DELIVERABLE

Project Acronym: **Europeana Cloud**

Grant Agreement number: **Europeana Cloud Deliverable 1.3 User Requirements Analysis and Case Studies Report. Content Strategy Report.**

Project Title: **Europeana Cloud: Unlocking Europe's Research via The Cloud**

Europeana Cloud Deliverable 1.3 User Requirements Analysis and Case Studies Report. Content Strategy Report.
[Joint Deliverables 1.3 and 1.6: User Requirements Analysis and Case Studies Report (1.3) and Content Strategy Report (1.6)]

Revision:

Authors:

   Stavros Angelis (Athena RC)
   Agiatis Benardou (Athena RC) – Work Package Leader
   Nephelie Chatzidiakou (Athena RC)
   Panos Constantopoulos (Athena RC)
   Costis Dallas (Athena RC)
   Alastair Dunning (Europeana Foundation)
   Jennifer Edmond (TCD)
   Stefan Ekman (UGOT/SND)
   Vicky Garnett (TCD)
   Dimitris Gavrilis (Athena RC)
   Caspar Jordan (UGOT/SND)
   Kika Kalogeropoulou (Athena RC)
   Konstantinos Karozos (Athena RC)
   Ilze Lace (UGOT/SND)
   Rufus Latham (UGOT/SND)
   Marian Lefferts (CERL)
   Dimitra-Nefeli Makri (Athena RC)
   Athanassios Mantes (Athena RC)
   Eliza Papaki (Athena RC)
   Dimitar Popovski (UGOT/SND)
   Owain Roberts (NLW)
   Norman Roger (University of Edinburgh)
   Martin Schaller (ONB)
   Björn Sjögren (UGOT/SND)
   Ingeborg Versprille (CERL)
   Kees Waterman (KNAW)

Revision History

Revision	Date	Author	Organisation	Description
1.	September 2015	Agiatis Benardou	Athena RC	Table of Contents, initial content of the report.
2.	October 08, 2015	WP1 members		Revision of ToC and structure
3.	November 2015	WP1 members		Amendments to the text, revisions and changes.
4.	November 2015	RCAB		Review and approval of the text.
5.	November 2015	WP1 members		Final editing.

D1.3 User Requirements Analysis and Case Studies Report. Content Strategy Report.

Executive summary[1]

The present Deliverable (1.3) comprises of two joint reports: former Deliverable 1.3 (User Requirements and Case Studies report) and Deliverable 1.6 (Content Strategy Report).

As a product of a multi-scale, interdisciplinary effort, Deliverable 1.3 (User Requirements and Case Studies report / Content Strategy Report) employs a multi-faceted approach to make sense of the information needs and behaviour of Humanities and Social Sciences researchers both within and outside the Europeana ecosystem, while achieving profound understanding of the ways these communities interact with existing Europeana content and metadata. Through extensive and widely-levelled empirical research (Case Studies, Web Survey, Interviews, Focus Groups) complimented by thorough desk research (Literature review, study of particular thematic areas), and building on work previously conducted in the context of other Digital Humanities Research Infrastructures (DARIAH, EHRI, ARIADNE, NeDiMAH) Work Package 1 managed to reach a long list of non-prioritized User Requirements as well as a set of flexible Content Recommendations for the upcoming development of Europeana Research.

---

[1] The members of Work Package 1 would like to thank the Research Communities Advisory Board (Prof. Lorna Hughes, Dr. Leif Isaksen, Prof. Karina Van Dalen-Oskam) for their insightful comments and supportive ideas throughout this endeavour.

Table of Contents

List of figures

1. Introduction

Unlocking the value of Europeana for research in the humanities and social sciences involves earnest consideration of the information needs, practices and methods of scholars as they pursue their research, but also of the suitability of web resources accessible through Europeana for research and scholarship. This document presents the results of work by the Europeana Cloud project aiming towards two complementary objectives addressing these concerns:

a) to develop an effective research content strategy for Europeana, based on an evidence-based account of the potential usefulness of Europeana and European Library resources for research in the Humanities and Social Sciences, and,

b) to improve the understanding of digital tools, research processes and content used in the Humanities and Social Sciences, thus informing the development of tools and aggregation of content in the Europeana Cloud.

These two objectives correspond to content-driven and user-driven approaches to requirements analysis. Attainment of both objectives is critical in ensuring that the specification, design and operation of the Europeana Cloud infrastructure and complementary project outputs will leverage the information riches of Europeana to match the requirements for European research in the humanities and the social sciences. The Europeana Cloud project invested, therefore, considerable effort in planning and implementing a comprehensive evidence-based research programme towards these two objectives. As part of the activities of Work Package 1, and under the guidance of a Research Advisory Board established to validate their work, Europeana Cloud partners conducted a wide-ranging literature survey on the information practices and needs of European humanities and social science researchers, both in the non-digital and digital domain; they surveyed of the use, needs and attitudes of researchers towards information resources and digital technologies; they established case studies from across the humanities and social sciences, and organized expert forums, seeking to elicit expert knowledge, and, combined with other sources of evidence, produce user requirements; and, finally, they surveyed Europeana and European Library to produce a qualitative assessment of digital resources available, and develop a content strategy. The work was co-ordinated by the Digital Curation Unit, Institute for the Management of Information Systems – Athena Research Centre (DCU), with significant involvement and contribution by the Europeana Foundation (EF), Trinity

College Dublin (TCD), University of Gothenburg (UGOT/SND), the Consortium of European Research Libraries (CERL), National Library of Wales (NLW), the University of Edinburgh (UEDIN), the Austrian National Library (ONB), and the Royal Netherlands Academy of Arts and Sciences (KNAW).

In the methodological framework established as part of the planning stage, it emerged clearly that all parts of the work were interdependent: in other words, that the dimension of the work related to researcher practices and requirements is inseparable from the dimension related to assessing Europeana content. This report, therefore, brings together under a single overarching structure the content of two public deliverables, originally enumerated in the Europeana Cloud Description of Work as separate documents: <u>D1.3 User Requirements Analysis and Case Studies Report</u>, and <u>Deliverable 1.6 Content Strategy Report</u>, which is now treated as an integral part of D1.3. Chapters 2 (Literature review and related work), 3 (Research Communities Web Survey), 4 (Humanities and Social Sciences Case Studies) and 4 (APIs in Humanities and Social Sciences Research) relate mostly to the former, while Chapter 6 (Research Themes: Parliamentary Papers; Population Displacement; Political Propaganda; Children's Literature; School History; The Ancient Mediterranean) to the latter. Finally, the concluding Chapter 7 (User Requirements and Content Strategy Recommendations for Europeana Research) is based on a synthetic consideration of both dimensions addressed by research reported in earlier chapters.

2. Literature review and related work

An essential first step in the process of Improving our understanding of digital tools, research processes and content used in the Humanities and Social Sciences, one of the principal objectives of Work Package 1, was to take stock of the current situation as emerging from the outputs of relevant research projects and as reported in scholarly publication. An initial literature review was conducted towards three successive Milestones during the first nine months of the Project, to provide background to the main findings of Deliverable 1.2 - State of the Art Report on Digital Research Practices, Tools and Scholarly Content Use, meant to present the exploratory stage of our research. The current chapter, which further connects with reporting on tasks such as the Research Communities Web Survey (T.1.3.4) and the Case Studies Expert Fora (T.1.4) undertaken to provide evidence towards user requirements, is part of the confirmatory stage of our research.

The rising impact of digital technology on scholarly research

Scholarly work has been one of the first domains that attracted the interest of information behaviour research. The inception of digital technologies by researchers, and the methodological promises and concerns related to the emergence of humanities computing and, further, digital humanities, shifted the interest to the question of how scholarly practice is affected by the use of digital resources, tools and services, and to what extent the digital mediation of accessing, managing and using scholarly resources by humanities researchers boils down to a standard set of granular micro-activities, or "scholarly primitives".[2] 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 since the 1950s, 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

---

[2] 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', 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).

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".[3]

Since the beginning of the 21st century, a growing range of research publications and projects has been addressing the impact of the inception of digital technologies on research practice in the Humanities and Social Sciences.[4] The wide interest in the impact of the digital age on the work practices of humanists and social scientists can be set in the light of their well-attested reluctance to adopt digital methods in the research process.[5] Nevertheless, the digital age has resulted in notable changes in the research practices of a growing community of researchers who increasingly use digital tools and content. 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",[6] and this promise has motivated increasing number of scholars to engage with digital approaches to scholarship. While in the context of broader changes, such as the adoption of digital infrastructures in the hard and natural sciences, the use of digital technology by the humanities and social sciences has been seen to have significant differences with regard to both researcher information behaviour and attitudes. While John Houghton, Colin Steele and Margaret Henty note the hesitation of humanists and social sciences to adopt digital sources and practices, merely acknowledging that "they have not ignored the technology" but insisting on the existence of a disciplinary divide with the sciences in terms of communication, collaboration, networking and in the form of sources used,[7] other studies describe humanities scholars in particular as increasingly enthusiastic users of digital resources.[8]

---

[3] Case, *Looking for Information: A survey of research on information seeking, needs and behavior*, 3rd ed., Emerald Group Publishing, 2012, p. 279.

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

[5] See for example the perceptions of digital content and practices within History in Houghton, Steele and Henty, Changing Research Practices in the Digital Information and Communication Environment, PhD Thesis, Department of Education, Science and Training, 2003, p. 10.

[6] Houghton, Steele and Henty, "Practices in the Digital Information", p. 52.

[7] Houghton, Steele and Henty, "Practices in the Digital Information", p. 51.

[8] Bulger et al., "Reinventing research?", p. 10.

The growing amount of literature, research projects and digital infrastructures for humanities and social sciences mirrors the rise of the impact of digital technology on scholarly research practices, including publication and collaboration. As technology evolves, digital research practices and related literature will closely follow the extent of technology adoption. A point made by Griffiths, Dawson and Rascoff perfectly captures this constant development: whereas humanities scholars 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".[9] The reason underlying this phenomenon may be the easy colonization and efficient use of the digital environment by younger scholars, who perceive it as a comfortable domain for conducting their own research. Unlike many researchers of the previous generation, younger scholars approach their research digitally since its beginning – this, in fact, is their "standard", or "traditional" way of conducting research. This generational change has a significant impact on the shift of attention, and value, from analogue to digital resources; as noted by the 2006 ITHAKA Report (2006), historians justifiably "worry that for undergraduates 'if it isn't electronic, it doesn't exist'".[10]

Digital content use by scholars

As researchers and scholarly resources constitute two core elements of the evolving scholarly process we sought to explore in detail the diverse dimensions defining and affecting how these two elements interact by looking at the active debate on the issue in scholarly literature. Nowadays, with the increasing domination of digital content in scholarly practice gives rise to various questions regarding the ways in which researchers identify, access, and use content. For instance: 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?

---

[9] Griffiths, Dawson and Rascoff, "Scholarly Communications", p. 5.

[10] Ibid, p. 12.

The topic of scholarly use of digital content figures extensively in information seeking behaviour literature.[11] The multidisciplinary study by Houghton, Steele and 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, including those in the arts and humanities, and the social sciences.[12] A number of its arguments are based on a survey conducted by the 2001 Research Assessment Exercise (RAE) Census in the United Kingdom. Another study, which also 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.[13]

Yet another study, conducted by the University of Washington in 2005, similarly approached scholarly content use in the context of information behaviour of scholars.[14] The study, based on an eapmination of information behaviour of researchers in the disciplines of humanities and social sciences seen as a unified field, focused on similarities rather than differences with regard to information approaches and methods used by scholars. In contrast, a survey on the use of e-resources for research in the humanities and social sciences, published by the British Academy in 2005, seeks to throw light on the differences in research approaches advanced, respectively, by the humanities, and the social sciences as they engage with digital resources.[15]

As early as 2003, L. Meho and Helen Tibbo agreed on the advantages offered by the Internet to social scientists in reaching digital information easily; however they noted that, at that time, "the problem is that only a small fraction of available information is digitized and made accessible to

---

[11] For the purposes of D1.2 the term "content" was defined based on the types of content found in Europeana (see examples of genres of Europeana's content visit http://www.europeana.eu/portal/usingeuropeana_results.html, accessed 3 June 2013; Information was also taken from the updated Europeana 2013 Vocabulary, unpublished report.)

[12] Houghton, Steele and Henty, Changing Research Practices.

[13] 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**/lib**use**/LUrep1.pdf, (Accessed: 15 March 2013).

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

[15] 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.

researchers".[16] Eight years later, the Research Information Network Report stressed the problem of assessing digital resources.[17] This time gap allowed 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, an observation that seems to agree with Claire Warwick et al. paper on "Library and Information resources and users of digital resources in the humanities" published in 2008.[18]

More discipline-focused studies were also examined in the context of Deliverable 1.2 such as the reports by the ITHAKA organization of 2006 and 2012 which present research activities and the measure of engagement of historians with digital content and practices.[19] A systematically theorized approach, based on the development of a formal conceptual model of scholarly information activity, is introduced by Agiatis Benardou, Panos Constantopoulos and Costis Dallas in their paper on working practices, digital content, its usability and privacy issues of the research community of Holocaust researchers as a study group.[20]

Other Research Infrastructures initiatives

The strong presence of the digital aspect in research is clear from the long list of research initiatives which have emerged during the last decade successfully exploring this area. More particularly, a significant number of research infrastructures in the Humanities and Social sciences directed their efforts towards identifying current research patterns, developing tools, and enlarging the pool of digital resources available to researchers. Among user-centric infrastructures, notable examples include the Digital Research Infrastructure for the Arts and Humanities (DARIAH) and its preparatory project Preparing DARIAH. In the context of DARIAH,

---

[16] Meho, 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, p. 586.

[17] Bulger et al., "Reinventing research? Information practices in the humanities", *Information Practices in the Humanities, A Research Information Network Report*, April 2011, p. 72.

[18] Warwick; Terras; Galina.; Huntington; Pappa (2008) Library and information resources and users of digital resources in the humanities. Program: Electronic Library and Information Systems, 42 (1) pp. 5-27.

[19] Griffiths, Dawson and Rascoff, "Scholarly Communications", pp. 1-14; Rutner and Schonfeld, "Changing Research Practices", pp. 2-52.

[20] 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.

various research results have been developed such as TextGrid and Isidore.[21] 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 an online system that collects, enriches and provides a unified data access for the digital humanities and social sciences. Another initiative has developed EASY, an Electronic Archiving System able to collect, archive and offer access to thousands of datasets in the humanities, social sciences and other disciplines.[22]

This brief list of research projects is further expanded by looking at other European infrastructure initiatives such as the European Holocaust Research Infrastructure (EHRI), the COnnecting 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. 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 may thus be considered as large aggregators of content. 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. Finally, NeDiMAH engaged in 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, centred around the concept of digital research method; a notable result of the project is the NeDIMAH Methods Ontology (NeMO), developed through a collaboration with the Digital Curation Unit, Institute of Management and Information Systems, Athena RC (DCU).[23]

---

[21] 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).

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

[23] Hughes, Constantopoulos, & Dallas. "Digital Methods in the Humanities: Understanding and De-scribing their Use across the Disciplines". In Schreibman, Siemens, & Unsworth (ed.). *A new companion to digital humanities.* Oxford: Wiley-Blackwell, 150-170.; More information on NeMO can be found at http://nemo.dcu.gr/

Focusing on digital resources, the Council of European Social Science Data Archives (CESSDA) constitutes the prime European infrastructure which contributes in locating, accessing and sharing data for the social sciences.[24] The Common Languages Resources and Technology Infrastructure (CLARIN) on the other hand focus more on linguistic resources while Bamboo Digital Research Tools wiki (BambooDiRT) is an initiative oriented towards an inventory of digital tools.

What about tools?

In 2003 Unsworth considered many of the questions regarding the development and use of digital tools.[25] He noted 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. In 2013, this vision is finally becoming a reality through initiatives such as The DHW project.[26] Other projects that have taken similar approach are the Salami Project[27], the Monk Project[28] and SEASR - The Software Environment for the Advancement of Scholarly Research.[29]

Tools and infrastructure in the Digital Humanities was discussed as part of the *The Cologne Dialogue on Digital Humanities 2012.*[30] Here, Anderson et al proposed the idea of *"…digital ecosystems[,...] 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*

---

[24] Council of European Social Science Data Archives – CESSDA, http://www.cessda.org/, 2012, (Accessed: 10 June 2013).

