements reports, due later on in the Project.



---

## 2. Desk Research: Scholarly Content Use State of the art

---

This milestone report is part of the Work Package 1 of Europeana Cloud, entitled “Assessing Researcher Needs in the Cloud and Ensuring Community Engagement”, and outcome of the Task 1.2.1 on state of the art on scholarly content use. Building on a survey of relevant user information behaviour literature, the report presents the results of desk research undertaken focusing on the research communities of the humanities and social sciences.<sup>1</sup> As researchers and content constitute two core elements in a research pattern which is constantly changing, a thorough exploration of all the dimensions defining and affecting their interaction is required in order to understand researchers’ use of content. This milestone will further contribute in the Deliverable 1.2 “State of the art report on digital research practices, tools and scholarly content use”.

### 2.1 State of the art: A literature review on scholarly content use

User information behaviour literature and surveys document a lively discussion on the interaction of researchers with content. This relationship is of great significance in terms of comprehending the research process and interpreting the research results produced. Nowadays, in particular, with the increasing domination of digital content various questions occur regarding the ways in which researchers approach and use content. How does digital content change research practices? How do researchers from different disciplines interact with different types of objects? Does this have an impact on the final outcome of a researcher’s work?

The discussion on scholarly content use is contained in a significant number of articles in the context of digital research practices. The desk research has revealed that the majority of the relevant literature examines the use of content by researchers only in a much wider context. In an effort to shape an overview of the information behaviour of researchers towards content, the literature review conducted aimed at gathering relevant points with regard to the fields of humanities and social sciences. The literature chosen to be explored is quite recent, published since the 1990s. While earlier works focused on the nature of the content preferred and used by researchers - analogue or digital - more recent articles have engaged in a new discussion on the potentials offered in the elaboration of this digital content. As technology advances and new research methods emerge, scholarly content use becomes critical in reaching innovative results.

---

<sup>1</sup> In an effort to avoid confusion and reach a clear definition of the terms “disciplines” and “fields”, this report adopts the term “field” to describe humanities and social sciences which cover a broad range of “disciplines” or “subject areas” following the definitions given in Europeana Cloud Deliverable 1.1 “Research communities identification and definition report”, 2.1 “Subject domains-Humanities” and 2.2 “Subject domains-Social Sciences”.

Prior to proceeding to a literature analysis, it is considered necessary to first define the term *content*. As this term can entail almost anything, the content of Europeana was used here in reaching a definition. Aiming to develop a Content Strategy for Europeana Research, this desk research focuses on the digital collections of Europeana in order to document the types of objects included in the term *content*. Therefore, in order to narrow this area to specific cases of interest for this milestone report, types of content will be presented alongside their genres which provide a more detailed view of the content to be addressed. According to Europeana Content Analysis Report of 2010 as well as Europeana's website, types of content are considered texts, images, video, sound and 3D.<sup>2</sup> More particularly, the category of 'text' is illustrated with genres such as books, periodicals and journals, diaries, poems, theses, newspapers, manuscripts and pamphlets while 'image' includes paintings, drawings, prints, photographs, maps, illustrations, drawings and furnishings. The rest of the categories include films, documentaries, newsreel and television programmes for 'video'; music, speech, art and sound effects for 'sound' and virtual 3D representations of objects, architecture or places for '3D'.<sup>3</sup> It is obvious that Europeana offers a wide variety of possible research content. On the other hand, despite the fact that these genres are not equally popular among researchers, they are all traced in the research process of humanists and social scientists suggesting thus the diversity of material required in both disciplines.

Exploring scholarly content use through literature led mostly to tracing articles on information seeking behaviour. This observation clearly connects content primarily with particular research activities, namely searching, chaining, browsing and accessing.<sup>4</sup> As researchers begin their journey, tracing and gathering sources are indispensable parts of the process and therefore this discussion occupies a wide area in bibliography on user information behaviour.<sup>5</sup> All these activities mark a common area of interaction between researchers and content; however, their contact is not limited in the main research activity of searching. Examining, for example, the Arts and Humanities Data Service Taxonomy of Computational Methods (AHDS), four stages of activity are suggested:

---

<sup>2</sup> Europeana, "Europeana Content Analysis Summer 2010", [pro.europeana.eu/documents/869571/874106/Content+A...](http://pro.europeana.eu/documents/869571/874106/Content+A...), 2010, p. 4, (accessed 14 October 2013). The category of 3D is only present in Europeana's website [http://www.europeana.eu/portal/usingeuropeana\\_results.html](http://www.europeana.eu/portal/usingeuropeana_results.html), (accessed 3 June 2013).

<sup>3</sup> For examples of genres of Europeana's content visit [http://www.europeana.eu/portal/usingeuropeana\\_results.html](http://www.europeana.eu/portal/usingeuropeana_results.html), (accessed 3 June 2013); Information was also taken from the updated Europeana 2013 Vocabulary, unpublished report.

<sup>4</sup> C. Palmer, L. Tefteau and C. Pirmann, "Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development", Report commissioned by OCLC Research, January 2009, pp. 9-15, published online at: [www.oclc.org/programs/publications/reports/2009-02.pdf](http://www.oclc.org/programs/publications/reports/2009-02.pdf), (accessed 10 April 2013).

<sup>5</sup> Examples of such bibliography: J. Rutner and R. Schonfeld, "Supporting the Changing Research Practices of Historians", *ITHAKA S+R*, December 2012, pp. 1-52, <http://www.sr.ithaka.org/sites/default/files/reports/supporting-the-changing-research-practices-of-historians.pdf>, (accessed 5 May 2013); R. Griffiths, M. Dawson and M. Rascoff, "Scholarly Communications in the History Discipline", *ITHAKA Report Commissioned by JSTOR*, August 2006, pp. 1-14, [www.sr.ithaka.org/download/file/fid/485](http://www.sr.ithaka.org/download/file/fid/485), (accessed 5 April 2013).

capture, structuring and enhancement, analysis, dissemination and presentation.<sup>6</sup> Matching research methods to content reveals the diversity of actions undertaken by each researcher to reach, access, elaborate and disseminate content and, in this way, the significance of more centralized studies in order to capture all dimensions of scholarly content use.

Initiating the discussion on literature, a multidisciplinary study by John Houghton, Colin Steele and Margaret Henty serves as an introduction to the changing research practices of all disciplines and on the significance attributed to digital or analogue sources by researchers in Arts and Humanities, Social sciences, Medical and Biological Sciences, Physical Sciences and Engineering and Area Studies and Languages.<sup>7</sup> References on content are present mostly with regard to how essential they are rated by researchers, aiming to distinguish, firstly, the popularity of digital and analogue sources and, secondly, the different approaches between disciplines. Despite the fact that this paper focuses mostly on the above issue and does not touch upon all research activities related to content, it contributes in shaping an overview on the various perceptions of digital resources and how fast these have evolved independently in each discipline. A number of its arguments are based on a survey conducted by the 2001 Research Assessment Exercise (RAE) Census in the United Kingdom which appeared to serve as sample survey in other papers as well. An example of a study which based its arguments – regarding the nature and range of required research material, nature of access to this material and its location – partly on RAE 2001 is “Researchers’ use of Libraries and other Information sources: current patterns and future trends” by the Education for Change LTD, University of Brighton and The Research Partnership.<sup>8</sup> This work has the same multidisciplinary character but is more content-focused as research material and its treatment from researchers are set at the centre of the analysis.

Moving on to studies focused on the Humanities and Social sciences, a study of the University of Washington in 2005 similarly approached scholarly content use in the context of the general information behaviour of scholars.<sup>9</sup> Examining the usage of digital sources, researchers in these two disciplines were perceived as a unity with regards to their approaches and the methods they adopted. In this way, similarities rather than discrepancies were stressed. In contrast, an interesting

---

<sup>6</sup>R. Speck, “The AHDS Taxonomy of Computational Methods”, Arts and Humanities Data Service, 2008, pp. 1-14, <http://www.ahds.ac.uk/>, (accessed 30 May 2013).

<sup>7</sup> J. Houghton, C. Steele and M. Henty, Changing Research Practices in the Digital Information and Communication Environment, PhD Thesis, Department of Education, Science and Training, 2003, pp. 1-186, [www.cfses.com/documents/Changing\\_Research\\_Practices.pdf](http://www.cfses.com/documents/Changing_Research_Practices.pdf), (accessed 20 February 2013).

<sup>8</sup> Education for Change Ltd, SIRU, “Researchers’ use of libraries and other information sources: current patterns and future trends”, The Research Partnership, University of Brighton, 2002, pp. 1-84, [www.rslg.ac.uk/research/libuse/LUrep1.pdf](http://www.rslg.ac.uk/research/libuse/LUrep1.pdf), (accessed 15 March 2013).

<sup>9</sup> A. Bass et al., “The Information Behavior of Scholars in the Humanities and Social sciences”, University of Washington, Washington, 2005, pp. 1-14.

survey conducted by the British Academy in 2005 on e-resources for research in the Humanities and Social sciences is more enlightening regarding the different research approaches of the two fields.<sup>10</sup> Presenting the questionnaire and the recorded answers, the authors distinguish the views expressed by humanists or social scientists and reach conclusions on their distinct characteristics with regard to digital content. Having a digital orientation, this survey is quite revealing on current scholarly content use and the changes traced in the digital world and will be thus examined in more detail later in this report.

Aiming to identify particular characteristics defining humanists and social scientists, in an effort to understand the research process, its needs and gaps, more specialized studies will be presented next. Regarding social sciences, a study by Meho and Tibbo on the Information-Seeking Behaviour allowed some revisions of earlier studies on researchers and their research methods as the emergence of digital content and the capabilities provided in Web browsers urged for examination and led to new conclusions.<sup>11</sup> In 2003, Meho and Tibbo agreed on the advantages offered by Internet in reaching digital information easily; however they noted that “the problem is that only a small fraction of available information is digitized and made accessible to researchers”.<sup>12</sup> On the other hand, the Research Information Network Report published eight years later, with a paper focused on the humanities, stresses the problem of assessing digital resources.<sup>13</sup> They suggest that the fact that many researchers prefer to cite printed rather than digital resources, despite the form actually used, underscores the difficulty of measuring the impact such resources have on scholarship and learning today. This time gap allows us to observe the various issues arising in the literature in different chronological periods mirroring thus a discussion accompanying digital research practices along its development. Claire Warwick et al. in their paper on “Library and Information resources and users of digital resources in the humanities” seem to agree with this observation as they note the same behaviour in a survey conducted by the Institute of Historical Research in 2005 and 2006. It is clear therefore that different issues arise in literature with regard to scholarly content use according to the community addressed and the authors’ aim.

The changes recorded following the diffusion of digital resources in a discipline such as History present great interest if viewed in different chronological phases. Two reports by the ITHAKA organization of 2006 and 2012 present, for example, different research activities and

---

<sup>10</sup>K. Jones and R. Bennett, “E-resources for research in the Humanities and Social sciences: A British Academic Policy Review”, *British Academy*, April 2005, pp. 1-116.

<sup>11</sup>L. Meho, H. Tibbo, “Modeling the Information-Seeking Behavior of Social Scientists: Ellis’s study revisited”, *Journal of the American Society for Information Science and Technology*, vol. 54, no. 6, 2003, pp. 570-87.

<sup>12</sup>Ibid, p. 586.

<sup>13</sup>M. Bulger et al., “Reinventing research? Information practices in the humanities”, *Information Practices in the Humanities, A Research Information Network Report*, April 2011, p. 72.

measure of engagement of historians with digital content and practices.<sup>14</sup> While the earlier report suggests slower adoption and production of electronic scholarship as well as a general hesitation in using digital content, this gradually alters in later reports with references to an enthusiastic engagement of historians with digital practices. However, even in 2012, this “discipline in transition”, as it is characterized, faces the lack of a satisfactory amount of digitally available primary sources. A more specialized approach is introduced by Agiatis Benardou, Panos Constantopoulos and Costis Dallas in their paper on working practices of research communities in the Humanities.<sup>15</sup> This article provides answers to questions related to digital content, its usability and privacy issues by focusing on the specific research community of Holocaust researchers as a study group. Enlightening the whole discussion by providing examples from a survey conducted for the purposes of the European Holocaust Research Infrastructure (EHRI) contributes in shaping conclusions in a particular context.

Information on scholarly content use is also enriched by the work done for the purposes of various European research projects and infrastructures. Particularly, a significant number of research infrastructures in the Humanities and Social sciences orientated their efforts in identifying current research patterns, developing tools and enlarging the family of digital resources. Focusing on this last aspect, the Council of European Social Science Data Archives (CESSDA) constitutes an example of such infrastructure which contributes in locating, accessing and sharing data for social sciences.<sup>16</sup> The content available through CESSDA is predominantly quantitative in nature, and while its portal does not contain information on how researchers use this content, its existence and apparent development indicates the increasing significance attributed to searching, analysing and disseminating digital data in this field.

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) on the other hand sets as its aim the long-term access and use of all European Arts and Humanities digital research data.<sup>17</sup> To this purpose various surveys have been conducted in an effort to document the changing research practices of humanities scholars and scholarly content use. In the context of DARIAH,

---

<sup>14</sup> Griffiths, Dawson and Rascoff, “Scholarly Communications”, pp. 1-14; Rutner and Schonfeld, “Changing Research Practices”, pp. 2-52.

<sup>15</sup> Benardou, Constantopoulos and Dallas, “An approach to analysing working practices of research communities in the Humanities”, *International Journal of Humanities and Arts Computing*, vol. 7, no. 1-2, Edinburgh University Press, 2013, pp. 105-127.

<sup>16</sup> Council of European Social Science Data Archives – CESSDA, <http://www.cessda.org/>, 2012, (accessed 10 June 2013).

<sup>17</sup> Digital Research Infrastructure for the Arts and Humanities – DARIAH, <http://www.dariah.eu/>, 2013, (accessed 5 June 2013).

various research programmes have been developed such as TextGrid and Isidore.<sup>18</sup> TextGrid constitutes a Virtual Research Environment which enables researchers in the text-based humanities to create, analyze, edit, store and publish their research data in one (user) environment while Isidore is a tool that collects, enriches and provides a unified data access for the digital humanities and social sciences. In addition, EASY is developed as an Electronic Archiving System to collect, archive and offer access to thousands of datasets in the humanities, social sciences and other disciplines.<sup>19</sup> All these initiatives underline the increasing predominance of digital research practices and digital resources.