[25] Unsworth (2003). Tool-Time, or 'Haven't We Been Here Already?'. *Presented at the Transforming Disciplines: The Humanities and Computer Science*. Washington, DC. Retrieved from http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html

[26] Bia, "The Digital Humanities Workbench," in *Proceedings of the 13th International Conference on Interacción Persona-Ordenador* (New York, NY, USA: ACM, 2012), pp. 50:1–50:4.

[27] http://ddmal.music.mcgill.ca/research/salami

[28] http://monkproject.org/

[29] http://www.seasr.org/

[30] M. Thaller (ed.), Historical Social Research Vol. 37 (2012), no. 3.

increased participation, sharing and building a social network of people, things, content and so on.[31]

3. Research Communities Web Survey

3.1 Methodology

The Europeana Cloud Research Community Web Survey was designed in order to facilitate an evidence-based account about the information practices and needs of humanists and social scientists in the digital environment, and the potential use of Europeana in the context of scholarly research. A questionnaire survey was designed by Athena RC in collaboration with NIOD in 2013, taking into account the protocol adopted in earlier questionnaire survey research in the context of the Preparing DARIAH project and the European Holocaust Research Infrastructure (EHRI). A purposive sampling approach was adopted, seeking to represent the population of humanities and social science scholars who are aware of and possibly involved with the Europeana community, as these match the target audience for Europeana Research. The link to the online questionnaire was disseminated to the Europeana network, through e-mails and social media. Following the identification and definition of research communities developed in D1.1, the target group included users from the fields of archaeology, history, law, linguistics, musicology, philosophy, social anthropology, social (and human, economic, political and cultural) geography, gender studies, economic and social history, political science and sociology.

The web survey was designed to measure specific aspects of research activity as a means of better understanding of the needs of digital users. It is divided in four sections, measuring: (1) the use of specific digital services by researchers, (2) the kinds of research activities users engage with, (3) the content as well as significant properties of resources favoured by users, and (4) the degree of agreement or disagreement with specific statements regarding the research process. Questions were complemented by a set of essential profile (demographic and socioeconomic) questions, aiming to help explain patterns identified in the results.

---

[31] Anderson and Blanke (2012). Taking the Long-View: From e-Science Humanities to Humanities Digital Ecosystems. *Historical Social Research, 37*(3), p. 147.

In order to measure the extent to which the participants agree or disagree with the questions presented in the questionnaire, Likert-type scales were used.[32] Likert-type scales most often range from 1 to 5 and are commonly used in order to measure evaluation. The evaluation statements (responses) are ordered and numeric values are assigned to each of them for the purpose of analysis on the ordinal scale. The scale chosen for this questionnaire comprises a middle value, allowing participants to indicate if their response to a question is considered to be neutral, thus allowing analysis to be more comprehensive. Most questions were of this type and could be answered using the following scale:

---

1 = not at all   2 = somewhat   3 = moderately   4 = very   5 = extremely

---

Here we report on the results of an analysis of descriptive statistics of closed questions in the questionnaire. In addition, some open (free text) questions were included in the questionnaire, allowing participants to provide examples and qualifications on their use of particular services or methods. We recoded responses to open questions to nominal variables, in the light of a qualitative data analysis approach, and in this document we analyze the results of descriptive statistics of these data, thus achieving more granularity and richness than originally addressed by closed questions in the questionnaire. The full contents of the Europeana Cloud Web Survey can be found in Appendix 1.

3.2 Dataset

The dataset of the Europeana Cloud Research Community Web Survey consists of 65 responses.[33] Regarding the field of study to which the respondents are attached, approximately three quarters of the sample (74% or 48 respondents) are attached to the humanities, while one

---

[32] For a detailed discussion on the use of Likert-type scales see Boone, & Boone. (2012). Analyzing Likert Data, *Journal of Extension*. Volume 50, Number 2, Clason & Dormody. (). Analyzing Data Measured by Individual Likert-Type Items. *Journal of Agricultural Education*. Volume 35, No. 4, and Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical education*. 38(12), 1217-1218.

[33] Indeed, the number of responses may not seem extremely satisfactory. However, one needs to keep in mind that the Web Survey results are complimentary to the rest of the empirical work (see below). Furthermore, this present Web Survey is planned to be compared and contrasted to other similar efforts across DHRIs in Europe (i.e. the DARIAH-EU and EHRI surveys) at a later stage, in the context of Europeana Research.

quarter (26% or 17 respondents) of the sample stated that their field of study was the social sciences in Figure 1.
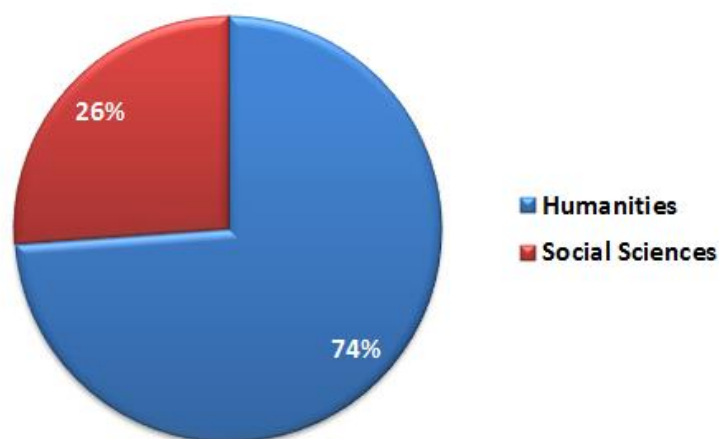
Figure 1. Respondents' field of study. N=65

Participants were asked to identify their primary and, if applicable, their secondary fields of research. They report belonging to a variety of disciplines. History alone represents a quarter of the total sample (26%, or 17 respondents). If taken together with classics (9%, or 6 respondents) and archaeology (11%, or 7 respondents), which contributed about a tenth each, those in one of the historical disciplines represent in total almost half of the sample (46% or 30 respondents). On the other hand, a smaller concentration of the sample can be found aggregated around social sciences and communication and media studies (9% or 6 respondents each) which combined would represent nearly a fifth of the sample of this survey (18% or 12 respondents).

Legend:
- Communications & Media Studies
- Social Sciences
- Archaeology
- Classics
- History
- History of Science, Technology & Medicine
- Language & Literature
- Linguistics
- Medieval Studies
- Philosophy
- Other

Other: one respondent each from Anthropology, Cultural Studies, Folklore, Geography, Music, Paleontology, and Theology & Religious Studies.
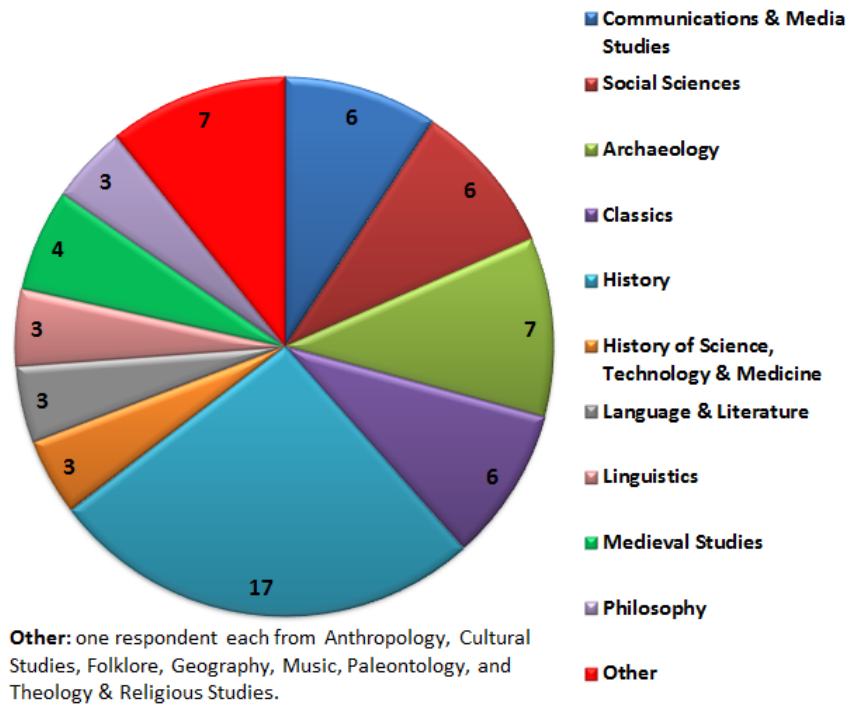
Figure 2. Disciplinary belonging. (Identify your primary and, if applicable, secondary fields of research). N=65

Participants were asked to identify their primary and, if applicable, their secondary professional affiliation and status. Since they could provide more than one response to this question, the percentages displayed below refer to the answers (N=74) and not to the individuals. Most respondents (35) stated that they work in an academic institution, while 9 work in a research institution outside the academic sector and 8 are working as librarians. The rest of the sample is distributed between PhD or postgraduate students (6), researchers, freelance PhD or postgraduate students (4), archivists (4), amateur researchers (4), undergraduate students (1), curators(1), museum professionals (1) and casual users (1).
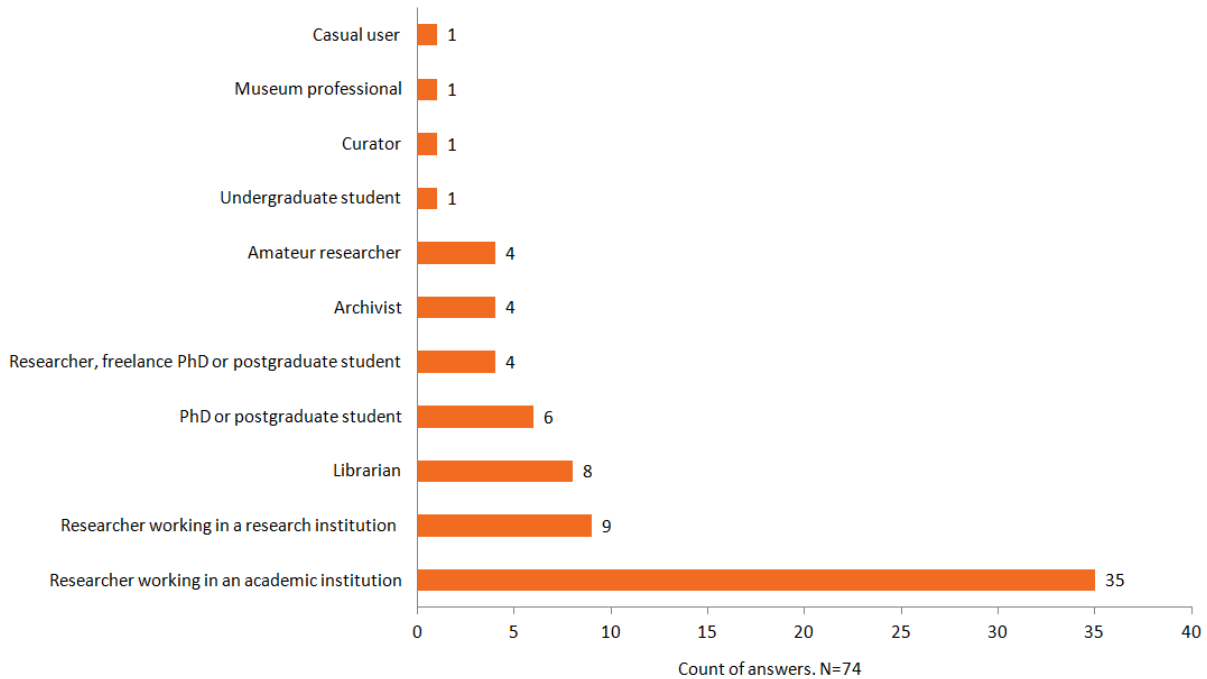
Figure 3. Participants' professional affiliation and status (How would you describe yourself?). N=74

Based on the data presented above, it is obvious that the vast majority of participants identify themselves as researchers. Overall the total percentage of those subjects that described themselves as researchers represents more than two thirds of the total sample (70% or 52 answers):

Figure 4. Total number of researchers in the sample. N=74

3.3 Findings and discussion

Use of specific services

In order to appraise the use of specific services made by researchers while working on their research projects, the respondents were asked to state how often they have used in the last 12 months the Europeana portal, the European Library, Google Scholar, some specific online digital archive related to their research, a commercial portal of a scholarly journal, museum websites and online library catalogues. The data in Figure 5 indicate that online library catalogues (84% or 52 respondents), specific online digital archives (83% or 52 respondents) and portals of scholarly journals (58% or 34 respondents) are mostly used by the respondents, followed Google Scholar (53% or 33 respondents), museum websites (33% or 19 respondents), the Europeana portal (10% or 6 respondents) and the European Library (5% or 3 respondents). The figures given are aggregates for using the resource once a month or more. Regarding the Europeana portal, two fifths of the respondents (41% or 24 respondents) state that they have not used it in the last 12 months, almost half (48% or 28 respondents) state that they have used it a few times, one in twenty (5% or 3 respondents) state that they use it at least once a month, 3% (2 respondents) state that they use it at least once a week and 2% (1 respondent) state that they use it several times a week. Finally, a significant proportion of the respondents state that they frequently use some other service, not mentioned in the list that was presented to them, for their research.
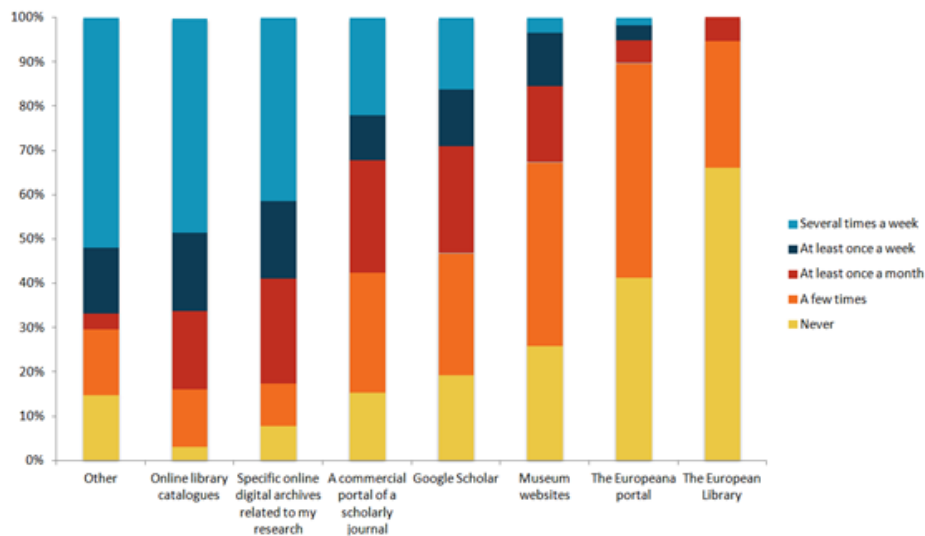
Figure 5. Frequency of use of services in the last 12 months (How often have you used the following services in the last 12 months?) Other, N=27. Online library catalogues, N=62. Specific online digital archives, N=63. A commercial

In a subsequent question the respondents were asked to identify specific digital archives or services they use. They were provided with three open free-text slots for digital archives and three further slots for other services. Figure 6 presents the frequency of use of digital archives and services that appear more than twice in the responses. The data indicate that Jstor, a digital library of academic journals, books and primary sources, is most frequently used, followed by the British Library, Gallica, Google Books, Perseus, E-brary, e-codices, TLG and Scopus.
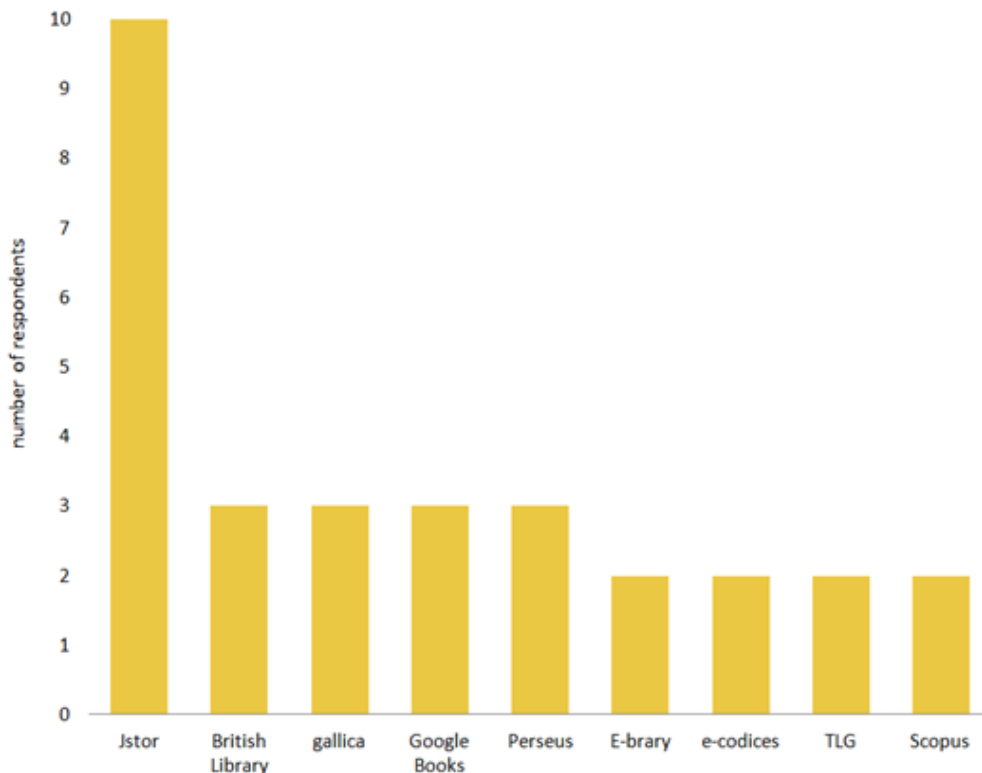
Figure 6. Count of respondents using specific digital archives and other services (Identify the digital archives or other services you stated you have used in your response to the previous question).