This brief list of research projects related to scholarly content use could be further expanded with references to European infrastructures such as the European Holocaust Research Infrastructure (EHRI), the COncecting REpositories (CORE) project, the Collaborative European Digital Archive Infrastructure (CENDARI) and the Network for Digital Methods in the Arts and Humanities (NEDIMAH). Aiming to develop an open, collaborative research environment which will provide integrated online access to dispersed (archival) resources relating to the Holocaust, EHRI constitutes a major research project which takes under consideration various aspects of digital research in order to lay the foundation for building such an infrastructure.<sup>20</sup> Furthermore, the CORE project in conjunction with the ServiceCORE project as well as the Digging into Connected Repositories (DiggiCORE) project aim in general to facilitate free access to content stored across Open Access repositories and are thus considered as large aggregators of content.<sup>21</sup> CENDARI on the other hand is a more specialized initiative which aims to integrate digital archives and resources for research on medieval and modern European history while NEDIMAH includes a series of activities and networking events that will allow the examination of the practice of, and evidence for, digital research in the arts and humanities across Europe.<sup>22</sup>

Examining the range of initiatives held towards improving research conditions in a digital environment, the literature produced discussing this topic and particularly their publishing frequency in recent years, leads to the conclusion that increasingly more European, national and private institutions set digital research practices, and in particular scholarly content use, as an issue urging

---

<sup>18</sup> Virtual Research Environment for the Humanities – TextGrid, <http://www.textgrid.de/>, 2012, (accessed 30 May 2013); Service Integration, data interconnect research and education - Isidore, <http://www.rechercheisidore.fr/>, 2013, (accessed 14 April 2013).

<sup>19</sup> Electronic Archiving System of Data Archiving and Networked Services – EASY, <https://easy.dans.knaw.nl/ui/home>, 2013, (accessed 5 April 2013).

<sup>20</sup> European Holocaust Research Infrastructure – EHRI, <http://www.ehri-project.eu/>, 2013, (accessed 5 May 2013).

<sup>21</sup> Connected Repositories - CORE Projects, <http://core-project.kmi.open.ac.uk/projects>, 2011, (accessed 5 May 2013).

<sup>22</sup> Collaborative European Digital Archive Infrastructure – CENDARI, <http://www.cendari.eu/>, 2012, (accessed 25 April 2013); Network for Digital Methods in the Arts and Humanities- NEDIMAH, <http://www.nedimah.eu/>, 2012, (accessed 25 April 2013).

study and analysis. This report, based on bibliographic references and work conducted in the context of European research infrastructures will proceed in analysing user information behaviour towards content by distinguishing type and subject, based on Europeana's content analysis, as described before.

## 2.2 Documenting scholarly content use

The literature review, as presented before, revealed interesting user information behaviours, which vary according to the field and content. Aiming to document how researchers from different disciplines interact with different kinds of objects, literature will offer illustrative examples of researchers' behaviour with regard to content types. This process will be developed based on the content types identified in Europeana's collections.<sup>23</sup> In this way, specific user behaviours will be recorded in conjunction with specific content types, namely images, text, video and sound. This structure will ultimately lead to conclusions about the use of digital content by humanists and social scientists according to object type.

### 2.2.1 Images

This is the largest content type in Europeana's collections and includes a wide variety of subcategories such as art-historical artefacts, photographs, paintings, sculpture, maps, engravings etc.<sup>24</sup> The analysis here will focus on digital images while references to analogue material, such as printed images, will be made only by way of comparisons in order to grasp differences in scholarly use of digital content. However, a problematic issue arising is the fuzzy distinction of the type of sources as analogue or digital in the literature and surveys examined. For example, the term "images" is not usually clarified as analogue or electronic images and thus the percentages attributed to this type of content cannot be always safely interpreted. Effort will be made to minimize any statistical errors or at least to state them in this discussion.

According to a survey of academic researchers and postgraduate students conducted in 2001 in the United Kingdom, which recorded the research behaviour of 424 researchers in the Social Sciences and 373 in the Arts and Humanities, 10% of the social scientists and 36% of the humanists surveyed rated photographs and still images as essential sources in their research.<sup>25</sup> Their distinct

---

<sup>23</sup> Europeana, "Content Analysis", p. 4; Europeana, [http://www.europeana.eu/portal/usingeuropeana\\_results.html](http://www.europeana.eu/portal/usingeuropeana_results.html), 2008, (accessed 3 June 2013).

<sup>24</sup> These genres are presented in more detail in Europeana, "Content Analysis", pp. 22-23.

<sup>25</sup> This survey approaches Area Studies and Languages as a distinct discipline, examined separately from Arts and Humanities. Considering that Linguistics is included in the field of Humanities, at least for the purposes of Europeana Cloud, as presented in Deliverable 1.1 "Research communities identification and definition report", this introduces a bias to the sample. For this reason, Area Studies and Languages are included here in the percentages attributed to Humanities, see Education for Change Ltd, *Researchers' Use of Libraries*, p. 20; The percentages presented are always  $n=100$ .

behaviour towards photographs is further traced in the use of artefacts. More particularly, 1% of social scientists and 19% of humanists regard artefacts as essential sources while this gap is trivial in the use of maps and charts with percentages of 11% and 15% respectively. Therefore, the significance attributed to each type of image seems to depend highly on the nature of research in each discipline; researchers in the Arts and Humanities being keener users of this type of content in general.

Digital images in libraries and archives add an interesting dimension in this discussion. Several memory institutions aim nowadays to offer digital material to researchers in order to serve best their research audience. However, in cases in which digital content is not already provided by the institution, a significant number of researchers undertake to digitize the material relevant to their research by using digital cameras. The impact of this practice is characterized by Jennifer Rutner and Roger Schonfeld in 2012 as “perhaps the single most significant shift in research practices among historians”.<sup>26</sup> A problem arising from this practice, despite its widespread adoption, is the fact that “because the digital images are typically JPEGs, there is no metadata inherently associated with the file that relates it to the content of the image. Scholars rely on complex file structures and good memories to access their files once home from the archive”.<sup>27</sup> The ITHAKA Report of 2012 engages, among other issues, with researchers using digital cameras in these institutions and creating digital material from analogue sources.

### 2.2.2 Text

Text is the second largest content type in Europeana’s collections and includes a wide variety of specific object types, namely books and articles, manuscripts, newspapers, magazines and archives.<sup>28</sup> Taking as point of departure the 2001 UK survey mentioned above, aiming to illustrate how attitudes to digital practices and content have changed during the last decade, the percentages of researchers using digital text vary in the two fields.<sup>29</sup> While Europeana’s collections consist mainly of primary material, the survey here will widen the discussion on the form of text used by researchers by presenting percentages for both secondary as well as primary material.<sup>30</sup> Electronic journals and

---

<sup>26</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 11.

<sup>27</sup> Ibid, p. 12.

<sup>28</sup> Europeana, “Content Analysis”, p. 4; Europeana, [http://www.europeana.eu/portal/usingeuropeana\\_results.html](http://www.europeana.eu/portal/usingeuropeana_results.html), 2008, (accessed 3 June 2013).

<sup>29</sup> The UK Survey distinguishes ‘sources’ as primary and secondary material, see Appendix 5 of Education for Change Ltd, *Researchers’ Use of Libraries*, pp. 48-64.

<sup>30</sup> While taking into consideration the scholarly discussion on the terminology of sources, this Deliverable report adopts the term ‘primary’ for computerised data sets, rare books and manuscripts, maps and charts, photographs, artefacts, moving image or sound recordings, while ‘secondary’ material to describe journals, newspapers, books and electronic



other electronic publications for example are valued as essential secondary sources in the Social sciences with a percentage of 57% whereas in the Arts and Humanities it only reaches 23%.<sup>31</sup> Similar user behaviour is recorded for electronic full-text services with percentages of 56% and 25% respectively. Regarding electronic pre-print archives on the other hand, the percentages presented in the survey are lower for both disciplines: this source was rated essential by 25% of social scientists and 11% of arts and humanists. Looking at printed text, that is books and journals, a different picture emerged. Books were rated as an essential source by 92% of social scientists and 95% of humanists while the ratios for journals were 97% and 89% respectively. This observation suggests that while digital forms were increasingly prevailing in images, analogue forms of text were still dominant in both research communities.

A different aspect of this discussion is presented in the interviews conducted by John Houghton et al. in their article “Changing research practices in the digital information and communication environment” few years later, in 2003, regarding the sources rated as ‘essential’ and the extent of their use. It is evident that while some forms of digital text are not rated as essential sources, they are in contrast extensively used. For example, while “online newspapers and press releases were not seen as an ‘essential’ source”, they “were used by almost 80% of social sciences, humanities and arts researchers”.<sup>32</sup> Therefore, this suggests that the percentages presented in the surveys rating sources as ‘essential’ are only indicative of the actual extent of their use by researchers. This observation is questioned in the most recent article of 2012 “Supporting the changing research practices of historians” by Jennifer Rutner and Roger Schonfeld in which they argue that “while it is still the case that the majority of interviewees would seek a print copy of a relevant source, the use of digitized texts -books, book chapters and article- was ubiquitous”.<sup>33</sup> Interestingly, scholarly content use and citation practices are developed independently in a measure as the form of sources used - digital or analogue - does not imply the form of source cited. In accordance with this notion, the Research Information Network Report of 2011 suggested that “many scholars seem to prefer to cite paper sources even when they have used an online version, largely to ensure a correct and durable record”.<sup>34</sup> Juxtaposing the main points of the literature produced during the last decade regarding the measure of engagement of researchers in the Social Sciences and in the Arts and Humanities with digital or analogue text, it can be observed that while in 2001 the rated

---

publications accordingly, as listed above in Appendix 5 of Education for Change Ltd. These terms were adopted based on the wider impression for defining sources in the literature review conducted.

<sup>31</sup> (n=100)

<sup>32</sup> Houghton, Steele and Henty, “Practices in the Digital Information”, p. 88.

<sup>33</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 17.

<sup>34</sup> Bulger et al., “Reinventing research”, p. 72.

essence of digital text was significantly less in comparison to analogue text the main focus nowadays is not on the form of text used than on the form of source cited. Therefore, while increasingly more researchers have already adopted, and continue to adopt, digital text in their work, their preference to cite the analogue form of the text could indicate a stronger reliance on the trustworthiness or scientific validity of printed forms of text-or is just another “analogue” practice soon to be transformed.

An interesting survey conducted by the British Academy Review in 2005 on “E-resources for research in the humanities and social sciences” discusses, among other issues, the availability of e-resources.<sup>35</sup> Listing journals, books/monographs and manuscripts as major categories of analogue sources used by humanists and social scientists, interviewees suggested that there is great lack of electronic forms of books and manuscripts with percentages reaching 78% and 76% respectively while by contrast the availability of journals reaches 55%.<sup>36</sup> As issues of availability and use of sources interact, greater availability of digital books or manuscripts, for example, could suggest greater extent of use from researchers. On the other hand, some of the problems that humanists and social scientists face, according to this survey, in accessing digital forms of text are restricted or inconvenient access, electronic records not online or accessible via Web, online versions not complete or insufficiently detailed and difficulties of using digital resources.<sup>37</sup>

A particular example of available digital text largely embraced by the research community is Google Books. Historians who were interviewed for the purposes of the ITHAKA Report of 2012 “widely acknowledged Google Books as a valuable tool for their work”.<sup>38</sup> Interestingly, Google Books are described as “a centrally important tool for accessing primary and secondary sources for research” for scholars focused on historical periods that are pre-1923.<sup>39</sup> The capabilities provided by digital text, in particular here Google Books, for full-text search for example create appealing conditions for researchers to adopt such material in their research. Several historians also referred to Google Books as a way to confirm quotes and citations during the writing and editing phases. Therefore, digital text has various uses and utilities. However, regarding Google Books, historians noted limitations of the available material in foreign languages.

---

<sup>35</sup> Jones and Bennett, “E-resources”, pp. 1-116.

<sup>36</sup> Ibid, pp. 39, 41.

<sup>37</sup> Jones and Bennett, “E-resources”, p. 42.

<sup>38</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 19.

<sup>39</sup> Ibid, p. 19.

### 2.2.3 Video and Sound

The object types of video and sound could be characterized as more recent additions to the palette of objects used by researchers in the Humanities and Social sciences. Their digital existence gradually emerged alongside the increasing presence of digital content. However, they still constitute minor categories of objects in research in these two fields, if compared with other content types, and for this reason they are usually handled together in literature and surveys. In the context of Europeana, sound and video occupy less than 1% of its content, a fact that reveals a substantial lack of such material among its collections. For these reasons, audiovisual material will be jointly presented here in order to avoid repetition of literature in which they are not discussed separately.

In 2001 in the United Kingdom only 5% of social scientists regarded video and sound as essential sources compared to 24% for humanists.<sup>40</sup> Despite the fact that this type of content was not considered as significant by both disciplines, it seems that it is increasingly perceived as a source for reaching innovative results. This observation emerges from a survey conducted by the British Academy in 2005 which addressed, among others, a question to humanists and social scientists on whether they use e-resources “to offer quite novel research possibilities”.<sup>41</sup> This question, which aimed to reveal the potentials attributed by researchers in digital resources, documented a lively/vivid interest towards sound and video material. For example, a researcher of Linguistics answered that “I use video-clips and animated short movies (to be played on a computer)...to elicit linguistic data from native speakers of Quechua. These tools make the elicitation of certain types of data much more efficient than without such tools”.<sup>42</sup>

“These types of artefacts have long been used as a source of content in history”<sup>43</sup>; with this notion Jennifer Rutner and Roger Schonfeld describe the relationship between historians and audiovisual material as primary sources in 2011. While specific numbers or examples of this content type are not given, it seems that these content types are highly valued among historians. For example, a researcher explained “how a mass digitization of government audio recordings and their availability in the public domain have shaped his career and his research”.<sup>44</sup> However, there are serious challenges which a researcher faces today when dealing with audiovisual material. More particularly, this article traces, among other, the lack of tools for capturing, presenting and citing such material while some scholars have expressed concerns regarding their scientific validity as

---

<sup>40</sup> This percentage corresponds to “moving images and/or sound recordings” in the survey, see Education for Change Ltd, *Researchers' Use of Libraries*, p. 20.

<sup>41</sup> Jones and Bennett, “E-resources”, p. 45.

<sup>42</sup> Ibid, pp. 45-46.

<sup>43</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 14.

<sup>44</sup> Ibid, p. 14.

artefacts within the academia. The answers given by some PhD students interviewed by Rutner and Schonfeld enlighten the statement above as they suggested that “they struggle technologically and methodologically to locate, capture, analyze and report on a variety of source types including audio, video, oral histories etc.”<sup>45</sup>

### 2.3 Conclusion

This first attempt to document and understand the way researchers in the humanities and social sciences approach and interact with content through literature led to some preliminary results. Some overall observations will be presented here in an effort to summarize the points traced in the literature review as well as to provide answers to the questions made earlier in the introduction. An observation that could be attributed in all types of content and corresponds to the question how digital content changes research practices and if this has an impact on the final outcome of a researcher’s work is that the availability of digital sources to the public has contributed in reaching content more easily and thus in producing research work faster. It is generally admitted that digital content has a significant impact in the traditional speed of the research practices. This observation is illustrated by the following experience of a researcher who described that “I was also able to do very broad searches that would have taken years of actually digging through the newspapers to find obscure references to [my topic]. So that is where I think I first started to use digital sources as a genuine research tool, rather than as a teaching tool”.<sup>46</sup>

Various user information behaviours were recorded regarding each content type, namely image, text, video and sound. Interestingly, ‘image’ could be described as the material concentrating more diverse behaviours, different for each discipline of the humanities and social sciences. In this way, the rated essence of this material type in surveys, for example, is presented to be higher for humanists than for social scientists. However, a common statement recorded is the need for metadata accompanying each digital image, otherwise the information is incomplete or even useless for researchers. Text is similarly highly valued for both fields while the popularity of its digital genres varies according to research needs. More specifically, e-journals seem to constitute an indispensable part of the research material gathered, described in 2009 as “the life-blood of UK research institutions”.<sup>47</sup> This is further illustrated with the following statement, “within four months, users at ten UK research institutions visited nearly 1400 ScienceDirect journals..., half a million times and

---

<sup>45</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 14.

<sup>46</sup> Ibid, p. 14.

<sup>47</sup> D. Nicholas et al., “E-journals: Their use, value and impact”, *Research Information Network*, London, 2009, p. 6.

viewed a million and a half pages” leading to the conclusion that “the number of titles and the number of article downloads have nearly doubled from 2001-2007 and 2003-2007, respectively”.<sup>48</sup> Particular emphasis in using e-journals is given to the abilities provided for full-text content. This is an issue particularly highlighted in user information behaviour studies while libraries and digital search engines, such as Google, increasingly invest in providing full-text content.<sup>49</sup> While the impact of this new service provided to researchers has not been studied in detail, it has been characterized as “a revolution in the scholarly communications landscape”.<sup>50</sup> This description could be accepted in terms of the enhanced abilities provided to researchers not only to trace digital content with key word searches but also to actually download and edit the material in favour of research purposes. In this way, digital text is traced faster, accessed easily and handled as real research material with all its dimensions provided as if using its analogue form. Therefore, it could be concluded that full-text services describe comprehensive digital libraries. Finally, in examining video and sound in literature and various research programmes an interesting dimension is added with regards to the new possibilities offered from these content types. While their popularity among researchers is far less than image and text, it is increasingly being adopted as a material to reach innovative results. This potential attributed to video and sound is illustrated in the recent effort of Europeana to enlarge its audiovisual collections as seen in the new research initiatives Europeana 1989 and Europeana 1914-1918.<sup>51</sup>

---

<sup>48</sup> Ibid, pp. 6, 14-15.

<sup>49</sup> Ibid, p. 10.

<sup>50</sup> Ibid, p. 10.

<sup>51</sup> These research programmes mostly focus in collecting interviews and other oral testimonies in order to record the social memory of these historical events and periods, see Europeana 1989, <http://www.europeana1989.eu/en/>, 2013, (accessed 5 May 2013); Europeana 1914-1918, <http://www.europeana1914-1918.eu/>, 2013, (accessed 5 May 2013).

---

### 3. Desk Research: Digital Research Practices State of the Art

---

Researchers and their digital research practices in the field of Humanities and Social Sciences are some of the core issues addressed in the Work Package 1 of Europeana Cloud, entitled “Assessing Researcher Needs in the Cloud and Ensuring Community Engagement”.<sup>52</sup> Understanding the principles driving research practices, with a particular focus to their digital aspect, is vital in achieving the goals set in the work package. In accordance with one of the objectives of Work Package 1, desk research was undertaken to understand, document and evaluate the current state of the art digital research practices. This milestone report, as part of Work Package 1 and outcome of Task 1.3.1, presents the results of this desk research and summarizes the relevant literature and research programmes identified in this context. The literature review will ultimately lead to an evaluation of the current research practices. In this way, gaps and barriers are aimed to be revealed, a process which will enable dealing with such issues in the future.

#### 3.1 State of the art: An overview of user information behaviour studies

The information behaviour of researchers has been a popular subject for scholarship since the 1950s. This is evident from the growing number of studies dedicated to this topic.<sup>53</sup> The critical perspectives of information behaviour studies have shifted several times since the beginning of the field in order to capture the changing nature of research practices and provide an overview of the various aspects defining and affecting the information behaviour of scholars. With regard to the emergence of the digital information practices, two key trends can be identified in the critical literature, trends which form a chronological and thematic sequence.

At the end of the twentieth century, an increasing number of scholarly papers referred to the widespread use of the Internet in research, thus making a distinction between analogue and web research practices.<sup>54</sup> Papers written only a few years before the turn of the century, describe research practices in the Humanities and Social Sciences that were fundamentally different to their present condition. An example of how the early practices can be expressed are found in Rebecca Watson-Boone’s 1994 article “The Information Needs and Habits of Humanities Scholars”, in which she observes that “humanists tend to work alone” while they “graze within texts and their colleagues’

---

<sup>52</sup> A common definition of the disciplines attributed to the Humanities and Social Sciences can be found in the Web Survey of WP1 of Europeana Cloud.

<sup>53</sup> Benardou, Constantopoulos and Dallas, “Analysing working practices”, p. 108.

<sup>54</sup> Bass et al., “The Information Behavior”, pp. 2-3.

minds rather than browse through collections, shelves or catalogs”.<sup>55</sup> A paper published by Mary Folster in 1989 provides an overview on social scientists’ information behaviour, and presents some of the most popular sources of information for the faculty as “locating citations, referring to one’s personal library, and consulting colleagues in one’s own department”, with no reference to any digital character of their research.<sup>56</sup>

Moving now from analogue research practices to the digital era and the mark it has made on research, the introduction of new technologies and their wide adoption by researchers initiated a discussion regarding their impact on scholarly practice and the measure to which scholarly primitives have been affected.<sup>57</sup> This discussion led to the emergence of new literature focusing on and exploring the changes that occurred, with an aim to record and understand the new form and dimensions of the research process. Apart from these perceptions traced in the information behaviour literature, various approaches constantly emerge inside the context of the digital scholarly information practice, characteristic of the growing nature of this field. A recent example is traced in the work of Donald Case, who suggests that the last thirty years witnessed a shift from studying “the use of channels and sources” to the “encountering and seeking of information and the interpretation of meaning from that information”, resulting in an expanding “scope of investigation to view the use and user preferences for digital tools and services”.<sup>58</sup> In order to document new research practices, the works presented above built on earlier literature. This report similarly builds on the wealth of papers discussing the changing digital research practices in an attempt to summarize their overall findings and to create an updated state of the art paper.

A significant variety of works and research programmes has been engaged with the introduction of new technologies in the Humanities and Social Sciences since the beginning of the 21st century. This literature has mostly focused on the influences of information and communication technologies on research practices.<sup>59</sup> The wide interest in the impact of the digital age on humanists and social scientists can be examined from the perspective of the traditional unpopularity of

---

<sup>55</sup> Ibid, p. 2; Rebecca Boone, “The Information Needs Needs and Habits of Humanities Scholars”, *JSTOR*, vol. 34, no. 2, Winter 1994, cited in Bass et al., “The Information Behaviour”, p. 2.

<sup>56</sup> Ibid, p. 3.

<sup>57</sup> The term ‘scholarly primitives’ was first introduced by Unsworth (2000) who used this term in order to describe ‘basic functions common to scholarly activity across disciplines’, J. Unsworth, “Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this?”, *Symposium on Humanities Computing: Formal Methods, Experimental Practice*, King’s College, London, 2000, <http://jefferson.village.virginia.edu/~jmu2m/Kings.5-00/primitives.html>, (accessed 11 September 2013).

<sup>58</sup> D. Case, *Looking for Information: A survey of research on information seeking, needs and behavior*, 3<sup>rd</sup> ed., Emerald Group Publishing, 2012, p. 279.

<sup>59</sup> J. Houghton, C. Steele and M. Henty, “Research practices and scholarly communication in the digital environment”, *Learned Publishing*, vol. 17, 2004, pp. 231-32.

technology in these fields. This notion results from the examination of analogue research practices of humanists and social scientists and their recorded hesitation in adopting new digital activities in their research process.<sup>60</sup> The digital age has resulted in a significant transformation of the research practices as a result of the increasing engagement of researchers with digital tools and content. Also, as William A. Wulf points out, “the humanities offer a new opportunity to explore how information technology can be employed in fundamentally different ways that will provide fresh insights and enrich research in other applications”.<sup>61</sup> Various opinions have been expressed regarding the reaction of the Humanities and Social Sciences to these changes. John Houghton, Colin Steele and Margaret Henty suggest that the two disciplines present different user behaviour and attitudes from the Sciences. This suggestion is further explained in terms of hesitation, of both disciplines, to adopt digital sources and practices, although “they have not ignored the technology”.<sup>62</sup> While, according to the authors, there is an apparent disciplinary divide in terms of communication, collaboration, networking and in the form of sources used, other studies describe humanities scholars in particular as increasingly enthusiastic users of digital resources.<sup>63</sup> The extent to which the Humanities and Social Sciences embrace digital research can be seen from how they perceive the changing research practices within the scholarly field. In this regard, an example concerning the overall impact electronic research methods is perceived to have on research quality (taken from a survey commissioned by the Research Support Libraries Group in the United Kingdom) is quite revealing.

---

<sup>60</sup> See for example the perceptions of digital content and practices within History in Griffiths, Dawson and Rascoff, “Scholarly Communications”, p. 10.

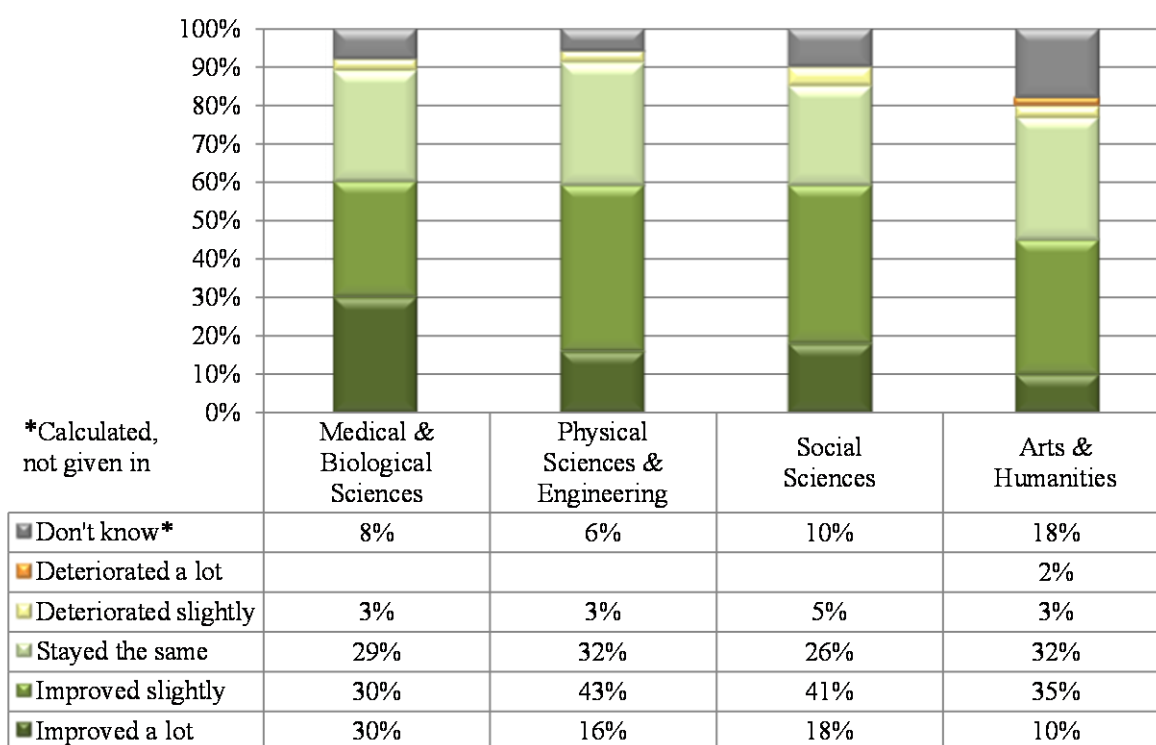
<sup>61</sup> Houghton, Steele and Henty, “Practices in the Digital Information”, p. 52.

<sup>62</sup> Houghton, Steele and Henty, “Practices in the Digital Information”, p. 51.

<sup>63</sup> Bulger et al., “Reinventing research?”, p. 10.



**Table 1: Overall impact on quality of research of electronic research methods**



Source: Education for Change Ltd, *Researchers' Use of Libraries*, p. 74, also available at <http://www.rslg.ac.uk/research/libuse/>, (accessed April 2013).

Examining Table 1 above, it seems that researchers in these four disciplines perceive and value differently the impact of digital practices on the quality of research. Firstly, the majority of respondents in the Medical and Biological Sciences favourably value the new research methods while 30% of their answers consider new technologies as having significantly improved research. Physical Sciences and Engineering as well as Social Sciences reach similar percentages in their answers which overall perceive these new methods as an improvement while the rates of the Arts and Humanities could be interpreted as slightly reluctant in agreeing with the previous answers.

The strong presence of the new digital research practices and their constant development is clear from the long list of research programmes which have emerged during the last decade successfully exploring this area, as introduced above in page 5. Of the wide variety of research programmes, this report only investigates user-centric infrastructures in more detail. Significant examples include the research programmes DARIAH and its preparatory project Preparing DARIAH as well as the European Holocaust Research Infrastructure (EHRI).<sup>64</sup> Operating in the context of Holocaust Studies, a deliverable report produced by the Digital Curation Unit in Athens has been concerned with documenting and understanding the research process of this particular community by

<sup>64</sup> For more information on these research programmes visit the websites DARIAH, <http://www.dariah.eu/> and EHRI, <http://www.ehri-project.eu/>

collecting data from literature review, from interviews with researchers and archivists, and by circulating questionnaire surveys.<sup>65</sup> Moreover, the Council of European Social Science Data Archives (CESSDA), the Network for Digital Methods in the Arts and Humanities (NeDiMAH), the Collaborative European Digital Archive Infrastructure (CENDARI), the research programmes CORE and DIGGICORE, the Common Languages Resources and Technology Infrastructure (CLARIN) and the Bamboo Digital Research Tools wiki (BambooDiRT) are some of the research programmes encountered in this literature review. With the exception of the final example, which is of American origin, they are all European. Relevant work done in the context of the user-centric research infrastructures will be further analysed below, in the discussion about the activities driving the research practice in the Humanities and Social Sciences.

The growing amount of literature, research programmes and infrastructures mirror the major emphasis on how the digital impacts on the Humanities and Social Sciences in relation to research practices, publication and collaboration. As technology evolves, digital research practices and literature related to this topic will advance in order to closely follow the new technologies adopted. A point made by Griffiths, Dawson and Rascoff perfectly captures this constant development; whereas humanities scholars, for example, were initially portrayed as hesitant users of digital technologies, “younger generations of scholars rely so heavily upon electronic resources that content available only in print has little visibility”.<sup>66</sup> The reason lying behind this statement could be the efficient residence of younger scholars in the digital environment which they perceive as a comfortable one for conducting their research. Unlike older researchers who have been researching for years in an analogue way, younger scholars approach their research digitally since its beginning as this has been their “traditional” way for research. An interesting example illustrating the above statement is also given in the ITHAKA Report of 2006 in which historians “worry that for undergraduates ‘if it isn’t electronic, it doesn’t exist’”.<sup>67</sup> As research practices constantly change, studies on their current state of the art are vital in order to discover the current needs of the researchers. An analysis of the scholarly activities will therefore follow, mapping the research process followed by humanists and social scientists and discuss the extent to which they use digital tools, content and practice, simultaneously identifying gaps and barriers to further enabling digital research in these disciplines.