Assessment of the importance of activities undertaken while seeking scholarly information

The respondents were asked to rate the importance of a series of ten activities they undertake as they seek relevant information for their research. More specifically they were invited to assess how important they consider to be activities such as searching specific query terms in order to find relevant information, searching using a combination of more than one query terms to find relevant information, conducting a new search within the results of a previous search,

consulting collection summaries to find information, consulting collection inventories or indexes to find information, browsing related sources, finding relevant information on the basis of resources one just happens to stumble upon, finding resources through footnotes in articles and/or books, asking peers and/or colleagues and asking reference librarians, curators or archivists. The respondents were presented with a numbered scale from 1 to 5, where 1 corresponded to "not at all important", 2 to "somewhat important", 3 to "moderately important", 4 to "very important" and 5 to "extremely important". The results, as seen in Figure 7, indicate that the use of queries, comprising either specific terms or a combination of terms, are judged to be very important. (90% or 57 respondents have considered specific terms as "extremely important" or "very important", and 87% or 55 respondents have considered combination of terms as "extremely important" or "very important".) Moreover, when seeking relevant information for their research, the respondents seem to consider important activities such as finding resources through footnotes in articles and/or books and using the results of a search in order to conduct a new search. (65% or 40 respondents have considered footnotes to be "extremely important" or "very important", and 56% or 35 respondents have considered searching within a previous search as "extremely important" or "very important".) Browsing related resources and consulting collection inventories, indexes and summaries are also rated as important (browsing related resources: "extremely important" or "very important" 33% or 20 respondents; consulting collection inventories etc.: "extremely important" or "very important" 38% or 23 respondents) while activities such as asking peers or colleagues, finding relevant information on the basis of resources one just happens to stumble upon and asking reference librarians, curators or archivists seem to be less central in the process of seeking information (fewer than one third of the respondents considered each of these to be "extremely important" or "very important").
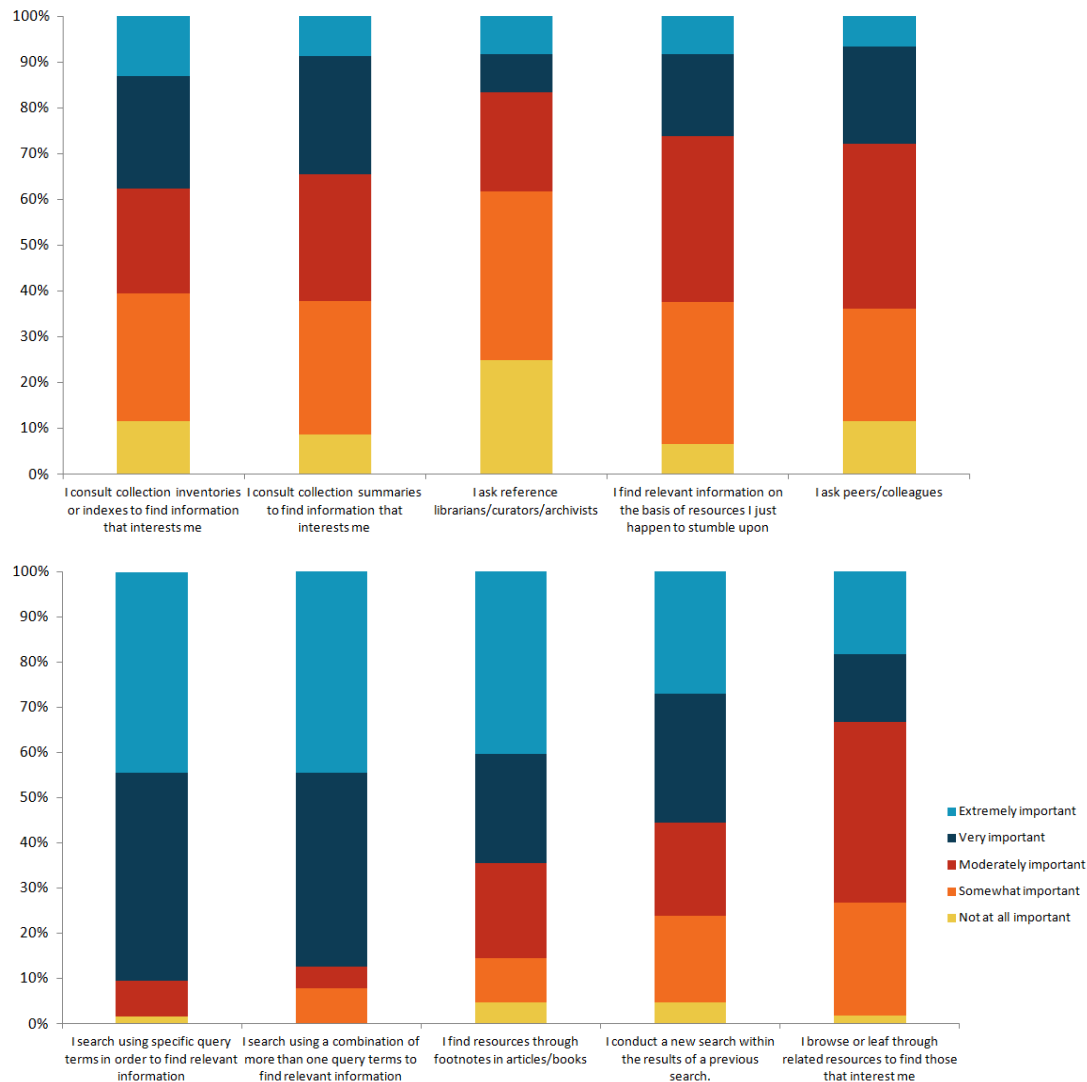
Figure 7. Importance of information-seeking activities (Assess how important are the following activities as you seek relevant information for your research). Specific query terms, N=63. Combination of query terms, N=63. Footnotes, N=62. Search within previous search, N=63. Browse related resources, N=60. Collection inventories/indexes, N=61. Collection summaries, N=58. Reference librarians/curators/archivists, N=60. Resources I stumble upon, N=61. Peers/colleagues, N=61.

Examples of queries asked by the respondents as they seek relevant information for their research

In a subsequent question, the respondents were asked to provide a common query they use to ask when they search for information about their research. The queries provided as examples differ significantly as regards their format and content. Most questions provided are thematic, looking for a specific subject, period of time or geographical region and are phrased affirmatively. Examples of this kind of query include "Digital Geography", "First World War",

"Iroquois Indians" or "Athens". These questions might also look for specific objects, such as "GIS" or persons, such as "Jakob Hurt". Another group of queries is phrased in the form of questions, such as "what is the etymology of word X?", "Where can I find a Photograph of X object" or "where is the first attestation of word X?". Finally Boolean queries are also used, such as "Caspar AND Joseph AND Dorer AND 1714" or "[author] AND [type-review];[subject] AND [geographic descriptors] OR [time descriptors, limits]". The word cloud below showcases the words used in the queries provided by the respondents.

Assessment of the aspects of a resource content while seeking scholarly information

The respondents were presented with a list of characteristics of a resource and asked to rate them as regards their importance. Those characteristics were the names of people mentioned in or represented by a resource, the kinds of objects, artefacts and art works, the classifications of people mentioned or represented, the specific places mentioned or represented, the dates, time spans or periods related to a resource, the names of specific events related to a resource, the classifications of events, activities or processes mentioned or represented and other things, ideas or entities related to a resource that were not mentioned in this list. The data in Figure 9 indicate that the dates, time spans or periods related to a resource are considered to be "very important" and "extremely important" by three quarters of the respondents (47 of 62 respondents), although this can probably be explained by the predominance of historians in the sample. Other aspects characterizing the content of a resource considered to be extremely or very important by most respondents include the specific places mentioned (55% or 34 respondents) and names of specific events (51% or 31 respondents). Classifications of events, activities or processes mentioned or represented as well as the names of people mentioned in, or represented by, a resource are considered to be extremely or very important by less than half of the respondents. Finally, the aspects of the content of a resource that seem to be considered less important by most respondents include the kinds of objects, artefacts and art works, the classifications of places mentioned or represented and the kinds of people mentioned or represented.
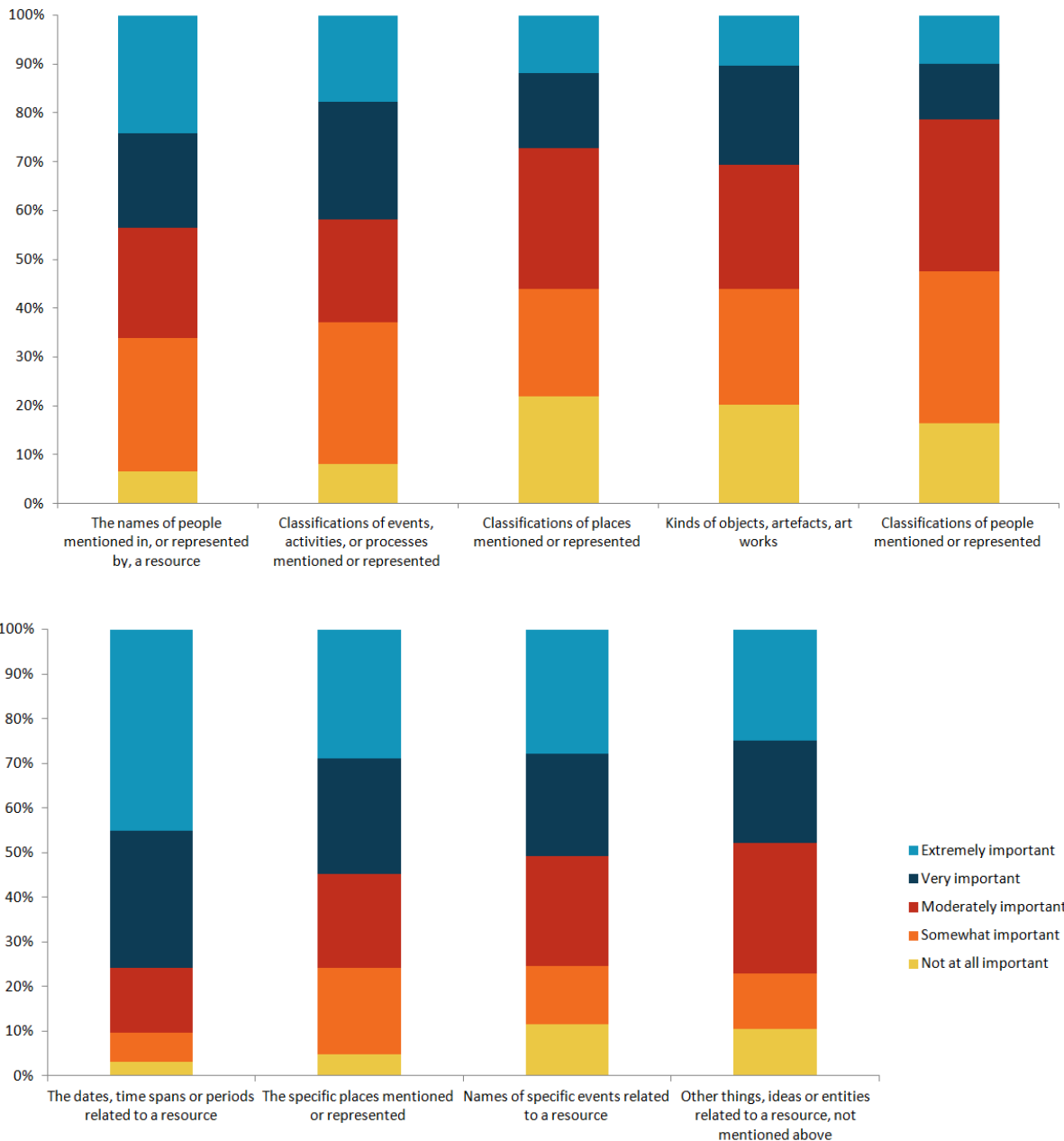
Figure 8. Assessment of resource characteristics in information-seeking (Assess the importance of the following aspects of the content of a resource as you seek relevant information for your research). Names of people, N=62. Classifications of events etc., N=62. Classifications of places, N=59. Kind of object etc., N=59. Classifications of people, N=61.Dates, time spans, periods, N=62. Specific places, N=62. Names of events, N=61. Other, N=48.

On the whole, the data presented above suggest that the respondents prioritize aspects of the resources that are very specific, such as the dates, places and events related to a resource. On the other hand they consider less important those characteristics which are indirectly or more loosely related to a resource, such as the kinds of objects, artefacts and art works, the

classifications of places mentioned or represented or the kinds of people mentioned or represented.