---

<sup>65</sup> European Holocaust Research Infrastructure, D 16.4 “Researcher Practices and User Requirements”, 2013.

<sup>66</sup> Griffiths, Dawson and Rascoff, “Scholarly Communications”, p. 5.

<sup>67</sup> Ibid, p. 12.

## 3.2 Documenting digital research practices

The methodology applied in approaching, understanding and documenting digital research practices lies in a desk research conducted which has tried to trace and gather all relevant sources, papers, research programmes, web pages and blogs, discussing digital research in the fields of Humanities and Social Sciences. In order to best document the outcome of this desk research, it was considered useful to record the research practices by observing the research process as followed by a humanist and social scientist when seeking, collecting, analysing and disseminating information. Wandering through this timeline of scholarly activities will best document the variety of dimensions affecting digital research practices aiming to reveal, by the end of it, specific problems that users deal with. Before proceeding to the analysis of the research practices, as these were traced in the literature, it is vital to define here the scholarly activities driving the research process of humanists and social scientists. As it is suggested in the deliverable report 16.4 “Researcher Practices and User Requirements” of the European Holocaust Research Infrastructure, “if viewed at an appropriate level of abstraction, working practices even in as broad a domain as the Humanities”-Social Sciences could be also added here-“tend to involve a finite set of fundamental processes common across disciplines”; these fundamental processes have been defined and approached several times leading to various interpretations of the scholarly activities. For the purposes of this analysis, the work published by Palmer et al. in 2009 will be adopted here as it summarizes best, in the opinion of the writer, the fundamental, scholarly “primitives”, namely *searching, collecting, reading, writing and collaborating*, as defined for all disciplines in an online environment. These are further refined in a wider list of twenty more detailed scholarly activities which thus elaborate and imprint the whole research process. In the course of this present report, the term “primitives” will be used to denote high-level classifications of research activities, while, for the lower-level, more granular stages of the research process we will be using the terms “activity” or, later on, in the case of use of research tools, “method”.

At this point, however, it needs to be emphasized that influences from other literature on defining these activities, such as the Taxonomy of DH Research Activities and Objects of DARIAH-DE, will not be ignored.

### 3.2.1 Searching

This research primitive is defined as the first step in a research process including a range of activities for *seeking, browsing* and *accessing* information. Interestingly, the new trends

characterizing *searching* have radically transformed the research process in both disciplines. According to a survey conducted by ITHAKA S+R on the research practices of historians, the majority of the interviewees agreed that “Google is the first port of call!”.<sup>68</sup> It is worth acknowledging here that Google is a corporation that offers a range of technologies, not a technology in itself. However, it is used here, and by the research community as recorded in the literature review, as shorthand to denote primarily their search services. Their answers suggested that researchers use general Google searches to start the discovery process for primary, archival and secondary material as information on these sources is “nearly always available on the open web”.<sup>69</sup> Google discovery tools, for example the general Google search and Google Books, are valued positively in the Humanities and Social Sciences for their ability to trace “hidden” or hardly detectable sources. The valuable dimension of these search engines is also praised in regards to “orienting scholars to a new field by helping them identify sources and gain access to a network of citations”.<sup>70</sup> The widespread adoption of online finding aids - and in particular Google - is increasingly evident in the literature on new research practices.<sup>71</sup> However, while the majority of recent surveys present the Web as a significant resource and starting point for research, this fact does not premise the abandonment of analogue research methods. According to the Research Information Network Report of 2011, while many researchers “start their searches for new information with Google...many still consult printed texts, from their bookshelf or university library”.<sup>72</sup> This observation is also traced in another survey by Harley et al. in 2009 which suggests that despite the fact that Google is widespread among humanities scholars, “traditional methods, such as citation chaining (83%) and learning from peers and experts (95%) remain the most significant ways of finding resources”.<sup>73</sup> This reinforces the opinion that “technology is not a substitute; it is a supplement”.<sup>74</sup>

Aiming to clarify and distinguish the measure of engagement of scholars with analogue and digital research practices, several surveys have been conducted which present the various ways of discovering and accessing information. Approaching this topic in an interdisciplinary way will allow broader observations and comparisons among disciplines.

---

<sup>68</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 18.

<sup>69</sup> Ibid, p. 18.

<sup>70</sup> Ibid, p. 19.

<sup>71</sup> For example, the survey conducted by the Digital Curation Unit, Athens for the European Holocaust Research Infrastructure Deliverable 16.4 on the use of online services rates Google as the most frequently used online service.

<sup>72</sup> Bulger et al., “Reinventing research?”, p. 70.

<sup>73</sup> Ibid, p. 68.

<sup>74</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 18.

For this reason, a sample of the survey by the Research Support Libraries Group in the United Kingdom in 2002, as introduced in Table 1, was chosen to further illustrate here the suggestion for mixed methods when seeking information. Table 2 presents the answers gathered regarding the discovery methods used and rated as essential by researchers in various disciplines during their research process. While the focus of this report is the digital research practices in the Humanities and Social Sciences, references to other disciplines as well as to analogue discovery methods are presented in an effort to grasp differences and the impact of the digital era on scholarly activities.

**Table 2: Essential access and discovery methods for percentage of survey**

	Medical & Biological Sciences	Physical Sciences & Engineering	Social Sciences	Arts & Humanities
Consult printed books or journals	91%	90%	93%	97%
Use inter library loans and/or document delivery	79%	65%	77%	71%
Search catalogues (online or card)	58%	66%	75%	76%
Use other library technology (photocopying, scanning, printing out etc)	62%	50%	58%	69%
Access online electronic books or journal articles	80%	69%	61%	32%
Browse through collections of books or other material	15%	24%	46%	66%
Consult primary materials	16%	16%	24%	65%
Use enquiry & research assistance	17%	15%	36%	32%
Consult surrogates?	3%	2%	10%	41%

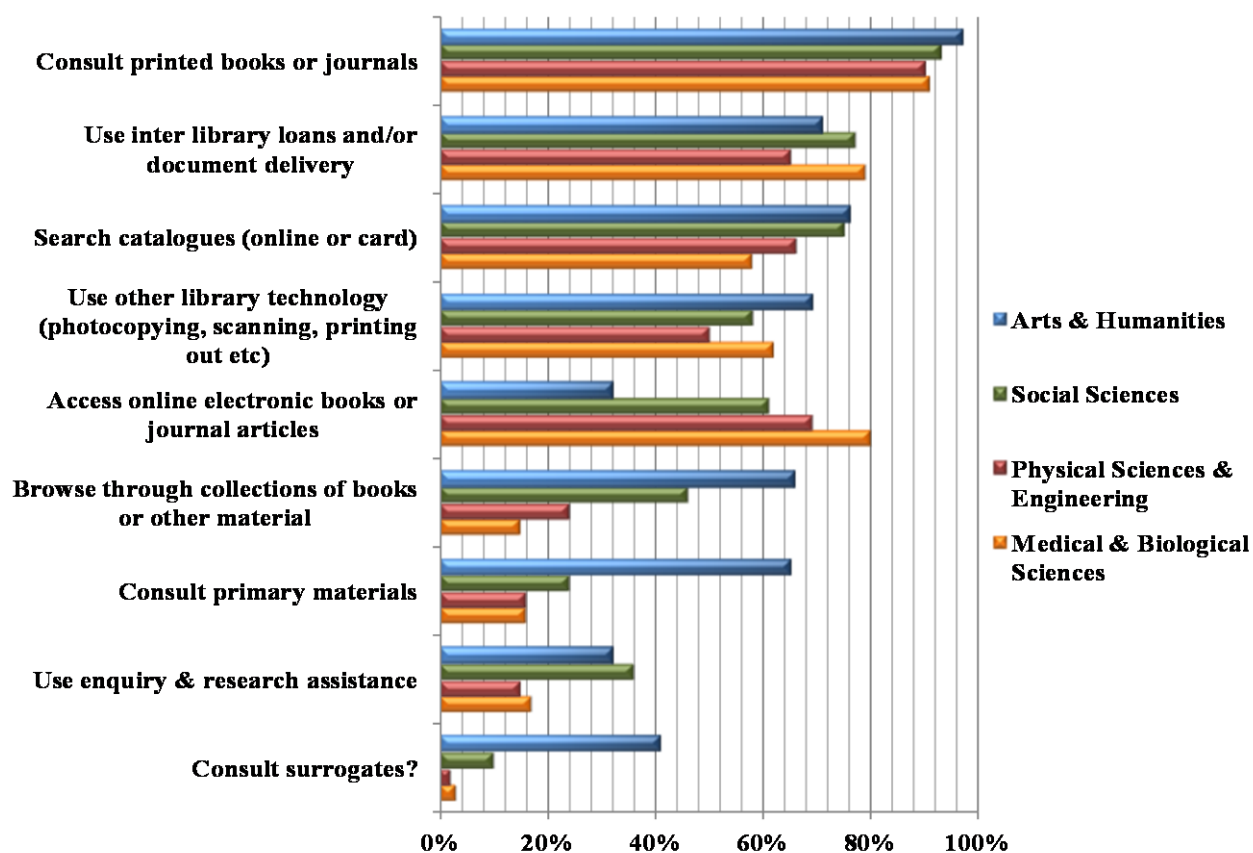
Source: Education for Change Ltd, *Researchers' Use of Libraries*, p. 24, also available <http://www.rslg.ac.uk/research/libuse/>, (accessed April 2013).

A quick view of Table 2 reveals significant contrasts in the rates following each discipline and discovery method. Observing, for example, printed books and journals on the one hand and online access to electronic books and journal articles on the other hand, it seems that humanists base their initial searches mostly on analogue form of sources, with a percentage of 96% in comparison to only 30% of the digital form. This gap is significantly smaller in the Social Sciences and almost insignificant in the Medical and Biological Sciences as researchers in these disciplines tend to rely

more on the digital form of sources, while maintaining the dominant position of printed books and journals in their research. A visual mapping of these percentages is presented in Figure 1 aiming to illustrate Table 2.

Another interesting observation drawn from table 2 refers to the use of libraries and archives when seeking information. According to the fourth listed method, a quite significant percentage of respondents in all disciplines uses library technology to print, scan or photocopy material. This activity most probably derives from a need to create a personal archive, either analogue or digital, during a research trip in a memory institution for later review and analysis. This trend, particularly for digitizing archival and research material, is also apparent in the ITHAKA Report on historians of 2012 which suggests that “researchers spend the majority of their time in archives informally digitizing materials” recognizing at the same time that “the widespread use of digital cameras and other scanning equipment to capture sources materials is perhaps the single most significant shift in research practices among historians”.<sup>75</sup>

**Figure 1: Essential access and discovery methods for percentage of surveys**



<sup>75</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 11.

Focusing on humanists, some suggest that “using archives meant costly travel and scheduling time to spend in the archives”, indicating cost, travel and time as significant barriers to their earlier research.<sup>76</sup> Apart from that, the challenge of tools suitable for digitizing material by scholars in an archival environment is also stated in the literature. More particularly, “the displacement of the intellectual engagement with the material appears to have some downsides, given the lack of software to facilitate the process of capturing and using digital photographs for scholars” while another challenge emerges from the integration of images with their textual notes, a process that adds another layer of format types.<sup>77</sup> (Digital research practices in an archival setting have been examined in detail for the purposes of the EHRI Deliverable report by DCU in 2012.)

While the whole range of literature agrees on the benefit of online information seeking for all disciplines, various differentiations are also recorded. Firstly, surveys conducted at national and international level present different rates of engagement with digital content and practices of scholars in different continents. In 2002, the findings of Education for change et al. demonstrated that web search engines were rated as ‘very important’ by 45% of researchers in the United Kingdom. In the same year, Friedlander showed that 80% of the faculty and students in her U.S.-based sample searched for information for their research online.<sup>78</sup> Disciplinary variations are also evident in the literature. Regarding electronic resources, for example, surveys showed that social scientists are more enthusiastic users of electronic resources, with 37% answering that they use such resources all or most of the time in comparison to 25% of humanists.<sup>79</sup> While these rates may be shaped by the varying perception of new technologies by each discipline, the issue of the digitized available content should not be acknowledged here. As Meho and Tibbo noted “the Internet has added a new dimension to the information-seeking behaviour of scholars...the problem, however, is that only a small fraction of available information is digitized and made accessible to researchers”.<sup>80</sup> In respects of content accessed, quite revealing is Table 3 which presents the sources rated as ‘essential’ by researchers in the fields of Social Sciences, Arts and Humanities as well as Physical Sciences and Engineering.

---

<sup>76</sup> Bulger et al, “Reinventing Research?”, p. 26.

<sup>77</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 12.

<sup>78</sup> Houghton, Steele and Henty, “Practices in the Digital Information”, p. 48.

<sup>79</sup> Ibid, p. 47.

<sup>80</sup> Meho and Tibbo, “Modeling the Information-Seeking Behavior”, p. 586.

**Table 3: Percentage rating sources ‘essential’ (Sample from U.K. RAE Census)**

	Medicine & Biological Sciences	Physical Sciences & Engineering	Social Sciences	Arts & Humanities
Printed refereed Journals	98%	96%	97%	89%
Books	60%	77%	92%	95%
Bibliographic tools, abstracting & indexing services	82%	75%	67%	62%
Electronic journals & other electronic publications	73%	62%	57%	23%
Electronic full text services	75%	57%	56%	25%
Other printed nonrefereed journals	19%	21%	38%	48%
Electronic pre-print archives	44%	45%	25%	11%
Computerised data sets	31%	28%	27%	13%
Photographs & still images	20%	16%	10%	36%
Newspapers	2%	2%	31%	29%
Rare books & MSS	3%	4%	9%	53%
Microfilm/microfiche	2%	3%	10%	39%
Moving images and / or sound recordings	9%	8%	5%	24%
Maps & charts	4%	7%	11%	15%
Artefacts	1%	4%	1%	19%

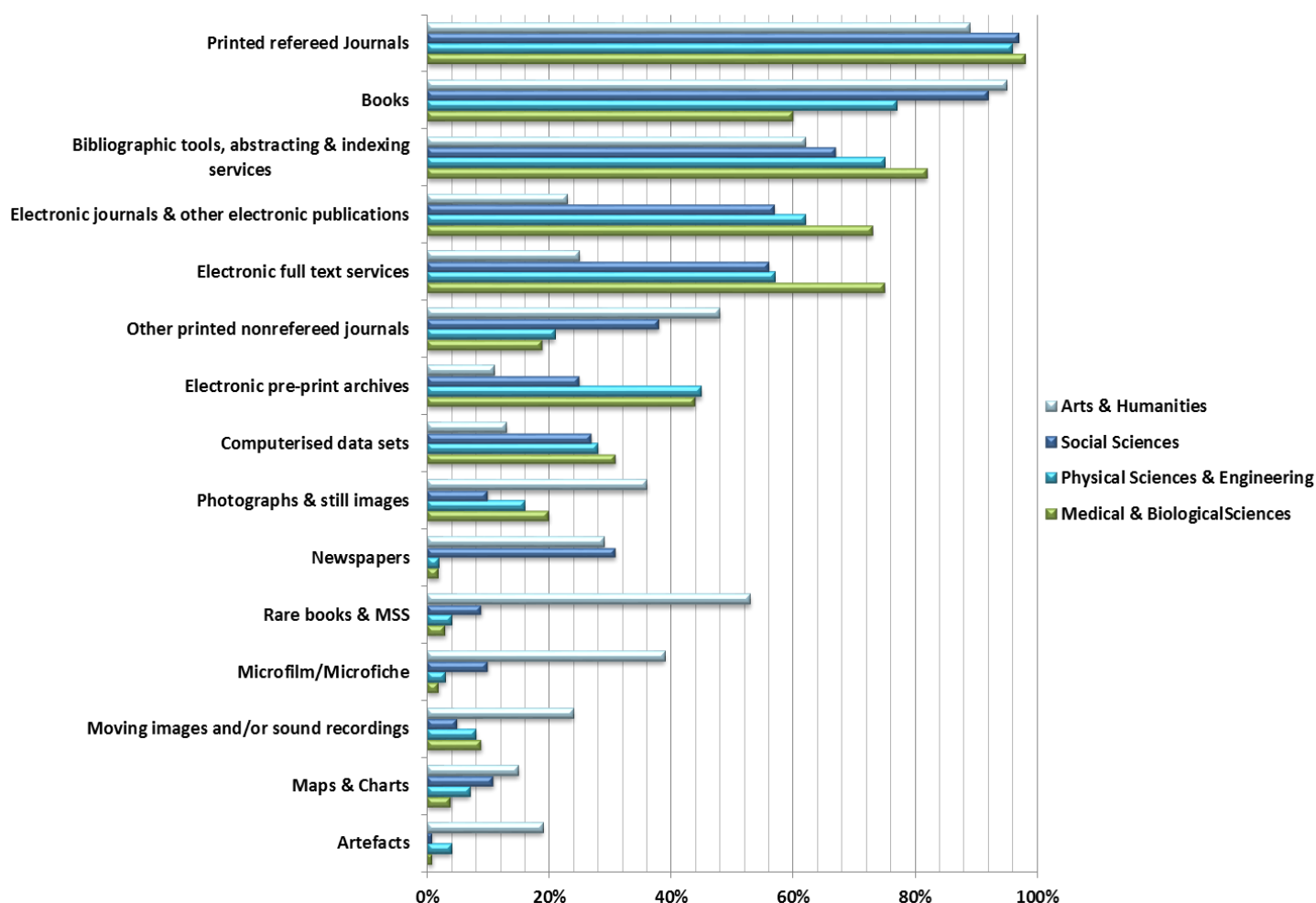
Source: Education for Change Ltd, *Researchers' Use of Libraries*, p. 20, also available <http://www.rslg.ac.uk/research/libuse/>, (accessed September 2013).

Examining Table 3 above, it seems that the analogue form of sources, for example books and printed journals, is still considered by all three disciplines as extremely essential while the rates of the digital form of text mark sharp contrasts to the printed form as well as among the disciplines, as is also illustrated in Figure 2 below. Electronic journals and books are rated quite lower to printed books and journals in Physical Sciences and Engineering, Social Sciences and Arts and Humanities; however, humanists seem to rely even less on the digital content than researchers in the other two fields with a minor percentage of around 20%. Regarding e-journals, it is interesting to add here that another aspect that differentiates humanities from the sciences is their information seeking behavior



and use of e-journals. More particularly, the Research Information Network Report of 2009 reported that historians, for example, “are more likely to access e-journals via Google, and to use search tools, especially menus, once they are inside the publisher’s platform”, a practice which is not widespread among Life Sciences.<sup>81</sup> To sum up, the overall findings here indicate that the analogue form of sources is still dominant in all three disciplines presented in Table 3 with a growing presence, on the other hand, of the digital content. In terms of wider engagement with the digital form of sources, Medical and Biological Sciences would be ranked first, followed by the Physical Sciences and Engineering, Social Sciences and last by the Arts and Humanities. It could be said, though, that “scholars engage with a range of resources and technologies, moving seamlessly between print and manuscript and digital resources”.<sup>82</sup>

**Figure 2: Sources rated as ‘essential’**



<sup>81</sup> Nicholas et al., “E-journals”, p. 25.

<sup>82</sup> Bulger et al., “Reinventing research?”, p. 72.

The web character that is currently attributed to information seeking behaviour was not adopted without question. Some typical behaviours recorded in the literature present researchers from both disciplines to have “actively dismissed Internet search engines and other web-based resources such as Wikipedia and electronic-only journals as unreliable sources of information”.<sup>83</sup> This behaviour most possibly lies not in the proven scientific invalidity of digital research practices than in an overall hesitation in adopting new practices. The discussion surrounding the scientific validity of resources is also mirrored in the citation practices as discussed before in chapter 2.<sup>84</sup> On the other hand, other researchers expressed their concerns as to whether this increased access to information resources has the equivalent impact in enabling them to ask and answer new questions or if the work conducted is similar to the pre-digital era though faster.<sup>85</sup> The potentials provided to the researchers by the use of digital search engines and sources as genuine research tools is met with disputable reactions as the researcher nowadays has a huge amount of information readily available, a fact which can be perceived either positive or negative. In general, transformations in practice, following the adoption of new technologies in research, were mainly described in terms of ease and speed of access to information as well as the possibilities given to conduct research with keyword search.<sup>86</sup>

### 3.2.2 Collecting

The second research primitive, according to Palmer, is *collecting*. This term includes the activities of *gathering* and *organizing* information for the purposes of a research topic. Following searching, browsing and accessing information, the range of sources traced, as presented in table 3, is stored in an effort to create personal collections that will lead, by rereading and elaborating, to results. This stage of research is closely interconnected with the form of sources gathered which vary as illustrated before.<sup>87</sup> In this way, “humanities scholars in particular, have been shown to cultivate and take pride in their personal collections of books and other print sources”, which are considered as a necessity since “rereading is a significant part of their interpretive work”.<sup>88</sup> Primary research materials in the humanities are collected amongst a wide range of heterogeneous sources in the form of texts, images or artefacts, a fact that requires the development of the relevant tools and services in

---

<sup>83</sup> Bass et al., “The Information Behavior of Scholars”, p. 5.

<sup>84</sup> See p. 16.

<sup>85</sup> Bulger et al., “Reinventing research?”, p. 28.

<sup>86</sup> Ibid, p. 28.

<sup>87</sup> Regarding the form of sources used by researchers see Table 3, p. 32.

<sup>88</sup> Palmer, Tefteau and Pirmann, “Scholarly Information Practices”, p. 17.

order to handle and store them.<sup>89</sup> Personal collecting and archiving of research information is described as a necessity, which is further developed in the context of digital resources, tools and services. Despite the initial reliance of humanists on collections of analogue sources, researchers in all disciplines have expressed positive feelings for the possibilities provided by new technologies to gather large quantities of digital data and have this information mobile on their laptops.<sup>90</sup> Echoing the interviewees in a survey conducted by Bass, Fairlee, Fox and Sullivan on the *Information Behaviour of Scholars in the Humanities and Social Sciences*, researchers “use some form of electronic filing system to keep track of their information”.<sup>91</sup> This is particularly described as “saving to disk locally or on a shared network, storing websites as favorites, and using the EndNote database”.<sup>92</sup> In this way, analogue record of sources gathered seems to rapidly decline among the list of research practices of both faculties. As scholarly activities interact and influence each other, “the specification of the structure and affordances of databases, and the procedures established for information management, rely heavily on established information seeking, annotation, versioning and knowledge making practices”.<sup>93</sup>

Researchers in the Humanities and Social Sciences are presented in the literature to manage and store information mainly on their desktops and laptops while, by contrast, they do not engage “with more advanced tools for text-mining, grid or cloud computing, or the semantic web”.<sup>94</sup> Aiming to specify the primitive of *collecting* for humanists and social scientists, the report chooses to focus here on a case study regarding the use of Old Bailey Online by researchers.<sup>95</sup> According to this sample of humanists and social scientists, in order to organize their sources and data “most of them maintain some sort of mini-database on their personal computers. For some, this database is simply a Microsoft Word file or collection of Word files. Others report maintaining lists of relevant trials using Microsoft Excel, Zotero (an online resource manager), or keeping a folder of each trial. One respondent received special permission from the developers of the Old Bailey Online to download the entire database and re-programme parts of it to meet his research needs”.<sup>96</sup> This survey suggests that researchers engage with a spectrum of approaches and technology use, which is increasingly growing and digitally orientated.

---

<sup>89</sup> S. Angelis et al., “Classical philology, text-based and Holocaust studies facing digital research infrastructures: from practice to requirements”, p. 3.

<sup>90</sup> Palmer, Tefteau and Pirmann, “Scholarly Information Practices”, p. 17.

<sup>91</sup> Bass et al., “The Information Behavior of Scholars”, p. 7.

<sup>92</sup> Ibid, p. 7.

<sup>93</sup> Angelis et al., “Classical philology”, p. 3.

<sup>94</sup> Bulger et al., “Reinventing Research?”, p. 7.

<sup>95</sup> Old Bailey Online is a database containing criminal proceedings of London’s central criminal court from 1674-1913 see *ibid*, p. 23.

<sup>96</sup> Ibid, p. 24.

Organizing information is presented in the literature alongside the creation of personal collections of data. The emerging personal digital libraries that researchers create face the challenge of managing analogue as well as digital research notes and sources while, according to the ITHAKA report, “digital systems do not appear to address all the needs of even those scholars who seek to use them”.<sup>97</sup> However, is this statement justified enough in the literature? In an effort to trace the most frequently used tools to organize sources by researchers, following the example presented before, it seems that Microsoft Word and Excel occupy the highest point in this list followed by databases and other digital tools. Regarding digital images for example, a historian said that his personal way of collecting and organizing such sources is firstly by recording them in order in paper notes and then in Excel spreadsheets in which he tries to fill information about the images.<sup>98</sup> On the other hand, while databases constitute a quite widespread way of seeking and tracing sources among the Humanities and Social Sciences, these are valued differently in regards to organizing information. Examining the answers of historians, in the context of the ITHAKA report, “while some interviewees reported using a database to organize and access their research materials, the operating system’s file search functions seemed to supersede this practice”.<sup>99</sup> It could be concluded therefore that the significance of organizing the research material lies in its ability to make each text, image or audio more easily detectable among the large amount of sources gathered. Focusing again in the same survey conducted by ITHAKA on humanities scholars, and more particularly on historians, they reported using a number of different approaches to organize content, mostly folder-based, where topic or author were the dominant criteria.<sup>100</sup> Another interesting practice of humanists when organizing content from a book, for example, is their strong reliance in the structure of the book which characterizes the whole act of organizing sources and notes. In general, scholars in both disciplines are increasingly “amassing incredible personal libraries of digitized material, alongside the content they are producing as part of the research process (notes or writings)”.<sup>101</sup>

### 3.2.3 Reading and Writing

The scholarly primitives of *reading* and *writing*, which include the lower-level activities of *scanning*, *assessing*, *rereading* and *assembling*, *co-authoring*, *disseminating* accordingly, could be considered as the essence of the analysis of the sources traced, gathered and organized during the

---

<sup>97</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 25.

<sup>98</sup> Ibid, p. 23.

<sup>99</sup> Ibid, p. 26.

<sup>100</sup> Ibid, p. 23.

<sup>101</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 26.

previous stages of a research process. If viewed in regards to digital research practices, these two activities can be interpreted as one stage of research which can be entitled as *analysing data*. This new primitive combines all the distinct activities of *reading* and *writing* while it could also include the activities of *annotating* and *translating* as these are considered significant- and in some cases necessary- steps in the analysis of the sources and the production of a new and innovative piece of work. Initiating the discussion on *reading*, it could be said that the form of sources gathered-digital or analogue - predefines in some measure the form of the information to be read. According to the survey presented by Houghton, Steele and Henty, almost 90% of reading and studying is conducted using printed documents.<sup>102</sup> As this sample of answers covers the whole spectrum of disciplines, it is interesting to specify here that the percentage of social scientists and humanists reading and studying the sources online was at least slightly higher than in the other sciences. Brockman et al. seem to agree with the first statement as they also noted that “scholars read in different media, but rarely read extended texts directly from a computer screen”.<sup>103</sup> Various reasons underlie the practice of reading sources in printed form; exploring literature and the above interdisciplinary survey in order to illuminate the process “few respondents mentioned the difficulty of reading from a screen”.<sup>104</sup> However, the majority of the answers given connect reading with annotating as the interviewees “mentioned the desire to place stickers and/or white notes on what they were reading, or the inclination to read at home over the weekend”.<sup>105</sup> These answers suggest, therefore, that the barrier to reading online might relate more to annotating and portability than to problems emerging from screen-based reading.

Expanding further the above suggestion, note-taking can have a significant impact on the way in which researchers approach their sources and work. Closely connected with the primitive of *reading*, annotating has been gradually transformed into a digital research practice adopting various methods that co-exist with traditional, analogue ways of note-taking. An example of this suggestion is given in the Research Information Network Report on humanists in 2011 in which interviewees described three ways of annotating; “to annotate the texts, most copy and paste relevant passages into Microsoft Word and then use the comment and highlight function. A few describe handwriting notes while reading the passages on their computer screens. Others report toggling between Microsoft Word and the files they are reading, taking notes while reading the texts online”.<sup>106</sup> These three

---

<sup>102</sup> Houghton, Steele and Henty, “Practices in the Digital Information”, p. 92.

<sup>103</sup> W.Brockman et al., *Scholarly work in the humanities and the evolving information environment*, CLIR, 2001, p. 7, also available <http://www.clir.org/pubs/reports/pub104/pub104.pdf>, (accessed September 2013).

<sup>104</sup> Houghton, Steele and Henty, “Practices in the Digital Information”, p. 92.

<sup>105</sup> *Ibid*, p. 92.

<sup>106</sup> Bulger et al., “Reinventing Research?”, p. 25.

methods are also evident elsewhere in the literature while a new method is added regarding the use of databases and other software tools for note-taking.<sup>107</sup> In general, the majority of researchers seem to combine paper and digital notes which are often imposed in lengthy processes for re-writing and organizing them.

An interesting aspect in terms of analysing multilingual sources is illustrated in the ITHAKA Report of 2012. A number of interviewees in that survey shared that “while they use Google Books during the writing and editing phases of a project to confirm quotes and citations” this is difficult for researchers working on corpus of foreign language material as its availability is limited.<sup>108</sup> For this reason, “many continue to rely on subscriptions databases which provide access to collections of foreign-language materials in these cases”.<sup>109</sup> The dominance of digital sources in English forms a significant barrier for researchers willing to conduct their research in other languages. For this reason, humanists and social scientists adopt various translation tools to help them when dealing with multilingual sources. However, a gap is noted in the literature regarding this issue as only minor references were traced in the bibliography explored.