Assessment of specific properties of a resource while seeking scholarly information

The respondents were subsequently asked to rate a series of specific properties of a resource while seeking scholarly information. The measurement was in the same scale, from 1 to 5 where 1 corresponds to "not at all important" and 5 to "extremely important". The properties proposed were the style, period, group or movement related to a resource, the format of a resource, the object type related to a resource, the genre, the name of the author, creator or issuing authority of a resource and the fonts, collection or holding institution to which the resource belongs. The data suggests that the name of the author, creator or issuing authority of a resource is considered to be the most important property of a resource by far (80% or 48 respondents consider them very or extremely important), followed by the fonts, collection or holding institution to which the resource belongs and the format of a resource (47% or 28 respondents consider them very or extremely important). The rest of the properties under consideration are considered to be relatively less important, with between half and a third of the respondents considering them very or extremely important. However, no property is considered "not at all important" by more than one fifth of the respondents.
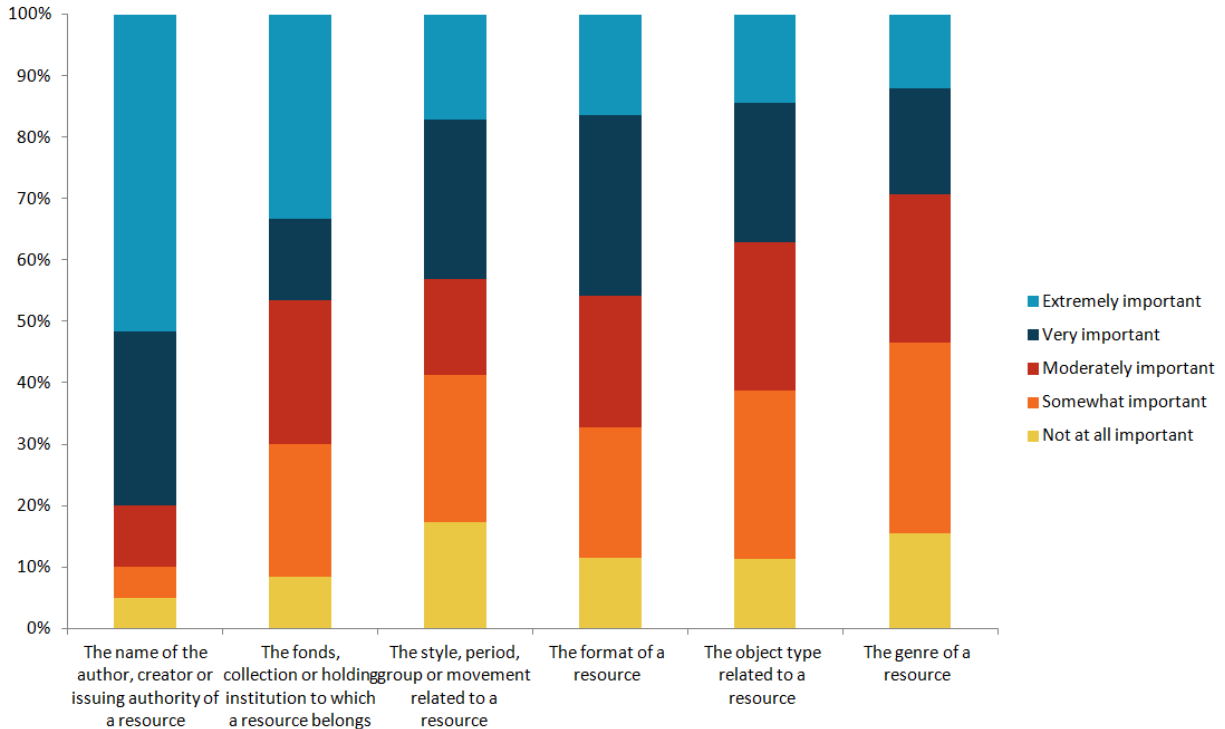
Figure 9. Assessment of specific properties of a resource in information-seeking (Assess the importance of the following properties of a resource as you seek relevant information for your research).Name of author etc., N=60. Fond etc., N=60. Style etc., N=58. Format, N=61. Object type, N=62. Genre, N=58.

Assessment of the importance of specific activities undertaken while organizing unpublished materials

The respondents were presented with an extended list of digital and non-digital activities undertaken while organizing unpublished materials and were asked to rate their importance. The activities proposed were: collecting lists of references to primary resources in a paper document or word processing file, manually copying or retyping the content of resources, filing together photocopies of resources, filing together digital copies of resources, using catalogue cards to manage resources, collecting and keeping references to resources, using keywords in order to identify the topic of resources, maintaining a manual index of keywords on resources, using a software application to organize resources, keeping a list of keywords for references to primary material. The data presented in Figure 10 suggests that filing together digital copies of resources is considered to be extremely or very important by most respondents (87% or 53 respondents). Moreover, the activities related to the collection of references are also considered highly relevant, as most respondents consider collecting and keeping references to all resources of interest and collecting lists of references to primary resources in either digital or analogue form to be "very important" or "extremely important" (59% [36 respondents] and 64%

[39 respondents], respectively). Using a software application to organize resources, manually copying or retyping the content of resources and filing together photocopies of resources are considered to be "moderately important", "very important" or "extremely important" by most participants (52% [32 respondents], 57% [35 respondents], and 56% [34 respondents], respectively), while the other activities proposed are considered less important. In particular keeping a list of keywords for references to primary material, maintaining a manual index of keywords on resources and using catalogue cards to manage resources are not considered to be important by most respondents – more than half of the respondents marked these as "not at all important".
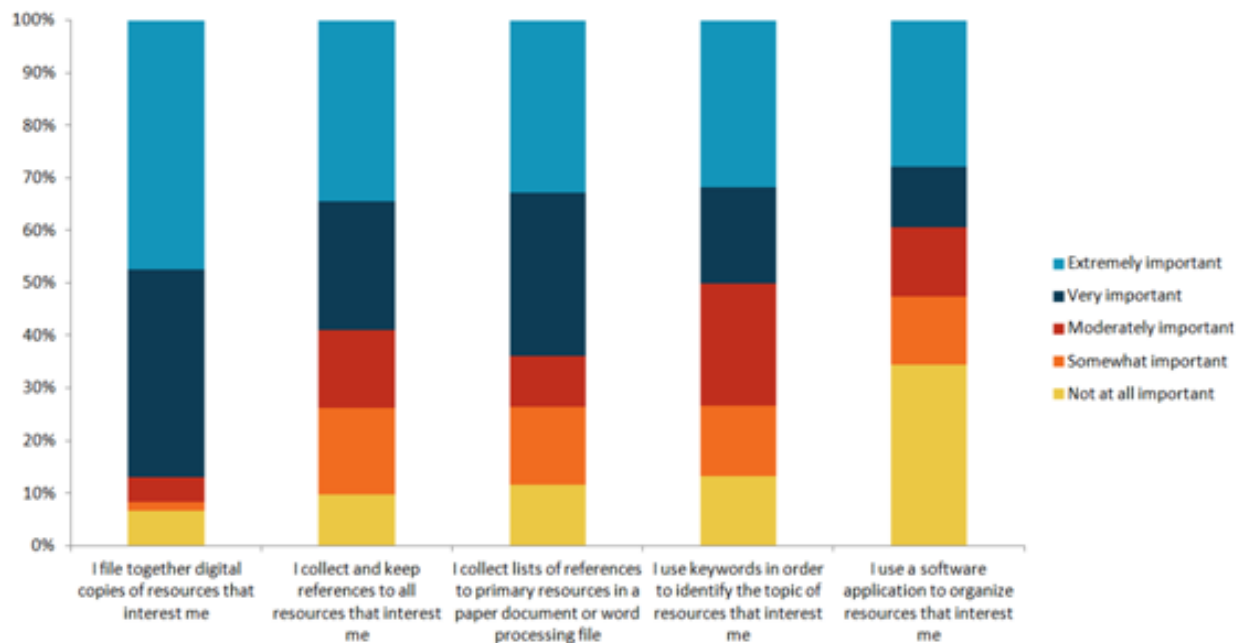
Figure 10. Assessment of specific activities for organizing unpublished materials (Assess how important is each of the following activities you undertake as you organize unpublished materials). File together digital copies, N=61. Collect and keep references, N=61. Collect list of references, N=61. Use keywords, N=60. Use software application, N=61. Manually copy content, N=61. File together photocopies, N=61. Keep list of keywords, N=60. Maintain manual index of keywords, N=60. Use catalogue cards, N=60.

Assessment of the importance of specific activities for organizing published materials

For organizing published materials, the respondents consider downloading and storing digital copies the most important of the proposed activities (80% or 49 respondents considered it extremely or very important). Collecting relevant bibliographic references is also considered "very important" or "extremely important" by most respondents (58% or 35 respondents). Less important are considered activities such as using a bibliographic reference management application to manage references and/or published materials, obtaining, copying and/or printing and storing physical copies of published materials and keeping a list of keywords for collected bibliographic references (considered very or extremely important by 39% [24 respondents], 48% [29 respondents], and 22% [13 respondents], respectively).
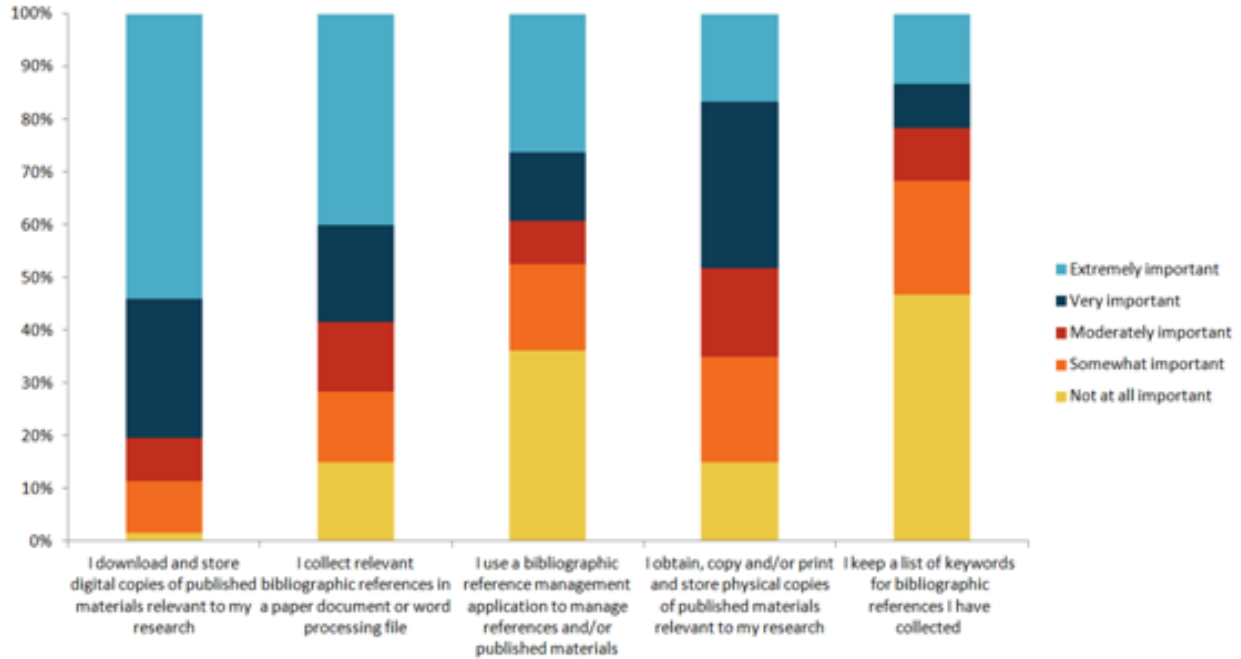
Figure 11. Assessment of specific activities for organizing published materials (Assess how important is each of the following activities as you organize published materials). Download and store digital copies, N=61. Collect bibliographic references, N=60. Use bibliographic reference management application, N=61. Obtain and store physical copies, N=60. Keep a list of keywords, N=60.

Assessment of the importance of specific activities for studying and annotating scholarly information

In a subsequent question the participants were asked to rate the importance of a series of activities undertaken as they study and annotate information relevant to their research. According to their answers the respondents believe that the most important relevant activities are underlining or highlighting relevant passages and scanning texts quickly (considered very or extremely important by 49% [30 respondents] and 66% [41 respondents], respectively). Nevertheless, the rest of available activities, as shown in Figure 12, are also considered moderately, very or extremely important by most respondents (between 60 and 80% [39 to 48 respondents]).
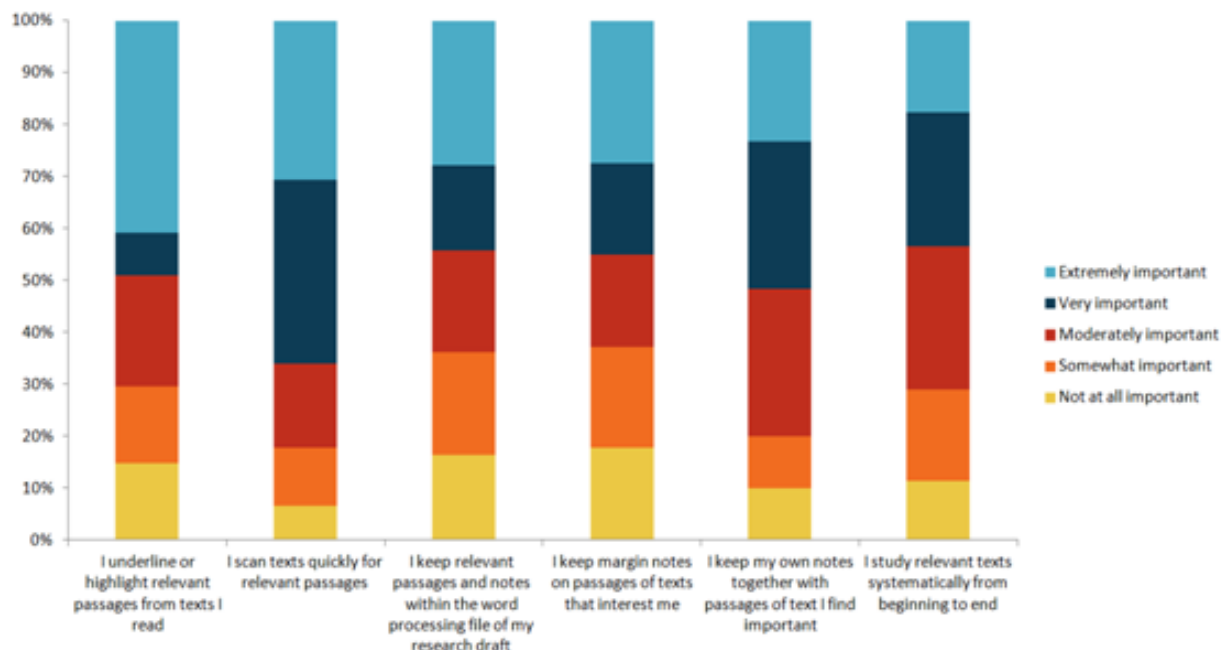
Figure12. Assessment of the importance of specific activities for studying/annotating information (Assess how important is each of the following activities as you study and annotate information relevant to your research). Underline or highlight, N=61. Scan texts, N=62. Keep relevant passages, N=61. Keep margin notes, N=62. Keep own notes, N=60. Study texts systematically, N=62.