### 3.2.4 Collaborating

Collaboration among researchers is characterized as a key change in humanities research and is widely discussed throughout the literature. Recent papers and reports seem to emphasize the enhancement and growth of more formal and systematic collaboration between researchers as a result of the new technology.<sup>110</sup> Interestingly, Houghton, Steele and Henty refer firstly to the communication among scientists when listing the fundamental changes affecting the practice of research.<sup>111</sup> Again, while research projects in the social sciences are described as mainly collaborative, there exists an impression that research in the humanities, in the contrary, is a lonely process. This impression is increasingly abandoned as various joint authorships, research programmes and networks emerge. Networking has also been encouraged in the social sciences, as a result of the Internet. According to participants in the study of Meho and Tibbo on the information-seeking behaviour of social scientists, “the easy access that the Internet allows to, and between, individuals and groups all over the world plays a significant role in further enhancing the importance of networking for research purposes among academic social scientists”.<sup>112</sup> Therefore, the

---

<sup>107</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 26.

<sup>108</sup> Ibid, p. 19.

<sup>109</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 19.

<sup>110</sup> Bulger et al., “Reinventing research?”, p. 7.

<sup>111</sup> Houghton, Steele and Henty, “Practices in the Digital Information”, p. 41.

<sup>112</sup> Meho and Tibbo, “Modelling the Information-Seeking Behavior”, pp. 13-14.

introduction of new technologies has in general facilitated collaborative research, contributed to richer communication and has thus promoted interdisciplinary research, researcher productivity and quality.

The innovative possibilities offered by the digital age in research are mirrored in the geographic scope of collaboration. More particularly, as Walsh and Maloney suggested in their work in 2002, “computer network use was associated with more geographically dispersed collaborations as well as more productive collaborations”.<sup>113</sup> Focusing here on the geographic aspect, enabling collaborations despite location and geographic placement of the researcher has naturally expanded the number of opportunities provided for communication in a research field and has therefore enabled the creation of richer networks. Since then, networking is further cultivated in the various research programmes presented before, such as the European Holocaust Research Infrastructure or the Network for Digital Methods in the Arts and Humanities, which explore a thematic field of research collaboratively from different geographic horizons. The emphasis given in this aspect by humanists and social scientists can be traced by the number of tools increasingly adopted in favour of a more efficient collaboration. For example, Bulger et al. in their article in 2011 noted that “remote collaboration is usually accomplished through e-mail and phone conversations, with drafts and supplementary documents shared via e-mail and, in rare cases, using online dropboxes”.<sup>114</sup> However, the tools used in research for this purpose vary according to the needs of each research team. The case studies presented in this paper of 2011 are indicative of this suggestion; “for the Birmingham English Department, along with the full range of simple collaboration tools such as e-mail and Skype, scholars are using RSS feeds and Twitter for current awareness... Likewise, the UCL Philosophy department, which is less likely to engage with collaborative computational information uses, nevertheless relies on resources such as blogs and newsfeeds to engage with other researchers”.<sup>115</sup> It could be therefore concluded that while both humanists and social scientists increasingly engage with collaboration and collaborative tools, their specialised needs may require accordingly specialised tools which will be easily accepted and embraced by the community as well as of maximum benefit to them.

Following this last suggestion on the best possible approach when creating collaborating tools for each research team, Griffiths, Dawson and Rascoff, in the same spirit of networking, make a different comment on the best approach to content. More particularly, they suggest that “the content would not have to reside on one server, but it would have to be ‘networkable’ and ‘sharable’ to be of

---

<sup>113</sup> J. Walsh and N. Maloney, “Computer network use, collaboration structures and productivity”, *Distributed work*, 2002, cited in Houghton, Steele and Henty, “Practices in the Digital Information”, p. 44.

<sup>114</sup> Bulger et al., “Reinventing research?”, p. 26.

<sup>115</sup> *Ibid.*, p. 62.

maximal use to the community”.<sup>116</sup> Moving on opposite direction from the suggestion made before on tools, which urges for more specialised collaborating tools, this suggestion here calls for a ‘sharable’ content for every researcher. Towards this end, various organisations are jointly working aiming to create a pool of dispersed content for all researchers from which content can be easily traced, retrieved, added and collaboratively used for the purposes of other research projects. In some cases, knowledge of collaboration tools is considered as a different expertise which a researcher does not possess. For example, the ITHAKA Report of 2012 observes that “the historian typically holds content expertise, while collaborators are likely to have expertise on the particular technology tool or method that is being applied to the research. In some cases, larger teams may collaborate together, with a variety of experts supporting work”.<sup>117</sup>

Scholarly communication is promoted and conducted through various channels that deal either with sharing content or knowledge expertise. All these forms of collaboration are, interestingly, closely interconnected with issues of security and trust. As was explicitly noted at the kick-off meeting of Europeana Cloud promoting security and trust will have a positive impact on collaboration among researchers who, according to the views expressed there, need quality assurance. Therefore, barriers are clearly traced and stated which, if raised, will strengthen and further promote digital networking of researchers. On the other hand, the views expressed in the literature explored for the purposes of this deliverable did not contain references on security issues, but rather they described “collaborative work as a positive experience”.<sup>118</sup> In both cases, however, there is a vivid interest for raising all possible barriers and creating the best possible conditions for the expansion of networks connecting researchers in dispersed geographic locations as well as for bringing together dispersed sources.

---

<sup>116</sup> Griffiths, Dawson and Rascoff, “Scholarly Communications”, p. 11.

<sup>117</sup> Rutner and Schonfeld, “Changing Research Practices”, p. 32.

<sup>118</sup> Bulger et al., “Reinventing research?”, p. 26.



---

## 4. Digital Tools State of the Art

---

The remainder of this report presents the results of desk research that aims to establish a theoretical basis for the more empirically focused case-studies in Task 1.3.5.

Scholarly work based on the use of digital content, tools and methods (often referred to as the ‘Digital Humanities’) is often described and categorized using methods. A comprehensive listing of methods can be found on [www.arts-humanities.net](http://www.arts-humanities.net).<sup>119</sup> Although the idea of methods is a very important concept – the application of these methods are reliant on digital tools. Better tools therefore, in theory, should lead to better research output. Very often these digital tools are developed in the context of a project which means that they are often developed with very specific project-specific applications in mind.

Digital tools aid researchers by:

- Facilitating and enhancing existing research, by making research processes easier via the use of computational tools and methods
- Enabling research that would otherwise be impossible: addressing research questions that would have been impossible to resolve without digital methods and tools
- Allowing them to ask new research questions i.e. questions that are driven by insights that were only achievable through the use of new tools and methods

### 4.1 What is a digital tool?

A glance at the literature regarding digital tools shows that the idea of what constitutes a ‘digital tool’ is not very well defined. One possible reason for this is that digital tools come in all shapes and sizes and are often outputs of specialised research projects. In some cases, the term ‘digital tool’ is used to refer to generic publishing platforms such as WordPress that aid the researcher with the dissemination of their work.<sup>120</sup> However, the majority of the tools listed as digital tools are used in the process of extracting new knowledge from digital content. Some tools, such as citation software etc. sit at the boundary and have implications for resources such as Europeana since the ability to cite digital objects consistently is an important issue for scholarly research.

One definition of a digital tool has been proposed by Hughes<sup>121</sup>:

---

<sup>119</sup> “Arts-humanities.net: Guide to Digital Humanities & Arts,” (accessed July 19, 2013), [www.arts-humanities.net](http://www.arts-humanities.net).

<sup>120</sup> <http://wordpress.com/>, (accessed 19 July 2013).

<sup>121</sup> L. Hughes, “Using ICT methods and tools in arts and humanities research”, in L. Hughes (ed.), *Evaluating and measuring the value, use and impact of digital collections*, Facet 2012.

---

*"I would suggest that a digital tool is any piece of software that can be used to gather, analyse and/or process data. A useful distinction can be drawn between tools that enable existing (i.e. analogue) research processes to be conducted better and/or faster, and tools that enable researchers to ask, and answer, completely new research questions. As digital objects are rarely only processed with just one tool (and should be re-useable and re-purposable in the future), interoperability and in/output interfaces are crucial for evaluating tools, whichever category they fall into"*

The use of digital tools is a critical component of the scholarly lifecycle since these tools allow the researcher to perform the tasks that lead to the creation of new knowledge and to answer new research questions based on digital content. In addition, a demonstration of a particular tool is often the way researchers are introduced to new methods.

## **4.2 Early digital tools**

The earliest Digital Humanities tools were primarily used to interact with electronic texts; they were often integrated into computer based learning packages that had very specific tasks. In 1994 the CTI Project at Oxford University Computing Services produced a 'Resources Guide' that listed the resources available at that time. It might not be a surprise that many of these packages listed in the Guide do not exist anymore. This is not very surprising since many of the learning packages that these tools were built for have also disappeared. However, it seems that philosophically we have come full circle since the Europeana Cloud project is looking at building tools based upon a specific (albeit much bigger) platform. Nevertheless, this is not the full story. Some of the legacy packages listed in the 1994 guide has been rendered obsolete since their functionality has been included in everyday software packages such as word processors. For example, the need for the Oxford Concordance Program tool for building concordances has been negated by the concordance building function of programs such as Microsoft Word. The contribution of Digital Humanities community to the development of standards such as SGML and later XML is not to be understated. Again, the latest versions of Microsoft Word allow users to export documents directly into an XML format.

Digital tools continued to be developed to accommodate an increasing number of methods. This work was done mainly through discrete projects hence there was a tendency to build project-specialised (hence highly customized) tools rather than generic tools (which can be easily re-used

and improved). This phenomenon still exists and it is generally the case that a specialized tools will give you a more useful tool but at the expense of usefulness to the rest of the community.

### 4.3 The role of tools in the scholarly lifecycle

In 2000 John Unsworth introduced a set of Scholarly Primitives as a set of “basic functions common to scholarly activity across disciplines...”.<sup>122</sup> In his 2003 paper he gave examples of specific tools that allowed researchers to perform these scholarly functions.<sup>123</sup> It is interesting to note that after 10 years many of these tools do not exist anymore (unless the URLs for them have changed). Here is the full list from the paper:

- **Discovering**<sup>124</sup>  
[Olive Software](#)  
*NITLE search engine*
- **Annotating**  
[Inote](#)  
*Imarkup*
- **Comparing**  
*IBabble*  
*The Versioning Machine*  
*The Virtual Lightbox*
- **Referring**  
*Scout Portal Toolkit*  
[Historical Event Markup Language](#)  
[Omnigator](#)  
*WebBrain*
- **Sampling**  
*TAPoR tools\**

---

<sup>122</sup> J. Unsworth, “Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?” (presented at the Humanities Computing: formal methods, experimental practice, King’s College, London, 2000), <http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html>.

<sup>123</sup> J. Unsworth, “Tool-Time, or ‘Haven’t We Been Here Already?’” (presented at the Transforming Disciplines: The Humanities and Computer Science, Washington, DC, 2003), <http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html>.

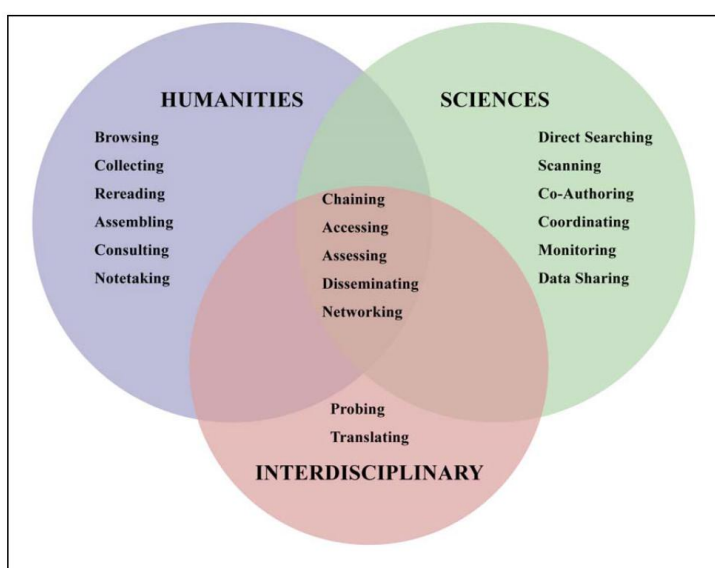
<sup>124</sup> Note: Tools in italics did not have a live URL link at time of writing this report. \* TAPoR exists at different URL and is cited in this report.

[TextArc](#)

*Graphviz and Ivanhoe*

A 2009 OCLC report by Palmer *et al* mentioned in Milestone 1.2 and 1.4 of the WP1 work package took the same approach and identified primitives most relevant to the disciplines of the humanities and sciences.<sup>125</sup> They also identified interdisciplinary primitives that were relevant to both (see Figure 3 below).

**Figure 3 - Scholarly primitives associated with disciplinary approach**



Source: Palmer et al.<sup>126</sup>

**4.4 Digital tools in the context of scholarly primitives**

Below is an attempt to establish links between the primitives and the tools that enable those primitives to be enacted in the real world. Detailed definitions of the primitives can be found in the OCLC report.<sup>127</sup>

Primitive	Functions Required	Comments and examples of tools
-----------	--------------------	--------------------------------

<sup>125</sup> Palmer et al, *Scholarly Information Practices*, <http://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf>, (accessed July 2013).

<sup>126</sup> Palmer et al, *Scholarly Information Practices*.

<sup>127</sup> Ibid.

<p><b>Searching</b></p> <p>1.1 Direct searching</p> <p>1.2 Chaining</p> <p>1.3 Browsing</p> <p>1.4 Probing</p> <p>1.5 Accessing</p>	<p>Full text indexing</p> <p>Interlinking between objects</p> <p>Ability to cite objects consistently</p> <p>Ability to browse and make serendipitous discoveries</p> <p>Ability to give access to the content (e.g. full text versions)</p>	<p>Google Scholar</p> <p>TextGrid</p> <p>Web of Knowledge</p> <p>Europeana Portal</p>
<p><b>Collecting</b></p> <p>2.1 Gathering</p> <p>2.2 Organizing</p>	<p>Ability to collect and store citations in one place</p> <p>Ability to collect and store content in one place</p> <p>Ability to store and organize a personal collection (online and offline)</p>	<p>Zotero, Endnote, BibTex</p>
<p><b>Reading</b></p> <p>3.1 Scanning</p> <p>3.2 Assessing</p> <p>3.3 Rereading</p>		
<p><b>Writing</b></p> <p>4.1 Assembling</p> <p>4.2 Co-authoring</p> <p>4.3 Disseminating</p>		<p>Dissemination software such as Omeka<sup>128</sup> CMS (with plug-ins such as Neatline<sup>129</sup>)</p> <p>Social Media</p>
<p><b>Collaborating</b></p> <p>5.1 Coordinating</p> <p>5.2 Networking</p> <p>5.3 Consulting</p>		<p>Social media</p>
<p><b>Cross-cutting Primitives</b></p>		

<sup>128</sup> <http://omeka.org/>

<sup>129</sup> <http://neatline.org/plugins/neatline/>

6.1 Monitoring		Web of Knowledge <sup>130</sup> (for scholarly output)
6.2 Note taking	Annotation of digital content	Many tools available (see below)
6.3 Translating		Could Europeana play a role in translating between domains?
6.4 Data Practices		If we expand this to include processing of content this is where most of the digital tools reside since they aid researchers to extract new knowledge from content/data (and in some cases - metadata). These tools are the focus for much of the remainder of this report.