Assessment of the importance of specific activities while working with others on a research project

Most activities related to scholarly collaboration and working with others on a research project proposed to the participants were considered to be fairly important, as more than half of the respondents rated each of the activities as moderately important or more. Of those, sharing copies or resources with colleagues, asking colleagues for their expert opinion on specific resources at a late research stage, and collaborating on joint publications or conference papers are considered most important, considered very or extremely important by 73% (44 respondents), 52% (31 respondents), and 55% (31 respondents), respectively. They are followed by asking colleagues on their expert opinion on initial research ideas (50% or 30 respondents), collaborating with colleagues on developing shared information resources and databases (51% or 30 respondents) and finally sharing one's own notes on specific resources with colleagues (33% or 20 respondents).
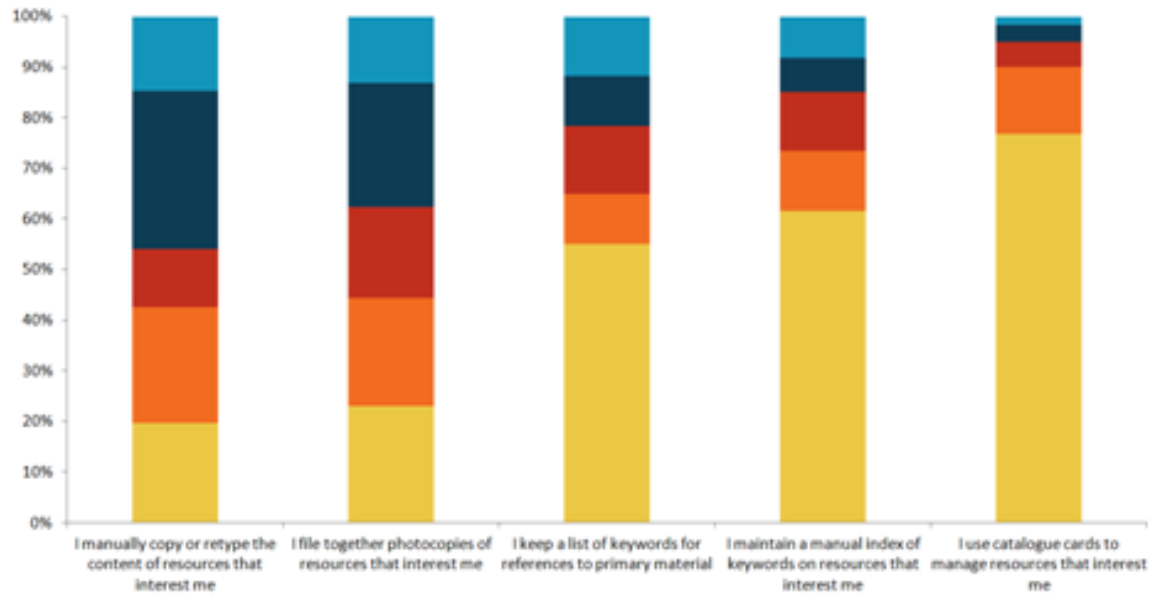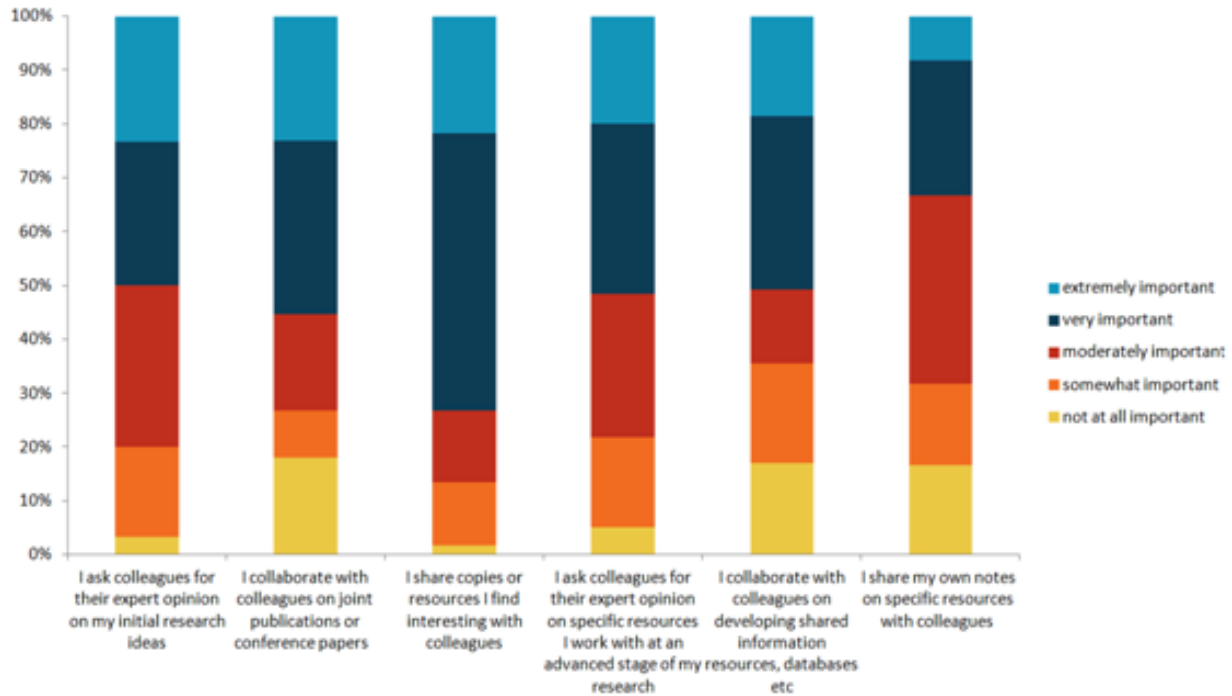
Figure 13. Assessment of specific activities for working with others (Assess how important is each of the following activities as you work with others for your research). Ask opinions on initial ideas, N=60. Collaborate on joint publications, N=56. Share copies or resources, N=60. Ask opinions on specific resources, N=60. Collaborate on shared information resources, N=59. Share own notes, N=60.

Researchers perceptions and normative views

The respondents were also invited to state if they agree or disagree with a series of statements related to scholarly work and collaboration. The results suggest that most of them (more than 70%) would be interested to know which scholars in their field work on a particular source or research question (54 respondents), they would share interesting resources with colleagues if they were allowed to (51 respondents), they would be prepared to share information with colleagues on the sources or research questions they work on (48 respondents) and that they would like to work in collaboration with others towards joint publications or common research results (47 respondents). Almost half (or 30 of 61) of the respondents regard copyright or privacy issues as important obstacles for their research, while the rest are divided between disagreeing with this statement and being uncertain about this issue. On the other hand, most respondents express that they don't find journal papers and books more trustworthy than online publications (32 of 60 respondents), that they don't find resources in a physical archive or collection more trustworthy than those in a digital archive (42 of 60 respondents), and that they don't find paper finding aids more trustworthy than online finding aids (44 of 60 respondents).
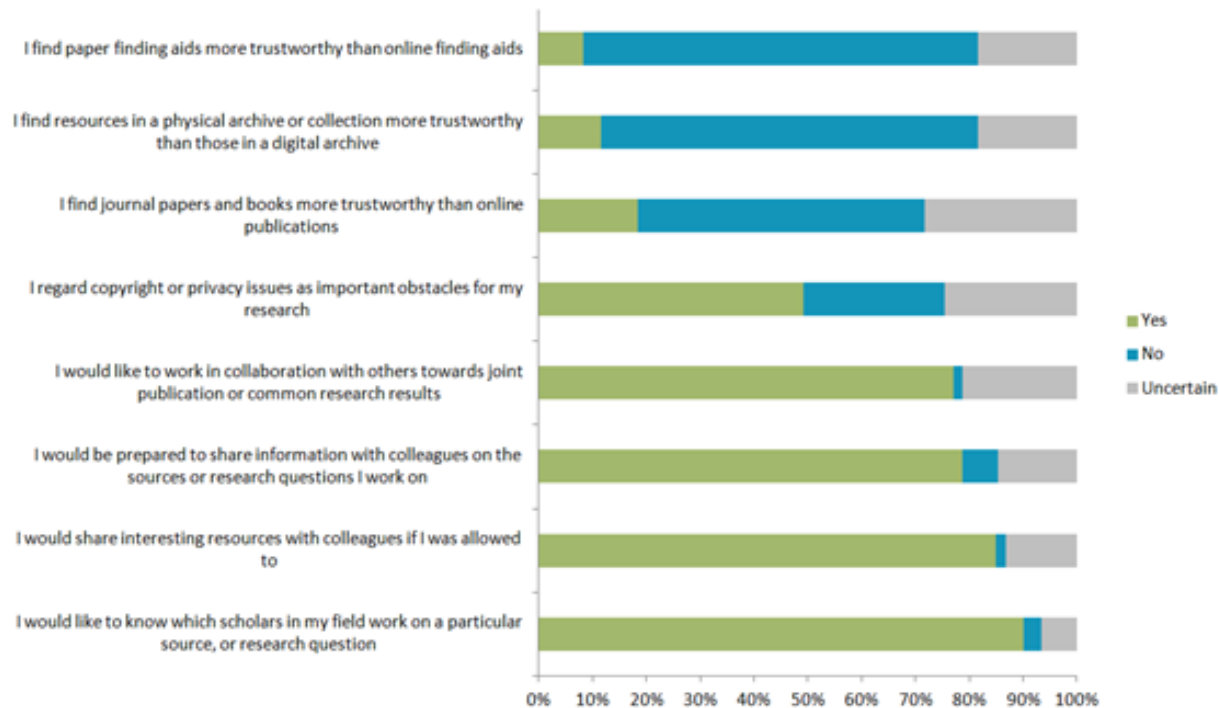
Figure 14. Researchers perceptions and normative views (Specify how much you agree with each of the following statements). Paper finding aids …, N=60. Resources in a physical archive …, N=60. Journal papers and books …, N=60. Copyright or privacy issues …, N=61. Work in collaboration …, N=61. Prepared to share information …, N=61. Share interesting resources …, N=60. Know which scholars …, N=60.

Assessment of the importance of functions related to the retrieval of resources

The last question of the web survey asked from respondents to rate a series of functions related to the retrieving of different kinds of resources to their local environments. According to their responses, the participants seem to consider more important the following activities (with figures given for those who considered the alternatives very or extremely important): fetching and automatically importing bibliographic references to bibliography applications (59% [32 respondents]), fetching and automatically importing search results on online digital resources to a word processing file (51% [30 respondents]), receiving automatic notifications on the existence of online digital metadata and resources of interest (54% [27 respondents]) and fetching and automatically importing online digital resources (e.g. images) to a research repository or database (33% [18 respondents]). As for metadata, the respondents seem to consider important activities such as fetching and automatically importing metadata into a search repository or database (42% [23 respondents]) and fetching and automatically reporting metadata into a word processing file (45% [26 respondents]). On the other hand, the related activities that are considered moderately important are fetching and automatically importing

online digital resources (e.g. digital images) into a word processing file (37% [21 respondents]), fetching and automatically importing metadata into a notetaking program (33% [18 respondents]) and fetching and transferring metadata in XML or RDF form into a computer (35% [20 respondents]). Finally, the activities that are considered to be less important are fetching and automatically importing online digital resources (e.g. images) into a notetaking program (22% [12 respondents]), fetching and automatically importing search results on online digital resources into Excel (24% [14 respondents]) and fetching and automatically importing metadata about online digital resources into Excel (19% [11 respondents]).
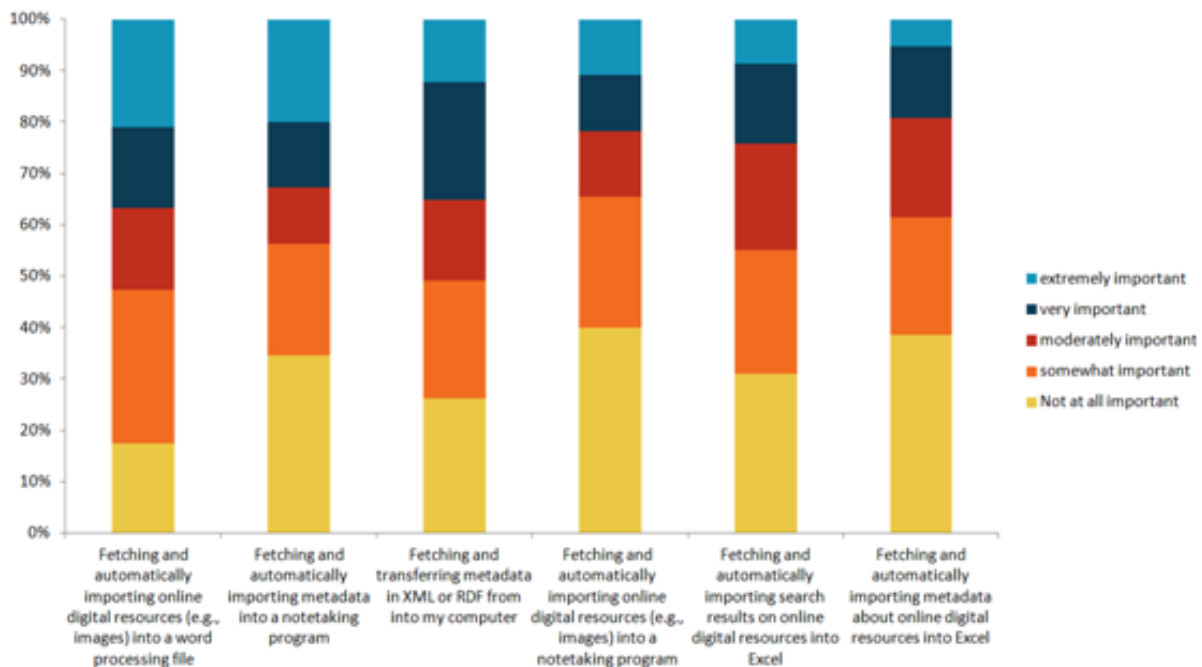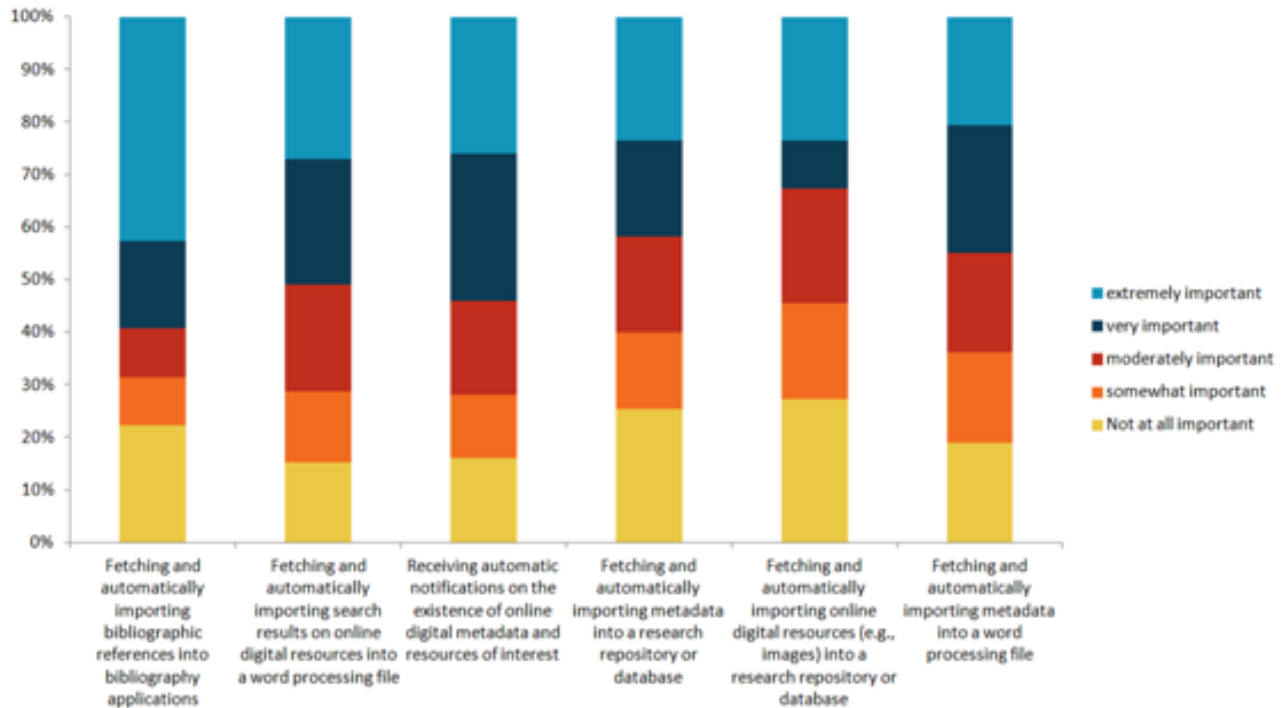
Figure 15. Assessment of functions related to the retrieval of resources (Assess how important is each of these functions to help you retrieve to your local environment different kinds of resources). Fetching and automatically importing online digital resources into a word processing file N=59. Fetching and automatically importing metadata into a notetaking program N=55. Fetching and transferring metadata in XML or RDF from into my computer N=57. Fetching and automatically importing online digital resources into a notetaking program N=55. Fetching and automatically importing search results on online digital resources into Excel N=58. Fetching and automatically importing metadata about online digital resources into Excel N=57. Fetching and automatically importing bibliographic references into bibliography applications N=54. Fetching and automatically importing search results on online digital resources into a word processing file N=59. Receiving automatic notification on the existence of online digital metadata and resources of interest N=50. Fetching and automatically importing metadata into a research repository or database N=55. Fetching and automatically importing online digital resources into a research repository or database N=55. Fetching and automatically importing metadata into a word processing file N=58.

3.4 Conclusions

This research has investigated the specific practices and needs of the Humanities and Social Sciences researchers included in the sample and contributes to our understanding of the use they make of digital tools and content. More specifically, this study adds to our knowledge about research queries, where researchers look for scholarly material and how they formulate their queries, about the use they make of specific services, the characteristics of the content and the properties of a resource that are considered important by the researchers while they seek relevant information for their work, the way researchers organize their published as well as their unpublished materials, the annotation of resources, the collaboration with others while working on a research project, and finally their needs regarding specific functions that might help them

retrieve different kind of resources to their local environment. Overall, the content of the questionnaire covers most of the fundamental practices comprised in the research process, and more particularly the main "scholarly activities" identified by Palmer et al. (2009) which are searching, collecting, reading, writing and collaborating.[34] Although the study does not provide much information for some steps of the research process, such as writing, it offers detailed account of user behaviour and priorities regarding other activities, namely collecting and collaborating, which are more directly related to digital research practices. This have been said, future related research should concentrate on other aspects of the scholarly research process in order to have a full account of the behaviour and needs of researchers.