It is obvious that there is a strong link between methods and tools. Categorising the tools by methods aids researchers to identify the most suitable tool(s) for their project/research question. This is already done through websites such as arts-humanities.net and more recently the Bamboo DiRT registry.<sup>131</sup>

#### 4.5 State of the Art Tools

In 2003 Unsworth considered many of the questions surrounding the development and use of digital tools.<sup>132</sup> He identified the potential of using web services to avoid researchers the hassle of installing software and dealing with hardware issues by requiring them only to use a web browser.

<sup>130</sup> <http://wokinfo.com/>

<sup>131</sup> "Bamboo Dirt", <http://dirt.projectbamboo.org/>, (accessed September 2013).

<sup>132</sup> Unsworth, "Tool-Time".

In 2013 this is finally becoming a reality through projects such as The DHW project.<sup>133</sup> Other projects that have taken similar approach are the Salami Project<sup>134</sup>, the Monk Project<sup>135</sup> and SEASR - The Software Environment for the Advancement of Scholarly Research.<sup>136</sup> Tools and infrastructure in the Digital Humanities was discussed as part of the *The Cologne Dialogue on Digital Humanities 2012*.<sup>137</sup> Here, Anderson et al proposed the idea of a ‘Digital Ecosystem’:

“...digital ecosystems are derived from communities rather than technologies. As open systems, digital research ecosystems will rely on communities and community involvement in a scenario where anyone can participate. The digital ecosystem is not for the specialist few but is instead about increased participation, sharing and building a social network of people, things, content and so on.”<sup>138</sup>

While van Zundert questions the benefits of large scale digital infrastructures<sup>139</sup>:

“The necessary generalizations and standardizations, management and development processes that large infrastructures need to apply to cater to wholesale humanities are at odds with well-known aspects of innovation. Moreover, such generalizations close off many possibilities for exploring new modeling and computing approaches. I argue that methodological innovation and advancing the modeling of humanities data and heuristics is better served by flexible small-scale research focused development practices.”

#### 4.6 Tools in relation to Europeana Cloud

Currently Europeana only holds metadata about digital objects. However, this project (Europeana Cloud) plans to ingest content (in addition to its metadata) into a new cloud infrastructure. This is an important development since most digital tools are designed to work with content as opposed to metadata and ingesting content will make it easier for existing tools to be utilised in the Cloud.

A previous study of Europeana content confirmed that metadata for the following broad categories of content are present in the Portal<sup>140</sup>:

<sup>133</sup> A. Bia, “The Digital Humanities Workbench,” in *Proceedings of the 13th International Conference on Interacci&ocute;n Persona-Ordenador*, INTERACCION ’12 (New York, NY, USA: ACM, 2012), pp. 50:1–50:4.

<sup>134</sup> <http://ddmal.music.mcgill.ca/research/salami>

<sup>135</sup> <http://monkproject.org/>

<sup>136</sup> <http://www.seasr.org/>

<sup>137</sup> M. Thaller (ed.), *Historical Social Research* Vol. 37 (2012), no. 3.

<sup>138</sup> *Ibid*; S. Anderson, p. 147.

<sup>139</sup> *Ibid*; J. Zundert, “If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities”, *Historical Social Research*, vol. 37, no. 3, (2012), p. 165.

1. **Texts:** books, letters, archival papers, dissertations, poems, newspaper articles, facsimiles, manuscripts and music scores
2. **Images:** paintings, drawings, prints, photographs, pictures of museum objects, maps, graphic designs, plans and musical notation
3. **Video:** films, news broadcasts and television programmes
4. **Sound:** music and spoken word from cylinders, tapes, discs and radio broadcasts
5. **3D:** virtual 3D representations of objects, architecture or places

#### 4.6.1 Text

Alongside images, text is typically the most common type of material present in cultural heritage collections and this is reflected in the number of digital tools that are available to aid researchers with the study of text. These range from small standalone (freeware and commercial) applications to applications that run within a large software environment such as SEASR (The Software Environment for the Advancement of Scholarly Research). AntWordProfiler is an example of a standalone freeware application that does not require extensive technical expertise (apart from an understanding of text profiling).<sup>141</sup> It enables a researcher "...to generate detailed statistics of the inserted text, while using the File Viewer and Editor Tool you [the researcher] are able to view and highlight different levels of vocabulary".

A number of commercial applications offer the ability to analyse text for semantic features. One example, called Diction describes the ability to "...that uses a series of dictionaries to search a passage for five semantic features—Activity, Optimism, Certainty, Realism and Commonality—as well as thirty-five sub-features".<sup>142</sup> Visualising and comparing different versions or translations of the same text can be very useful for researchers. Example of this is the Translation Array Tool and Visualizing Version Variation (prototype tool) developed at the University of Swansea.<sup>143</sup> This allows researchers to "... obtain an overview of the relationships of similarity between parallel segments in different versions of Shakespeare's Othello".<sup>144</sup> Tools that only require a web-browser are accessible to users with very little technical expertise. Voyant-tools.org is an example of such a tool.<sup>145</sup> It provides a simple analysis of a text that is uploaded or is accessible via an URL. More

---

<sup>140</sup> "Results in Europeana", [http://www.europeana.eu/portal/usingeuropeana\\_results.html](http://www.europeana.eu/portal/usingeuropeana_results.html), (accessed July 2013).

<sup>141</sup> [http://www.antlab.sci.waseda.ac.jp/antwordprofiler\\_index.html#sthash.6PikGkch.dpuf](http://www.antlab.sci.waseda.ac.jp/antwordprofiler_index.html#sthash.6PikGkch.dpuf)

<sup>142</sup> <http://www.dictionsoftware.com/>

<sup>143</sup> <http://delightedbeauty.org/vvv>

<sup>144</sup> Z. Geng et al, "Visual analysis of segment variation of German translations of Shakespeare's Othello Information Visualization", 1473871613495845, first published in July 2013.

<sup>145</sup> <http://voyant-tools.org/>



sophisticated analysis based on XML-TEI encoded texts can be performed using the TXM Beta an Unicode-XML-TEI text/corpus analysis platform.<sup>146</sup> An online implementation of this platform has been constructed by *Centre National de la Recherche Scientifique*.<sup>147</sup> Analysis of topics in a corpora of text is also of interest to researchers. MACHine Learning for Language Toolkit (MALLET) is a tool hosted by the University of Massachusetts Amherst.<sup>148</sup> The Lincoln Logarithms project based at The Emory Center for Digital Scholarship (ECDS)<sup>149</sup> used this toolkit to analyse the contents of sermon texts and describes the functionality as follows: "...MALLET shatters a group of texts into their individual words and then uses probability statistics to identify groups of words that frequently occur together. These groups of co-occurring words might indicate the presence of a theme or topic across the corpus. Its primary output includes lists of the top words for each topic group and a table that specifies the topical percentages in each document". In certain areas of research researchers rely on the ability to transcribe text to make them available for analysis. A number of transcription tools are registered on tools registries, one such example is T-Pen.<sup>150</sup> These tools are used by Humanities scholars to aid this type of work.

Alongside specific tools, portals or frameworks are being established that allow researchers, in the words of the SEASR Project "to design, build, and share software applications that support research and collaboration". The TAPoR project takes the same approach and allows tools developers to make their tools available to a wider audience via web services.<sup>151</sup>

## CASE STUDY #1 – TEXTUAL CONTENT

**Dr Tom Cheesman, Reader in German, Swansea University**

### 1. Can you suggest a good digital tool for use with text?

[delightedbeauty.org/vvv](http://delightedbeauty.org/vvv) ('Version Variation Visualization') - very much an experimental prototype, but points towards interesting possibilities for working with a particular kind of corpus: sets of versions of a work such as differing editions and translations. Looking ahead we want to be able to unravel intertextual relations and explore how reversioning is/was performed in different contexts, with different purposes.

### 2. What kind of things could you do with metadata?

<sup>146</sup> <http://wiki.tei-c.org/index.php/TXM>

<sup>147</sup> <http://portal.textometrie.org/demo/?locale=en>

<sup>148</sup> <http://mallet.cs.umass.edu/>

<sup>149</sup> <http://digitalscholarship.emory.edu/>

<sup>150</sup> <http://t-pen.org/TPEN/>

<sup>151</sup> <http://tada.mcmaster.ca/Main/TAPoR>

You can't do much without metadata. Provenance is just a sine qua non. Intertextual metadata interests me more: it points to contexts, and contexts are everything in studying texts. For our kind of corpora, FRBR-type metadata (e.g. 'this text is a re-use of a version which was also re-used [at these dates], [by these publishers], [in these types of publications], [in these theatres]' / 'this is a unique MS item' / 'this text was composed by redaction of text Y and text Z, 50% each') could eventually be produced by algorithmic comparative readings and checked against specialists' 'ground truths' (assuming rich and well organised provenance metadata too). Also, alignments ('this sentence or phrase is a translation or redaction of that other sentence or phrase') are metadata we need to train machines to create.

**3. What could you do with current Europeana Metadata? (for an example record see <http://www.europeana.eu/portal/record/9200119/2528EDB78EBA1DD87CB64A5623E8CBB4A9B465BD.html>)**

This is basic provenance metadata + a very basic "type" assigned to the image. I can't tell whether Wellcome has a run of *The Builder* or this is a loose page cut out of it, or whether it's part of a collection of pictures of schools or of the work of that architect or in fact anything about the context it was torn from. Also I don't get a link to any other images of the subject (or similar subjects) nor indeed to the NLW item.\* NLW provides much more 'formal' metadata (more work done on 'type', 'subject') but, 'description' is used in a confusing way (print subtitle followed by digitisation statement) and there is less provenance, probably because the item lacks the 'intrinsic context' ('*The Builder*', vol./no.) which would be my starting point for 'reading' the image in a context.

**4. What could you do with full-text content?**

I'm a literary critic and linguist, and my job title is "Reader in German". For me, full-text content is what it's all about: I read. When big data tools find interesting texts among the 'great unread', I want to read them or hear from someone who has. Algorithmic reading tools should enable people like me to read more, better. Including 'reading' languages I can't read, and comparing more versions of something than can remember.

## 4.6.2 Images

When dealing with images the ability to annotate is powerful and there are a number of tools that offer this functionality. The 'Annotation' category is one of the most populated categories in the Bamboo Dirt tools registry with 79 entries as of 10 October 2013. The DM project is an example of an environment for the study and annotation of images and texts.<sup>152</sup> Users are able to create image and text 'targets' for digital images and text. These targets can be a polygon of a single image or an annotation that refers to whole corpora of text as a set of images. Beyond annotation, computational algorithms can be used to analyse images. Although this technology has come a long way Terras in a recent book chapter states that applications are “...are generally bespoke, requiring specific application of techniques to suit the individual images and research question”.<sup>153</sup>

### CASE STUDY #2 – IMAGES

**Dorothy Porter, Curator of Digital Research Services, University of Pennsylvania**

#### **1. Can you suggest a digital tool for working with images?**

My favourite tool for working with images is the UVic Image Markup Tool<sup>154</sup>. I have used it for several different projects over the years.

#### **2. What does it allow you to do?**

It's an image annotation tool. Basically, you load in an image, put a bounding box around some part of the image, and then apply an annotation to it. You can create different classes/types of annotations, and then change the appearance of the bounding boxes depending on the type of annotation. You can also create annotations that are actually transcriptions (that is, the tool differentiates between transcriptions and pure editorially created annotations), and they are treated differently in the back-end (the back-end being TEI encoding). Once you have created your annotated image, you can generate a webpage that displays your image and annotations. I really like that it's based on a simple format of TEI, because it means I can do things like take a project that was

<sup>152</sup> <http://ada.drew.edu/dmproject/>

<sup>153</sup> M. Terras, (2012) “Image Processing and Digital Humanities” in M. Terras, J. Nyhan, and C. Warwick (eds.), *Digital Humanities in Practice*, Facet 2012.

<sup>154</sup> [http://tapor.uvic.ca/~mholmes/image\\_markup](http://tapor.uvic.ca/~mholmes/image_markup)

originally built in ColdFusion, export the data in spreadsheets, convert the spreadsheets to XML, translate the XML to the UVIC-TEI format, and then load that into the IMT and generate a new website that looks and functions much like the old one but depends on open, well-formatted data (I'm doing this at the moment).

### 3. What could you do with Europeana metadata? (for an example record see

[http://www.europeana.eu/portal/record/9200105/wellcomeimages\\_org\\_record\\_V0014532.html](http://www.europeana.eu/portal/record/9200105/wellcomeimages_org_record_V0014532.html))

Well, you could load in images and then create annotations on that image from the metadata about that image. Given how the tool works it would make most sense if you could either apply annotations to the whole image (which you can, in the IMT), or if relevant you could also select those areas of an image which the rich metadata described and apply annotations that way. However the first approach (applying annotations to the entire image) could probably be automated while the second approach would need to be manual (unless your rich metadata already included coordinates for bounding boxes).

#### 4.6.3 Video and Sound

The field of video and sound cover an extremely broad range of possible applications. Fields such as dialectology and phonetics, film studies, musicology, and media studies all rely on these types of sources. Transcription of sound and video is one area where digital tools can aid researchers working with these types of content. Tools such as HyperTRANSCRIBE do not conduct automatic transcription; however, they do aid the task of transcription.<sup>155</sup>

The analysis of music files is another field relevant to this category of material. The SALAMI (Structural Analysis of Large Amounts of Music Information) is a good example of the way a shared environment (such as the cloud) allows analysis of large data sets.<sup>156</sup> One of the tools in this project is creates bubble diagrams allowing labelling for formal analysis of music. The project describes the advantages of this approach:

*“To date, musical analysis has been conducted by individuals and on a small scale. Our computational approach, combined with the huge volume of data now available from such source as the Internet Archive, will: a) deliver a very substantive corpus of musical analyses*

<sup>155</sup> <http://www.researchware.com/products/hypertranscribe>

<sup>156</sup> <http://ddmal.music.mcgill.ca/research/salami/background>

*in a common framework for use by music scholars, students and beyond; and, b) establish a methodology and tooling which will enable others to add to this in the future and to broaden the application of the techniques we establish.”*

#### 4.6.4 Spatial Technologies (includes 3D)

Geographic information systems (GIS) have started to make inroads into humanities and social science research with archaeologists, in particular, making extensive use of the technology. One example of a tool that can be used to create virtual 3D representations of objects, architecture or places is QueryArch3D.<sup>157</sup> QueryArch3D is a project based at the University of New Mexico and is described as an “online, searchable repository ... that brings together GIS data, 3D architectural models, and virtual environments for teaching and research on ancient architecture and landscapes”. GIS is also a major component of ORBIS: The Stanford Geospatial Network Model of the Roman World.<sup>158</sup>

#### 4.7 Visualisation and Europeana Content Analysis

The use of visualisation either to analyse or present results is a theme that straddles all content types and deserves a mention in this report. Although visualisation is itself a method many visualisation tools are available. A basic example is the PaperMachines<sup>159</sup> plug-in for Zotero whereas Circos<sup>160</sup> is an example of a more advanced tool. PaperMachines was used by the Lincoln Logarithms Project to analyse the content of sermons.

A good example of where visualisation tools could be used in the current (metadata only) Europeana corpus would be to analyse the topics currently present in metadata records. Visualisation of the different topics or subject headings present in the current Europeana portal could be done using a tool such as Circos.<sup>161</sup> Network analysis could also be used to examine the relationships between different records/objects in Europeana.