The results are significant in three respects: they provide a detailed account of the searching and collecting processes of researchers in Humanities and Social Sciences, they draw attention to the ways researchers organize, study, annotate their resources and collaborate with each other and finally they provide insights about the needs of the researchers in Humanities and Socials Sciences about a series of statements regarding collaboration, sharing, the use of online publications or copyright issues, thus providing a useful account of current perceptions and opinions about key issues concerning the digital scholarship community.

For more concrete recommendations pertaining to user requirements, please see Chapter (7) below.

---

[34] Palmer, Teffeau and Pirmann (2009). Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development, *Report commissioned by OCLC Research*, pp. 9-15. Retrieved from www.oclc.org/programs/publications/reports/2009-02.pdf

4. Humanities and Social Sciences Case Studies

4.1 About the task: purpose and method

The work feeding into this deliverable D1.3 greatly benefited from the conclusions presented in Deliverable 1.4 on content priorities for Humanities and Social Sciences research communities (D1.4)[35] - itself the result of work carried out in Task 1.2.2 Matching Europeana content to the Research Communities. The aim of this task was to match existing content in Europeana and The European Library to the humanities and social sciences research communities . The report built on the outcomes of tasks 1.1.3 Research Communities identification and definition (D1.1) and 1.2.1 Desk research into the state of the art on scholarly content use (D1.2).

Many of the D1.4 report's recommendations have been further explored in this deliverable:

- D1.4 noted the importance of defining the intended user group of the data available in Europeana in order to determine the service level to be provided by Europeana Research, an approach that was adopted for this report.

- In the course of the tasks underlying the report, the goals of Europeana Research and the plans for its future development were, as recommended by D1.4 further define. The proposal to develop collection descriptions (see pages 178 below) will go towards the need for contextualisation of data that was signaled in D 1.4.

- The need to provide high quality content, improve the handling of multilingual resources and metadata are part of the recommendations for the Europeana Content Strategy as a whole and the content offerings for Europeana Research in particular.

- D1.4 explored which types of content would be desirable, and found that full-text, sound and video would be needed, that more data could be sourced from Eastern European countries, and that the 19th and 20th century are relatively underrepresented in Europeana.

---

[35] See Europeana Cloud Deliverable 1.4
http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.4%20Content%20priorities%20for%20Humanities%20and%20Social%20Sciences%20research%20communities.pdf

- Both D1.4 and the current report are clear that research communities interacting with Europeana Research will require tailored research corpora with critical mass focused on specific subjects. D1.4. concluded that readily available content for the humanities includes history, musicology and philosophy. For the social scientists: political science, world politics, ideology and propaganda; social history, sociology, (popular) culture, religion, possibly genealogy.

- D1.4 expressed concerns about availability of data through an API only and the development and sustainability of tools developed by each group of academics engaging with Europeana Research data. The current report follows up on the use of tools in this chapter 4 and APIs in chapter 5 below.

- D.1.4 concludes that the target audience for Europeana Research is fluid with ever changing research questions being asked by interdisciplinary groups of scholars in an ever changing mix of academic disciplines (historians and economists, or art historians and scientists, or social sciences and sociology students, etc.), and with an ever evolving set of skills, including those related to data mining. With the current report, Europeana Research has taken the first important steps on the path of constant communication with their scholarly users, in order to work with them to prepare the best tailored service based on an extremely rich but variable resource.

For the purposes of the current report, and according to the Description of Work, sub-task 1.3.5 aimed to "identify, study, document and analyse practical case studies of actual use of innovative digital tools and services, in the Humanities and Social Sciences respectively."[36] This fell within Work Package 1's task 1.3, which was concerned with investigating user requirements; in particular their digital research practices, and the tools and content types that users of Europeana might typically employ.

The purpose of this task was to investigate what researchers use innovative digital tools for when they work with the types of material that a future Europeana will contain. We achieved this by analysing actual cases of how researchers within three disciplines work with three innovative digital tools. The three cases were selected through what can be referred to as *purposive*

---

[36] Description of Work (Annex I) for Europeana Cloud: Unlocking Europe's Research via The Cloud.

sampling,[37] that is, the cases are selected as illustrative specimen to provide a useful basis for the discussion. It is important to note here that the qualitative approach employed aimed to offer general insights into researchers' activities through detailed investigations of particular instances, as well as the implications of these insights on the future development of Europeana and Europeana Research. It should be noted here that these instances do not constitute a review of digital tools or research practices, nor should they be taken to be representative or comprehensive.

T1.3.5 within the context of other tasks in Work Package 1

The case study as a methodological approach involves multiple sources and methods of gathering data that allows for an integrated and in-depth analysis. The case study of T1.3.5 provides a useful complement to the three methods previously applied in Work Package 1, namely desk research (T1.1.3, T1.2.1, T1.3.1, T1.3.2), expert forums (T1.4), and web surveys (T1.3.4)[38]. Through a case study approach, some research practices and activities were illuminated from a new perspective and in more depth than has been done in the other tasks.

Methodology: Definition of Task 1.3.5

Case studies employed a mixed-method approach to a varied material and although they are not possible to generalize to populations, they are generalizable to theoretical and analytical propositions[39]. For an investigation of "actual use", case studies thus offered a highly effective way of approaching the material.

Two other central concepts in the Description of Work for T1.3.5 were "innovative" and "digital tools and services". Digital tools and services (henceforth referred to simply as *digital tools*) have been defined in an earlier Europeana Cloud Deliverable.[40] That report emphasises the multifaceted nature of digital tools. In short, they assist the researcher in various ways at various

---

[37] For a more thorough discussion of this, see Teddlie & Yu,"Mixed Methods Sampling: A Typology with Examples", *Journal of Mixed Methods Research*, vol. 1:1 (January 2007), pp. 80-83.

[38] User-requirements and tools have also been investigated within Work Package 3, as they seek to develop tools and services that can be used within Europeana.

[39] Yin, *Case Study Research: Design and Methods*, (5th ed.), Thousand Oaks, CA: Sage, 1994, pp. 40-44. See also Flyvbjerg, "Five misunderstandings about case-study research", *Qualitative inquiry*, vol. 12, no. 2, 2006, pp. 219-245.

[40] Europeana Cloud Deliverable 1.2 State of the art report on digital research practices, tools and scholarly content use, pp. 38-39.

stages of the research process. Specifically, digital tools (a) facilitate existing research procedures, (b) enable new research methodologies, and/or (c) give rise to completely new research questions.

An innovative digital tool is, in this deliverable, taken to refer to a tool that either performs functions that were previously unavailable, or that performs already available functions in a qualitatively different way. Examination of how researchers used the tools required them to be well acquainted with the tools, which, along with the inclusion of research publications as material, meant that there was a limit to how recent the tools were that could be included in the cases.

The case studies of T1.3.5 focused on the Europeana "content types" (image, text, sound, video, and 3D – the final type is somewhat vaguely defined in Europeana, however, and represents only a small proportion of the content. For those reasons, it has been largely excluded from this study). The actual and practical use of the tools should thus involve material of types corresponding to the Europeana content types. Rather than trying to find researchers and tools that work with current Europeana material for this task, the cases selected aim to investigate tools of kinds that might be used in future research on Europeana content.

Definition of the cases and research questions

The case study research approach involves a variety of data collection and analysis techniques allowing for in-depth analysis and representation of complex phenomena.[41] Even though the approach lacks a strictly defined set of rules and procedures, it is a suitable methodology for combining different data sources and research procedures for in-depth analysis of interrelated aspects of social phenomena. This makes the case study useful for examining innovative tools and services in social sciences and humanities research.

One of the most comprehensive and widely cited works on case-study research methodology is Robert Yin's "Case Study Research",[42] which discusses a number of central components in the research design of case studies. For the purpose of the present study, two of these are of particular relevance. The first of these is the study's questions, which are of great importance since they will have implications on every other aspect of the study (for example, the cases

---

[41] Yin. "Case Studies", pp. 9-19.

[42] Ibid.

selected in this task will be presented following the structure of the research questions). The second component regards the unit of analysis. At this stage, the researcher has to define what the case is, e.g. an individual, an event or some kind of process. These two components are highly interrelated as the nature of the question(s) must depend on the unit of analysis, and vice versa. The following sections define the cases and describe the questions for the case studies of this task.

Definition of the case

Three disciplines, History, Sociology and Education, were selected as they were likely to work with material from, or similar to material available through, Europeana, that is, cultural heritage material of the four major Europeana content types. Moreover, History and Sociology were selected because they are fairly typical for the fields of Humanities and Social Sciences, respectively, whereas Education is a research field that straddles the border between the two fields.

The cases themselves centre on one particular tool and the way in which a researcher or group of researchers make use of that tool. The case study thus focuses both on the tool and its documentation as well as researchers' experiences of using the tool with one or more types of research material in several stages of research.

A broader perspective on the case and its relation to Europeana is offered through the satellite cases – examples of application of alternative tools to those selected for T.1.3.5 case studies. As such, two alternative tools that support or facilitate similar scholarly activities have been selected, based on the information provided by the researchers interviewed while collecting information for the primary cases.

Research questions

The application of an innovative tool in research can be characterized by a variety of interrelated dimensions involving the research traditions in a particular scientific discipline and the role of digital tools in it; new research trends and the digital tools; functionality of the tools and skills and experiences of the users; compatibility of the innovative tools and other tools typically used by the researchers in the discipline etc. The overarching research question is therefore how researchers in particular disciplines use digital tools, and it can be broken down in five related sub-questions:

- **How does the tool work?** Data types that the tool works with; functionality of the tool in different stages of the research process; the level of competences needed to work with the tool etc.

- **In what way has it been used?** Disciplinary background of the users, types of data they have worked with; description of the research project(s) the tool has been used in; the scholarly activities the tool has been used for, etc.

- **What are its strengths and weaknesses?** The motivation for choice of the particular tool; the problems encountered while using the tool, improvements needed for the tool and the gaps in the discipline that could possibly be addressed by new tools.

- **What is its further potential**? How could it be used with Europeana-type material? The potential of the tool in applying it to Europeana material; the changes needed for the tool and for the material.

- **What kind of new research (methods) does it facilitate/enable?** The issues and questions in the discipline that can be addressed better by using the tool.

The satellite (alternative) cases aimed to answer the following questions:

- **How has this alternative tool been used in the discipline in relation to the main tool?**

- **What are the strengths and weaknesses of the alternative tool** in the discipline, working with different data types, etc.

- **How can the alternative tool be potentially used with Europeana material?**

- These questions have been elaborated in data collection instruments (as interview questions) and explored in the analytical process in data material.

Selection of cases

From each discipline, one digital tool was selected: NodeXL for Sociology, Transana for Education and HyperImage for History.[43] Suggestions for appropriate tools were solicited largely

---

[43] Thus we are in compliance with suggested sample size of 3 to 5 case studies in order to be able to perform analytic generalizations based on the results. For further discussion on sampling schemes and sample sizes, see, Onwuegbuzie and Collins, "A typology of mixed methods sampling designs in social science research", *The*

by querying researchers in our network, not least those who had in some way assisted in previous tasks (in particular the Expert Forums). The researchers participating in the Forum were selected by the relevance of their research to potential use of Europeana material. Other aspects taken into account in selection of cases were related to the available documentation of the tool and its application in research; coverage of Europeana content types; and relation of the case to new trends, new research environments and/or new research questions arising in the disciplines.

The application of the tool should be documented in publications. By its nature, a "cutting-edge" innovative tool is used by few researchers and there would be little if any documentation available while the well-documented tools may be considered mainstream rather than innovative. The tool selected for the case study needed to be new to researchers in the discipline but at the same time fairly well documented in publications.

It should be noted that no attempt was made to identify cases that engaged with current Europeana material, for two reasons. Europeana as it stands today, does not offer content and metadata of sufficient quality to make it an attractive tool for research or source of research material[44]. Moreover, identifying researchers, in particular social scientists, who worked with Europeana content as well as put digital tools to innovative use in their work quickly proved more than challenging (it would be hard if not impossible to identify such use, in particular in disciplines that are not currently involved in cultural heritage material).

More importantly, the tools that were selected are such that they – or tools with similar functionality – can quite conceivably be used for future research on Europeana content. The forum reports that an important feature of a future Europeana could be the ability to export content to third-party tools for analysis, and the three tools are the kind of tools to which Europeana could be expected to export content. It is of course impossible to predict exactly which particular programs will be used with Europeana even in a few years' time, but Transana, NodeXL, and HyperImage all belong to categories of tools that would likely be employed with exported content.

Qualitative Report, vol. 12, no. 2, 2007, pp. 281–316. http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf (Accessed: 18 February, 2014).

[44] Europeana Cloud Deliverable 1.5 (1 of 4) Expert Forum Case Studies Report; Deliverable 1.5 (2 of 4) Expert Forum Tools & Content for Humanities Research Report; Deliverable1.5 (3 of 4) Expert Forum Tools & Content for Social Sciences Research Report.

The three tools finally selected all included functionality that enabled or facilitated what was broadly considered innovative research methods and covered the major Europeana content types as well as metadata (it should, however, be emphasized that the tools selected for T.1.3.5 should not be taken to represent the whole range of the tools available for use within the disciplines). Transana facilitates collective work with audio and video transcription and qualitative analysis; NodeXL enables mixed-method network analysis; and HyperImage is a tool that makes it possible to annotate images and parts of images. For each of the tools, an appropriate researcher or group of researchers was chosen (mainly on the basis of publications or previous contacts) for closer study.

For Education two alternative tools that support or facilitate similar scholarly activities were selected in the course of the data collection process: NVivo, a commercial and well-established tool in the research community, and Voyant, an innovative web-based text reading and analysis environment. These tools form the basis of the two satellite cases.

Data collection

In order to capture the multifaceted nature of how digital tools and services are employed in research, the uses of innovative tools were studied by integrating data from different sources. Yin lists six sources of evidence (documentation, archival records, interviews, direct observations, participant-observations, and physical artifact),[45] of which three can be applied to the subject under examination here:

- analysis of published documentation (of tool and research applying the tool)

- semi-structured interviews with researchers

- observations (experiences) of how the tool works.

For each case, information was gathered via semi-structured interviews with the researchers, desk-study of the available documentation of the tool in research as well as observations of the demo versions or similar examples of how the tool works. Six semi-structured interviews were made between November 2013 and January 2014 (1 for Transana, 2 for alternative tools in Education, 2 for NodeXL, 1 for HyperImage). They were face-to-face interviews (4) and web-based interviews (2), and lasted about one hour each.

---

[45] Yin, "Case Studies".

Analytical process

The semi-structured interviews were recorded. The audio files were transcribed and coded using different methods and tools:

- transcribed and coded using Transana (for the Transana and NodeXL cases)

- transcribed and coded in MS Word and NVivo, (for the satellite cases)

- transcribed and coded in Windows media Player and MS Word (HyperImage).

Published research was reviewed in order to present the novelty and value of the tool in the discipline (what types of research questions are researchable because of the tool, what new research questions can the use of the tool facilitate and other), as well as the stages of the research in which the tool has been used, and the data import and export procedures.

Observations of the tool in function were made by using the documentation of the tool, personal experiences with the tool (HyperImage, NodeXL, Transana) using sample data sets (HyperImage, NodeXL) and actual research material (Transana). These were of particular importance in discussing the research activities supported by the tools and their possible applicability to the Europeana content.

Data was analysed and presented to provide answers to the main case study questions and highlight the aspects of application of the digital tool and its relevance to facilitating use of Europeana content. Finally, analytic generalizations were made, where appropriate.