#### 4.8 Conclusions on research into digital tools

- **The limitations of desk research** - During the composition of this section of the report it became clear that tools very often do not feature prominently in research papers due to the fact

---

<sup>157</sup> <http://mayaarch3d.unm.edu/queryarch3d.php>

<sup>158</sup> <http://orbis.stanford.edu/>

<sup>159</sup> <https://github.com/chrisjr/papermachines>

<sup>160</sup> <http://circos.ca/>

<sup>161</sup> Ibid.

that they are a means to an end for the researcher. Their evaluation and selection is part of the research process and is not reported in detail in papers. Another difficulty with evaluating digital tools by means of desk research is the fact that the latest tool is not synonymous with the state of the art (i.e. the newest tools are not necessarily the best tools). Meaningful evaluation of the tools mentioned in this report can only be achieved by empirical studies such as case studies. Two case studies have been included for the most common content types (namely text and images) in an attempt to partly address some of these concerns.

- **Offline vs. Online** – There is a steady shift towards tools that can be accessed centrally via a web browser (be that in a cloud environment or on a central institutional server) in a framework/environment where tools developers can plug-in to via APIs. Online working is fundamental to the concept of a Cloud concept however; there can be advantages in using standalone installations which means that consideration should be given to whether records/objects/data can be downloaded from Europeana Cloud to enable this kind of work. A standalone tool has advantages (e.g. more control over hardware and availability) as well as obvious disadvantages (issues with hardware etc.).
- **Creating new tools vs. facilitating the use of existing tools** – One of the aims of this project is to build tools. However, consideration should be given as to how existing tools (which users are already familiar with) could be accommodated. For example, TAPoR allows 3<sup>rd</sup> party tools developers to make their tools available to a wider audience via web services. The project needs to think about how it can “tap in” to existing communities. Governance, legal and economic context will affect the way users engage (or fail to engage) with Europeana Cloud. The previously mentioned concept of a ‘Digital Ecosystem’ may be useful guide in this context.
- **Content and Data vs. Metadata** - It was stated at the beginning of this section of the report that tools are a means to apply a method. A large proportion of methods are predicated on researchers having access to the content in addition to its metadata. If researchers are to apply existing tools to Europeana Cloud it will require content as well as metadata.
- **The richness of metadata** - The Europeana Data Model (EDM) and other schemas used in Europeana means that metadata is mapped from more expressive schemas resulting in metadata that is not as rich. A contributing institution often has to make compromises when mapping their data to the EDM hence potentially excluding useful data. This in turn limits what type of tools and methods that you can apply to the metadata and content in question.
- **The role of Europeana Cloud in the scholarly lifecycle?** – Answering this question will help the project to decide what tools to build.

- **Combinations of tools** - It is very often the case that research questions require the application of more than one tool. If it is to enable better research and engage researchers using multiple tools on one dataset/content/collection needs to be easy. There is great power in using tools in combination. This is illustrated very clearly in the Lincoln Logarithms Project.<sup>162</sup>
- **'Build it and they will come'** - Tools are evaluated and selected by researchers to achieve a specific goal therefore a generic 'build it and they will come' approach risks creating tools that are not useful to anyone. This was expressed clearly in 2005 at a Summit on Digital Tools for the Humanities<sup>163</sup>:

*"We closed the discussion by affirming, once again, that we should build for particular applications and users but also with consideration for the agreed-upon standards for interfacing software tools."*

Van Zundert recently reaffirmed the alternative approach to big standardised infrastructures<sup>164</sup>:

*"...modeling highly specific distributed web services is a more promising avenue for sustainability of highly heterogeneous humanities digital data than standards enforcement and current encoding practices."*

This project needs to ensure that it mitigates the risk of building tools that are too generic to be useful.

---

<sup>162</sup> <http://disc.library.emory.edu/lincoln/>

<sup>163</sup> Summit on Digital Tools for the Humanities - Report on Summit Accomplishments (University of Virginia, September 28, 2005), <http://www.iath.virginia.edu/dtsummit/SummitText.pdf>.

<sup>164</sup> Zundert, "If You Build It, Will We Come?", p. 165.

## Conclusion

---

The rise of digital content and its gradual adoption from researchers in the humanities and social sciences has inevitably brought changes in the research practices and user behaviour. Despite the fact that user information behaviour literature, aiming to grasp the impact of the digital age in research, is a growing field, scholarly content use is only studied in the context of more general topics. The shift in digital tools to web-based technologies only adds to these complexities. Therefore, further research in this particular field will illuminate and direct efforts to build research infrastructures which will best serve research needs.



---

## References

---

### Unpublished Sources:

European Holocaust Research Infrastructure, Deliverable 16.4 “Researcher Practices and User Requirements”, 2013.

Europeana Cloud Deliverable 1.1 “Research communities identification and definition report”, 2013.

Europeana 2013 Vocabulary, Unpublished Report.

### Published Sources:

Angelis, S., et al., “Classical philology, text-based and Holocaust studies facing digital research infrastructures: from practice to requirements”, *Supporting Digital Humanities 2011*

Bass, A., et al., “The Information Behavior of Scholars in the Humanities and Social sciences”, University of Washington, Washington, 2005, pp. 1-14.

Benardou, A., Constantopoulos, P., and Dallas, C., “An approach to analysing working practices of research communities in the Humanities”, *International Journal of Humanities and Arts Computing*, vol. 7, no. 1-2, Edinburgh University Press, 2013, pp. 105-127.

Bia, A., “The Digital Humanities Workbench,” in *Proceedings of the 13th International Conference on Interaccic&oacute;n Persona-Ordenador*, INTERACCION ’12 (New York, NY, USA: ACM, 2012), 50:1–50:4, doi:10.1145/2379636.2379685.

Brockman, W., et al., *Scholarly work in the humanities and the evolving information environment*, CLIR, 2001. Also available <http://www.clir.org/pubs/reports/pub104/pub104.pdf>, (accessed September 2013).

Bulger, M., et al., “Reinventing research? Information practices in the humanities”, *Information Practices in the Humanities, A Research Information Network Report*, April 2011, pp. 1-84.

Case, D., *Looking for Information: A survey of research on information seeking, needs and behavior*, 3<sup>rd</sup> ed., Emerald Group Publishing, 2012.

Education for Change Ltd, SIRU, “Researchers’ use of libraries and other information sources: current patterns and future trends”, The Research Partnership, University of Brighton, 2002, pp. 1-84, [www.rslg.ac.uk/research/libuse/LUrep1.pdf](http://www.rslg.ac.uk/research/libuse/LUrep1.pdf), (accessed 15 March 2013).

Europeana, “Europeana Content Analysis Summer 2010”, [pro.europeana.eu/documents/869571/874106/Content+A...](http://pro.europeana.eu/documents/869571/874106/Content+A...), 2010, pp. 1-42 (accessed 14 October 2013).

Geng, Z., et al., Visual analysis of segment variation of German translations of Shakespeare's Othello Information Visualization 1473871613495845, first published in 2013 doi:10.1177/1473871613495845

Griffiths, R., Dawson, M., and Rascoff, M., "Scholarly Communications in the History Discipline", *ITHAKA Report Commissioned by JSTOR*, August 2006, pp. 1-14, [www.sr.ithaka.org/download/file/fid/485](http://www.sr.ithaka.org/download/file/fid/485), (accessed 5 April 2013).

Houghton, J., Steele, C., and Henty, M., Changing Research Practices in the Digital Information and Communication Environment, PhD Thesis, Department of Education, Science and Training, 2003, pp. 1-186, [www.cfes.com/documents/Changing\\_Research\\_Practices.pdf](http://www.cfes.com/documents/Changing_Research_Practices.pdf), (accessed 20 February 2013).

Houghton, J., Steele, C., and Henty, M., "Research practices and scholarly communication in the digital environment", *Learned Publishing*, vol. 17, 2004, pp. 231-32.

Hughes, L., 'Using ICT methods and tools in arts and humanities research' in Hughes, L. (ed.) *Evaluating and measuring the value, use and impact of digital collections*, Facet 2011,

Jones, K., and Bennett, R., "E-resources for research in the Humanities and Social sciences: A British Academic Policy Review", *British Academy*, April 2005, pp. 1-116.

Meho, L., and Tibbo, H., "Modeling the Information-Seeking Behavior of Social Scientists: Ellis's study revisited", *Journal of the American Society for Information Science and Technology*, vol. 54, no. 6, 2003, pp. 570-587.

Nicholas, D., et al., "E-journals: Their use, value and impact", *Research Information Network*, London, 2009, pp. 1-52.

Palmer, C., Teffeau, L., and Pirmann, C., "Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development", Report commissioned by OCLC Research, January 2009, published online at: [www.oclc.org/programs/publications/reports/2009-02.pdf](http://www.oclc.org/programs/publications/reports/2009-02.pdf), (accessed 10 April 2013).

Rutner, J., and Schonfeld, R., "Supporting the Changing Research Practices of Historians", *ITHAKA S+R*, December 2012, pp. 1-52, <http://www.sr.ithaka.org/sites/default/files/reports/supporting-the-changing-research-practices-of-historians.pdf>, (accessed 5 May 2013).

Speck, R., "The AHDS Taxonomy of Computational Methods", Arts and Humanities Data Service, 2008, pp. 1-14, <http://www.ahds.ac.uk/>, (accessed 30 May 2013).

Summit on Digital Tools for the Humanities - Report on Summit Accomplishments (University of Virginia, September 28, 2005), <http://www.iath.virginia.edu/dtsummit/SummitText.pdf>.

Terras, M., 'Image Processing and Digital Humanities' in Terras, M., Nyhan, J., and Warwick, C., (eds.) *Digital Humanities in Practice*, Facet 2012.

Unsworth, J., “Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this?”, *Symposium on Humanities Computing: Formal Methods, Experimental Practice*, King’s College, London, 2000, <http://jefferson.village.virginia.edu/~jmu2m/Kings.5-00/primitives.html>, (accessed 11 September 2013).

Unsworth, J., “Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This?” (presented at the Humanities Computing: formal methods, experimental practice, King’s College, London, 2000), <http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html>.

Unsworth, J., “Tool-Time, or ‘Haven’t We Been Here Already?’” (presented at the Transforming Disciplines: The Humanities and Computer Science, Washington, DC, 2003), <http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html>.

### Websites:

“Arts-humanities.net: Guide to Digital Humanities & Arts”, [www.arts-humanities.net](http://www.arts-humanities.net), (accessed July 19, 2013).

“Bamboo Dirt”, <http://dirt.projectbamboo.org/>, (accessed 30 September 2013).

Circos, <http://circos.ca/>, (accessed 15 September 2013).

Collaborative European Digital Archive Infrastructure – CENDARI, <http://www.cendari.eu/>, 2012, (accessed 25 April 2013).

Connected Repositories - CORE Projects, <http://core-project.kmi.open.ac.uk/projects>, 2011, (accessed 5 May 2013).

Council of European Social Science Data Archives – CESSDA, <http://www.cessda.org/>, 2012, (accessed 10 June 2013).

Diction. The Text-analysis Programme, <http://www.dictionsoftware.com/>, (accessed 6 June 2013).

Digital Research Infrastructure for the Arts and Humanities – DARIAH, <http://www.dariah.eu/>, 2013, (accessed 5 June 2013).

Digital Technologies for Research in Maya Archaeology, <http://mayaarch3d.unm.edu/queryarch3d.php>, (accessed 12 October 2013).

Distributed Digital Music Archives and Libraries Lab, <http://ddmal.music.mcgill.ca/research/salami>, (accessed 15 June 2013).

DM: Tools for Digital Annotation and Linking, <http://ada.drew.edu/dmproject/>, (accessed 5 June 2013).

Electronic Archiving System of Data Archiving and Networked Services – EASY, <https://easy.dans.knaw.nl/ui/home>, 2013, (accessed 5 April 2013).

EMORY Libraries, Digital Scholarship Commons, <http://disc.library.emory.edu/lincoln/>, (accessed 20 September 2013).

EMORY University, Emory Center for Digital Scholarship, <http://digitalscholarship.emory.edu/>, (accessed 28 May 2013).

European Holocaust Research Infrastructure – EHRI, <http://www.ehri-project.eu/>, 2013, (accessed 5 May 2013).

Europeana's website, [http://www.europeana.eu/portal/usingeuropeana\\_results.html](http://www.europeana.eu/portal/usingeuropeana_results.html), 2008, (accessed 3 June 2013).

Europeana 1989, <http://www.europeana1989.eu/en/>, 2013, (accessed 5 May 2013).

Europeana 1914-1918, <http://www.europeana1914-1918.eu/>, 2013, (accessed 5 May 2013).

GitHub, <https://github.com/chrisjr/papermachines>, (accessed 17 June 2013).

Laurence Anthony's Website, [http://www.antlab.sci.waseda.ac.jp/antwordprofiler\\_index.html#sthash.6PikGkch.dpuf](http://www.antlab.sci.waseda.ac.jp/antwordprofiler_index.html#sthash.6PikGkch.dpuf), (accessed 7 May 2013).

MALLET, MAchine Learning for LanguagE Toolkit, <http://mallet.cs.umass.edu/>, (accessed 25 May 2013).

Metadata Offer New Knowledge, <http://monkproject.org/>, (accessed 5 May 2013).

Neatline, <http://neatline.org/plugins/neatline/>

Network for Digital Methods in the Arts and Humanities- NEDIMAH, <http://www.nedimah.eu/>, 2012, (accessed 25 April 2013).

OMEKA, <http://omeka.org/>

ORBIS: The Stanford Geospatial Network Model of the Roman World, <http://orbis.stanford.edu/>, (accessed 15 May 2013).

Portal Textometrie, <http://portal.textometrie.org/demo/?locale=en>, (accessed 7 June 2013).

Researchware Inc, <http://www.researchware.com/products/hypertranscribe/>, (accessed 12 October 2013).

Service Integration, data interconnect research and education - Isidore, <http://www.rechercheisidore.fr/>, 2013, (accessed 14 April 2013).

Share More, Learn More, <http://www.seasr.org/>, (accessed 3 April 2013).

Text Analysis Portal for Research, <http://tada.mcmaster.ca/Main/TAPoR>, (accessed 30 June 2013).

The UVic Image Markup Tool Project, [http://tapor.uvic.ca/~mholmes/image\\_markup/](http://tapor.uvic.ca/~mholmes/image_markup/), (accessed 10 May 2013).

TPEN, Transcription for Paleographical and Editorial notation, <http://t-pen.org/TPEN/>, (accessed 14 June 2013).

TXM, Teiwiki, <http://wiki.tei-c.org/index.php/TXM>, (accessed 5 June 2013).

Version Variation Visualisation, <http://delightedbeauty.org/vvv>, (accessed 7 June 2013).

Virtual Research Environment for the Humanities – TextGrid, <http://www.textgrid.de/>, 2012, (accessed 30 May 2013).

Voyant, <http://voyant-tools.org/>, (accessed 5 June 2013).

Web of Knowledge, <http://wokinfo.com/>, (accessed 5 June 2013).

Wordpress, <http://wordpress.com/>, (accessed 19 July 2013).