4.2 Case Study 1: Education and Transana

Transana is a piece of CAQDAS (computer-assisted qualitative data analysis software), developed by a group of researchers at the Wisconsin Center for Education Research in Madison. The program is intended mainly for researchers working with video data but also works with audio files. It is at its best when used to analyse large quantities of data, for example a series of interviews or classroom videos, since its main analytic tools are a range of means of categorising smaller parts ("clips") of the files into different analytic categories, creating possibilities to look at how a specific phenomenon is distributed across the data or to discover how a systematic relationship between different events recurs in varied settings. Since the program gives a very basic analytic structure without any pre-defined analytic conceptions it is very adaptable as far as theoretical frameworks are concerned.

The case centres on how Transana is used by Dr. Thomas Hillman at the Department of Education, Communication and Learning (University of Gothenburg, Sweden). Hillman's research is based around the way in which people adapt technologies to use as part of their learning processes, in formal as well as informal educational settings (the latter would include for instance museums and zoos). (For a list of publications, for which Hillman has used Transana as an analytical tool, see Appendix I – Documentation of Transana.)

Functionality and features: Data handling

Transana handles a wide range of audio and video source file formats. On the output side format versatility is more moderate but the formats of choice are very useful. According to the Transana project's webpage:

Transana 2.50 supports MPEG-1, MPEG-2, most AVI video, QuickTime MOV, MP4, and M4V formats, as well as MP3, WAV, and AAC audio on both Windows and OS X. Windows Media Video, WMV, and Windows Media Audio, WMA, formats are supported on Windows only. Older versions of Transana may not support all of those formats.

Transana versions from 2.50 onwards feature a media converter based on ffmpeg, which can convert a number of formats into more Transana-friendly ones. It is not clear which formats are and are not supported by the converter, but according to the screencast devoted to the tool the aim is to cover as many formats as possible.

Since Transana uses audio and, in particular, video files as data sources there is a potential steep learning curve at the beginning as importing video files into Transana may prove difficult. This has to do with the fact that digital video comes in a multitude of formats many of which are container formats and as such may contain data in varying states and types of compression, resulting in files with the same suffix behaving differently for no apparent reason.

Several data types can be exported from Transana. Transcripts may be exported in either RTF (.rtf) or XML (.xml) format with or without the time codes. The various kinds of reports that can be generated in Transana (see section 3.2.3, below) may be exported as RTF files and the visualisations (see section 3.2.3, below) can be exported as JPEG (.jpeg/.jpg) image files. Notes can be exported one by one as text files (.txt) but are probably more useful included in a report where their context is obvious. It is also possible to export an entire database into an XML file for later re-import into Transana.

Single- and multi-user functionality

There are three different versions of Transana: a single-user version, a lab version, and a multi-user version. The single-user and lab versions allow only one person at a time to work on any specific dataset. On the other hand, the multi-user version enables several researchers to analyse the same data simultaneously and from different locations by means of a MySQL database that stores the analytic data. These data are thus instantly accessible to all researchers who have access to the database and any changes they make are in turn instantly transmitted to the database and thereby to all other researchers. The database holds only the analytic data, which means that each researcher needs to have a local copy of the source data file and a Transana client installed on his/her computer.

Functions of the program

Transana's analytic tools provide a non-destructive and theory-independent way of looking for patterns in large data material. By dividing audio or video material into smaller chunks called "clips", tagging the clips and/or putting them into categories it is possible to discover and visualise patterns in the material.
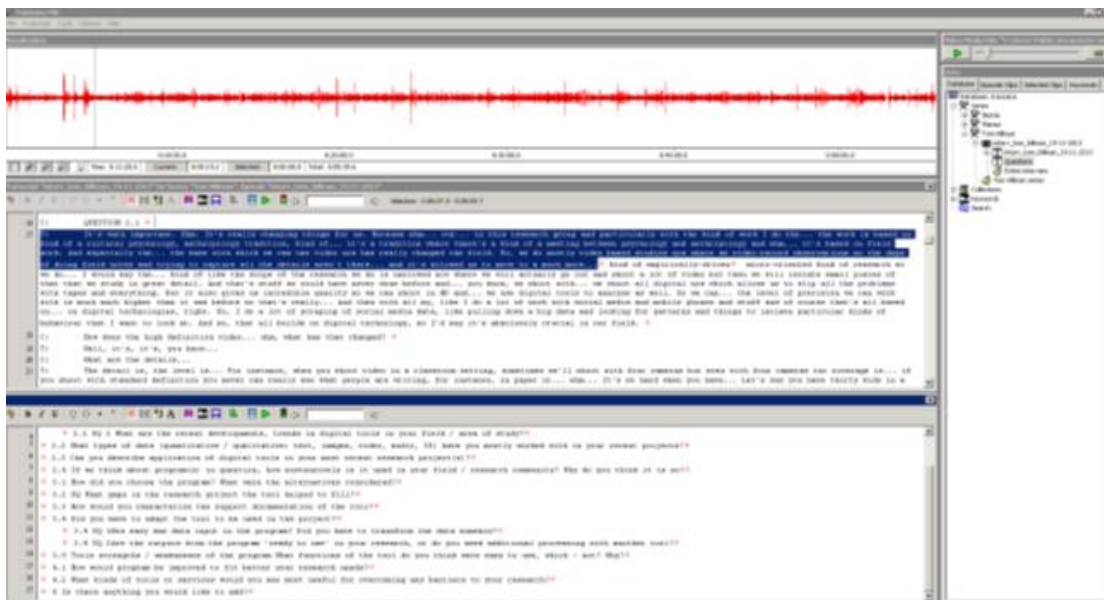
Figure 16. The main interface of Transana. At the top is a visual representation of the audio. Beneath it are two transcripts the upper of which shows the clip being played marked in blue. To the right is a content browser for the database that gives access to recordings, transcripts, keywords, collections etc.

In theory there is no upper limit in the number of analytic levels which can be employed in Transana. Media files can be divided repeatedly according to varying principles since it is possible to work in several transcript files at a time. Each transcript file contains references to time points in the media file and the period between two such references is a clip. Seeing as the references reside in the transcript files and do not intrude in the media file this process of division is non-destructive to the media file and thus infinitely repeatable.

This simple process gives the researcher the freedom to decide which theoretical framework is to be employed in the analysis. Once the decision has been made, the clips are tailor-made by the researcher to fit the model. Other ways of categorisation are also theoretically unbound. For example, every clip needs to belong to a user-defined category. Furthermore, it is possible to tag clips with equally user-defined keywords which in turn belong to a different set of user-defined categories.

Transana leaves researchers pretty much on their own when it comes to methodological decisions. What to code for, in what order to do it, which kinds of transcriptions should and should not be made etc. are questions that the researchers need to have answers to, at least preliminarily, unless they are prepared to spend a great deal of time on trial-and-error.

Transana's search function turns the keywords into a powerful analytic tool by enabling the researcher to perform complex searches based on them. This is one way of finding patterns in the data. Another way of looking for patterns are the visualisation functions of Transana. They allow the user to create different kinds of visual representations of the distribution of keywords in the data.

Transana offers the user the possibility of adding notes to different layers of the database (to collections, keyword groups and so on). The researcher can thus document the analytic process in-program, gathering all the information about analytic decisions and insights in one format and in one place.

The user can get an overview over the keywords, categories and notes by means of software-generated reports. These reports tell the researcher which categorisations have been made and why.

Using Transana for qualitative research

The research projects that Hillman describes are characterised by a large amount of data in the form of video footage. Hundreds of hours of video captured with several cameras provide the bulk of the material. The first task, once the video files have been ingested into Transana, is thus to gain an overview of the material. There is a triage process, rough coding of the data files that provides a structure of the material that facilitates the qualitative analysis:

We'll tend to do this like kind of triage coding, so we get like a hundred hours of video and then we'll go through and do [. . .] a content log of what's in it. [. . .] you'll end up with like essentially metadata, [. . .] maybe five to ten tags of what's going on so that the team can quickly go in and find all the pieces that have this kind of activity in them. [. . .] [And] that gets us into the material as well, so that tends to be the first thing we do once we've been out [in the field].[46]

A concrete illustration of this method can be found in a recent article.[47] There, Hillman describes how he uses first triage coding, and "initial sorting" where he uses Transana to mark instances where participants engage in the particular activity that he wishes to analyse. He then proceeds to add descriptive keywords to these instances. By using these keywords, he can search across all his material for similar activities and analyse them while maintaining a broad perspective.

Hillman also observed how this coding is similar to a cataloging process, turning the material into a library of analytically relevant segments. This part of the process also benefits from the collaborative functions of Transana. When several members of a team work with triaging or coding the same data in a joint database, synergies are created, because multiple people are coding the same material but also because "you can look at what someone else has found and add your own tags to it."

In relation to a research project that investigates the inclusion of handheld digital technology in maths teaching, Hillman explains how he makes audio-visual recordings of interviews and classroom situations. He then uses Transana to synchronise transcripts and recordings. In the method section of the publication, the importance of this connection between transcript and data

---

[46] For reasons of brevity no full transcriptions are included in this document.
[47] Hillman, "Tracing the construction of mathematical activity with an advanced graphing calculator to understand the roles of technology developers, teachers and students", *International Journal for Technology in Mathematics Education*, (forthcoming).

file to the analysis is emphasised: "This freedom to move between recording and transcript while maintaining synchronisation afforded flexibility to the analytic process during which multiple passes of each interview and video-recording were made."

Hillman's descriptions of how he uses Transana focuses on what he refers to as "the front-end of the process", the preliminary or early part of analysing the material. For other tasks, such as detailed video analysis and subtitling, he relies on other software. Within the educational sciences, there is the additional problem of anonymisation, and while Transana offers the possibility to take screenshots from a video file, such images must generally be treated further in for instance Adobe Photoshop in order to obscure faces or turn the image into a cartoon-like drawing.

Not every researcher employs other programs to supplement Transana, according to Hillman. Some people use it as their only analytic tool, depending on their level of research and unit of analysis. With a fairly broad unit of analysis, looking for major features such as structural elements of particular activities, Transana would provide sufficient functionality, he claims.

Transana's strengths and limitations

The main strengths of Transana as an analytical tool that Hillman as well as other researchers point out has less to do with particular affordances than with the program performing two of its defining features well: Transana is a tool for collaborative qualitative analysis of video and audio data; it is described as being good at doing just that. Hillman explains how the software is more video-based than other CAQDAS tools, and emphasises how easily the researcher can input, manage, and triage large volumes of video data. Dempster and Woods also note how quickly they organised, converted, and ingested their video data before analysis.

A particular strength Hillman experiences is that there is no particular need to transcribe the files before beginning to analyse them. Dempster and Woods discuss the different needs for verbatim transcription and gisting, and point out the usefulness of using several transcripts of the same file for different types of transcriptions or analytical annotations, which can then be looked at in parallel.[48]

---

[48] Dempster and Woods. "The Economic Crisis Though the Eyes of Transana" in *FQS*, vol. 12, no. 1, 2011, Art. 16, §11.

The basis for Transana's transcriptions and annotations is the use of time points. These are used to connect specific parts of the transcript or specific annotations with the data file. The data (video or audio) are never altered, so clips can be easily extended or shortened according to need, and there is no need for real-time file-sharing: each collaborator could have their own copy of the data file (as long as the file path is identical for all local copies).

The collaborative functionality of Transana is also considered one of its strong points. The transcripts, clips, and keywords are stored in a relational database, accessible to all members of a project, and Hillman explains how the collaboration with and sharing of transcripts, keywords, and analytical notes (a built-in function) creates synergies throughout the research process. The simplicity with which collaboration can take place is increased by the fact that the Transana client can run in Windows, Macintosh, and Linux environments.

Points that were particularly in favour of Transana, according to Hillman, were that the program was developed by researchers – he mentioned how the developers were very responsive to his questions and comments – and that it was developed as an Open Source project. The developers' responsiveness even stretches as far as implementing changes requested by researchers using Transana.

Hillman also discussed how, once you have understood the possibility of breaking down data by applying keywords to as well as categorising clips, there are "a million different ways" in which to use them, making Transana a very versatile tool, allowing for either a deductive or inductive analytical approach. This flexibility also shapes the way analysis is carried out, however, because "those coding tools are fixed and so you end up looking at the data for ways that you could add keywords or put clips into categories."

Apart from some technical drawbacks (he refers to the system as "fragile" without specifying exactly what he means by that; the writers of this report have not encountered any propensity to crash or lose data, for instance), Hillman's main criticism of the program is again tied to the fact that it its main functionality is connected to the qualitative analysis of video and audio. It does not allow text to be used as a data source and although images can be inserted into the transcript, using images as data source requires a work-around:

I made a video slide show of the photos I wanted to have in there and I created a blank audio file to go with them so I could code them and then inserted time codes for the changes in the

slide show. So that allowed me to create a fake transcript that actually had descriptions what was going on in the photos [. . .] [Transana] wants a time-based material all the time so if you have something static you have to create [a time-based version of it].

The problem of integrating text and images into the Transana system, for instance, can be taken to be an unfortunate shortcoming if the research material contains other than audio-visual material. In Hillman's (and presumably many other Educational science researchers') case, the use of recorded interviews and classroom sessions along with handouts and other teaching material, and even written assignments or diaries by pupils, is not uncommon.

Hillman also mentions how Transana is limited for a really detailed analysis of video material, as "it doesn't really allow you to do particularly detailed break-downs of small segments of video." And although Transana allows you to have multiple transcripts (capturing different aspects of the material) and display them in parallel, he would have liked a greater level of integration between these various transcripts and the data source. Currently, he uses ELAN when the analysis requires this.

Transana's further potential and future use with Europeana material

Earlier reports in this project have drawn attention to two areas where further development of Europeana is believed by researchers to have the potential to be especially beneficial. These areas are the enrichment of metadata and functionality for exporting Europeana content to third-party digital tools.

Enrichment of metadata is crucial to making the Europeana content researchable since richer metadata allows researchers to make a useful selection of data for a research project. Once the data to be used has been selected it needs to be exported in a format that the researchers can use in their analytic software of choice. This section will discuss these areas from a Transana perspective, but the discussion can easily be extended to apply to other, similar, digital tools.

Enrichment of Europeana metadata

Transana, like other CAQDAS software, adds new metadata to the material during the research process, metadata that can then be exported. If such enriched metadata can be imported back into Europeana to add to the analysed objects' metadata records, they could, conceivably, provide a valuable resource for researchers using the Europeana material.

One of the exportable data units in Transana are transcripts, which are exportable as RTF or XML files. Such transcripts could, if they were made searchable, greatly enhance Europeana's metadata and would add to the accessibility of audio and video files. Since Transana transcripts can contain all kinds of information, rather than being only verbatim, they are a potential holding place for a great variety of metadata, structured as well as unstructured.

Verbatim transcripts could also form the basis for making the material accessible in other languages by being the source for translations. If the time codes are maintained throughout the Europeana side of the life cycle, the translations would also enable researchers unfamiliar with the original language of the audio or video file to identify relevant clips.

Another way of incorporating Transana-generated metadata into Europeana could be to find a way to use the data behind Transana's visualisations to produce visual mappings of the content of a certain piece of audio or audio-visual data.

Exportability of Europeana material

Central for researchers who want to use Europeana material is that it is importable into whatever analytical software they use. In the case of Transana, and as far as the audio and video data are concerned, this means keeping to formats and encodings that are supported by Transana or can be converted in its media conversion tool. There is no exhaustive documentation to indicate which formats are supported by Transana, mainly because the multitude of possible combinations of formats and encodings make this unpredictable and laborious to test. Transana's developers recommend MPEG-1 or MPEG-2 but it is probably a good idea to keep Europeana's restrictions on formats as lax as possible in order to avoid unnecessary conversion that may result in loss of quality and also not to repel potential contributors.

Transana does not feature support for analysis of still image data, but Hillman suggests a work-around that could be easily implemented in Europeana Cloud. Hillman created a slideshow from a series of pictures and could then tag each picture in Transana. An application that creates such a slideshow from a series of user-selected image files could conceivably be incorporated in Europeana to facilitate analysis of a series of still images.

The metadata available via Europeana should also be in formats that can be easily imported by third-party software. This is to say that it should be possible to export all kinds of metadata in

human-readable text format. In the Transana case, these can then be imported into notes, keywords or into a transcript and used by the researcher directly in Transana.

Wishful thinking

Transana's keywords are associated with clips, that is, with parts of media files rather than the whole file at once. For this reason they are also useful in connection with these clips and rather than with the entire source files. Consequently, it would be ideal if the data transfer back and forth between Europeana and Transana were to take this into account, keeping the keywords' association with time codes intact and perhaps even offering a visualisation of the keywords' placement in the source file within Europeana.

Dedoose – toward a web-based analytic tool

Hillman mentioned Dedoose[49] during the interview, a web-based tool for qualitative data analysis that, according to Hillman, shows great potential to become a serious research tool in the future. The tool works independently of the user's operating system but does require a reasonably modern browser. Analytic and visualisation functionalities appear, at first glance, to be quite well-developed, offering more choices than in Transana.

It seems that a move towards more web-based tools is underway not only where the Transana kind of analysis is concerned but also for network analysis (cf. section 5.5.5). In fact, Hillman also mentions using the GoogleDocs[50] spreadsheet tool in combination with an online service called IFTTT[51] to automatically gather data from social networks like Instagram and Twitter, which can then be analysed using a network analysis tool. Web-based research tools are clearly growing in importance at the moment.

4.3 Case Study 2: Art History and HyperImage

Research into History, particularly research including images, requires a tool or suite of tools that can recognise and also annotate images in a way that allows for both qualitative and

---

[49] http://www.dedoose.com (Accessed: 6 January, 2014).

[50] http://docs.google.com (Accessed: 6 January, 2014).

[51] http://docs.google.com (Accessed: 6 January, 2014).

quantitative analysis. Traditional Art History research involving images has used print images. At best, these may have been photocopied or re-printed for annotation purposes. Digital tools are not typically involved in research practice within the field of Art History, however, they are increasing in usage. Most traditional work within Art History would instead be focused on the interpretation of the artworks rather than doing something digital with them.[52]

One tool that enables digital annotation of images is HyperImage. This tool is the result of collaborative research between the Humboldt-Universität in Berlin, Leuphana University Lüneburg, and Zweitwerk – a content software company. The tool was built in response to a significant dearth of tools for image archiving and annotation in comparison to what is available for texts. In 2007, just after the team began development on HyperImage, image annotation was limited to metadata text. The resulting software was to be a "single integrated system, whereby each stage will be based on open standards and therefore compatible with other systems.[53] "It is this claim to direct compatibility with other systems that prompted us to investigate its suitability for use with Europeana. The HyperImage software is available to download for free from SourceForge and is open source.

In investigating this particular tool, and its uses in History and Art History, we spoke to Karolina Badzmierowska, a PhD student in Art History who uses images as the main focus of her research. She uses HyperImage for annotating those images. We also looked at some pilot projects connected to the development of HyperImage.

Functionality and features

According to the HyperImage website, "[the] HyperImage platform supports the linking of (audio)-visual objects, texts and mixed-media documents".[54] The tool is web-based, which allows the user to access it from multiple computers, as information is stored online. There is

---

[52] It is an open source resource, and sits on an Omeka platform (see http://omeka.org) as an exhibition and archive. There are one or two exceptions. For example *Gothic Past* a project based in Trinity College Dublin, is an online image archive of artifacts and architectural items of significant historic value found around Ireland.

[53] Warnke, Kuper, Helmers. HyperImage – Image-Oriented e-Science Networks. In: Max Planck Digital Library / German e-Science Conference, ID: 315522.0, 2007.

[54] http://manual.hyperimage.ws/ (Accessed: 10 December, 2013).

the option to store the information on your local server, if so required, however. Users work with the HyperImage Editor and HyperImage Server to annotate and store images and associated data.

HyperImage allows the user to "drag-and-drop" images into the HyperImage interface ready for annotation, either from their image management software (such as Windows Explorer or Adobe Bridge) or from a website. HyperImage works by allowing the user to select items, or "elements", within an image and annotate them. It is then possible to link to similar elements within other pictures. For example, in the HyperMedia project, we can see highlighted elements within the image; the annotations on that image; and the links it makes to related images (see Figure 16).
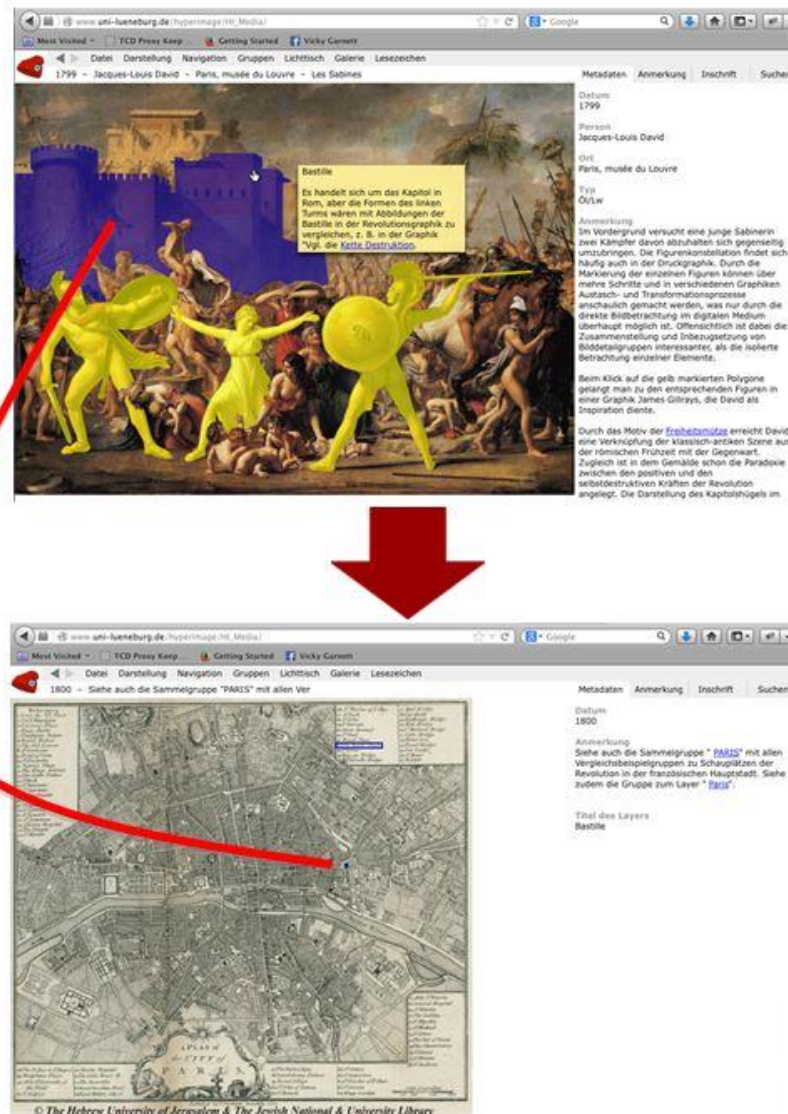
Figure 17. Annotated images in Hypermedia showing how layers in one image can link to a layer in another

As an image annotation tool, HyperImage includes space for metadata. It comes with three metadata standards out of the box: basic, Dublin Core and CDWA (Categories for the Description of Works of Art). Dublin Core is an international standard and is encouraged for richer metadata purposes. Metadata fields can be freely tailored and extended to the user's needs on a custom basis. Custom fields and further metadata standards can be incorporated using the built-in template editor. Different metadata templates can be used in conjunction to form a complex set of metadata fields.

The drag-and-drop function for populating the HyperImage catalogue allows for single and multiple images to be uploaded to the tool. These images can then be filed into groups, depending on their category type within the context of the research. External URLs can also be created as objects, which can later be linked to one or more images or layers within the project (see Figure 17).
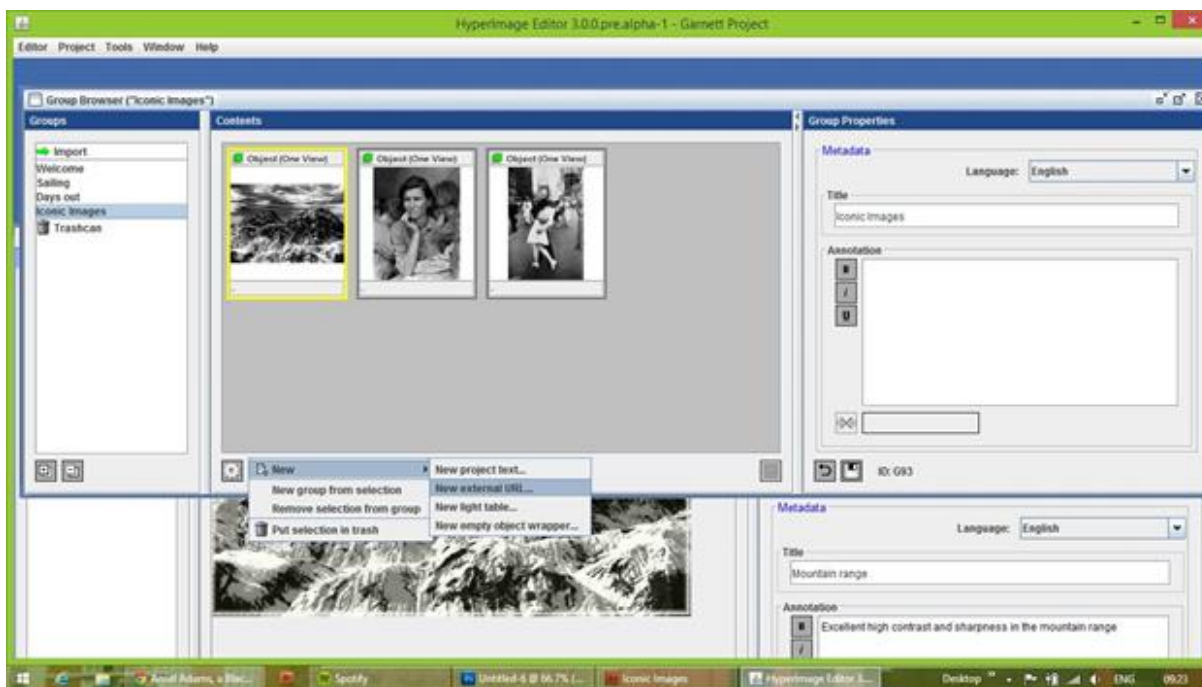
Figure 18. HyperImage Editor with dropdown menus showing how to create an external link within an object

HyperImage has its own terminology. Image files are referred to as "objects". Items within an object that are annotated, or tagged, are called "elements". These annotations are made via "layers" into which "polygons" are drawn around the shape of the element you want to annotate (see Figure 18).

Layers within images are not a new concept. Image manipulation tools such as Adobe Photoshop and Corel Draw have been using layers within images for some time. These, too, allow for changes to be made without altering the original digital image. The difference with HyperImage, however, is that these layers enable hyperlinks to be created on them. The hyperlinks can link to an explanation of the object in the image, another similar object in another image, or to an external website, thus allowing for interaction between the images and their wider context. For example, within the tool, annotations are made on the layer as polygons,

which can be colour-coded to identify individual annotations or groups of annotations around a topic within single image. Layers can have a transparency level set to prevent annotations from interfering too much with the visibility of the image.

Multiple polygons can be drawn into one layer. This is particularly useful where there are several repetitions of the same element within an image (e.g. trees) or where a single element is in separate parts within the image, as can be seen in Figure 20 where the sailor is partially covered by the girl he is holding. Links can be made on elements of images based on themes.
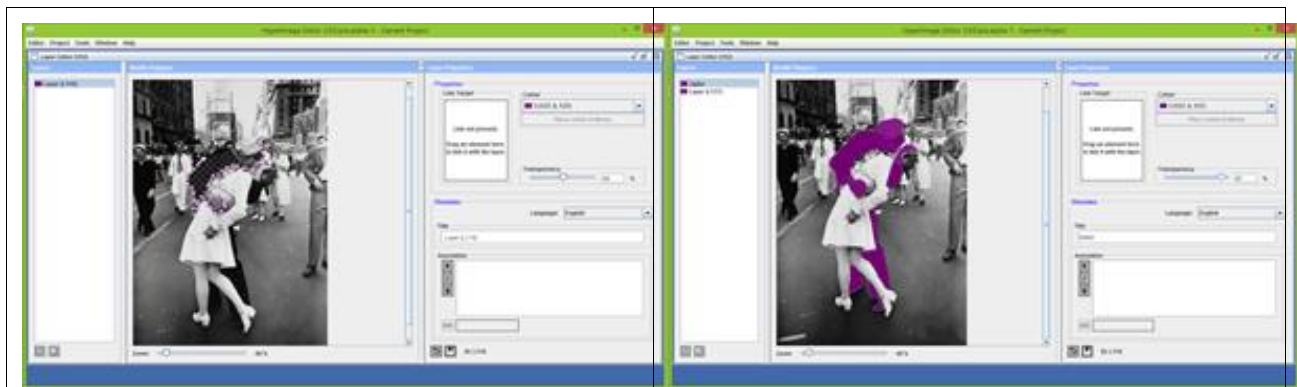
Figure 19. Polygons being drawn onto layer over digital image in HyperImage Editor	**Figure 20.** Multiple polygons drawn onto single layer over digital image in HyperImage Editor

Outputs

The outputs from the Reader and Editor can be downloaded in the XML mark-up language as a backup. This can then be used to recreate the information on a website, or user's server. This download takes the form of a directory or zip-file, which contains the full structure, metadata, and links and relationships in XML format. The images are supported as image files within the directory and are rendered in different sizes for use online. By exporting in this fashion, it allows the user to import the directory to their server, or even to a Wordpress site (or similar) for publication.

The HyperImage Reader essentially shows the fruits of one's labour as the published data. A preview mode for the HyperImage Reader is available which allows you to see how this online version develops as you annotate more and more images. HyperImage Reader is currently a separate platform, which can be used alongside the HyperImage Editor.

Using HyperImage in Art History

Karolina Badzmierowska

Our informant, Karolina Badzmierowska, is a PhD student working within Art History. She applies digital humanities techniques to her research, which looks at images within the Fagel collection. Her research requires annotation of these images, alongside the use of spreadsheets in Excel. She also uses Zotero to manage citations within her research. By her own admission, Badzmierowska had not done a great deal of research into alternative tools for digital image annotation, but had heard of HyperImage through Twitter.

Badzmierowska is using HyperImage to annotate digital images taken from the Fagel Collection, which is held in Trinity College Dublin's Long Room Library. The Fagel Collection is a single collection that takes up "a mile of shelf space".[55] The Fagel family in the Netherlands compiled it during the 18th century, comprising scientific drawings, maps and atlases, cartoons, broadsheet newspapers, and political writings. The entire collection contains over 20,000 items. Badzmierowska hopes to work on as many of the scientific images as possible to analyse them for their artistic merit, as well as thematic, iconic and scientific links. Once she has completed her annotation, she will export this as XML in the format available from HyperImage. Her overall ambition is to create an online archive that will allow the public to interact with the images, discovering links between them. Her ideal set up will be via touch screen or gesture recognition, but she realises this may be beyond the scope of her PhD.

Pilot projects

The majority of publications that involve the use of HyperImage are only available as the online output of the project. Any journal articles or conference papers on the user-experience for HyperImage have, for the most part, been written by the HyperImage team. For that reason, we looked to pilot projects for examples of uses of the tool.

To test HyperImage once it had gone through the initial stages of development, the team established a set of pilot projects. The complete list of pilot projects is too numerous to go into

---

[55] http://kalden.home.xs4all.nl/verm/fagel-dublin-lecturesENG.html (Accessed: 16 December, 2013).

detail here;[56] however, to see the basic principles of how HyperImage was used, we will look at HyperTaxon, HyperGiotto and HyperMedia.

HyperTaxon was a project dedicated to studying a small group of insects known as *Peloridiidae.* The aim of the project was to compile "existing information" about the insects in one virtual space, thereby making it accessible to members of the public as well as academic researchers.[57] The published version showcased images of the insects, the regions in New Zealand in which they can be found, and anatomical information about the creatures.[58] The HyperImage Editor allowed the HyperTaxon team to highlight different parts of the insect on the photographs (see Figure 21).
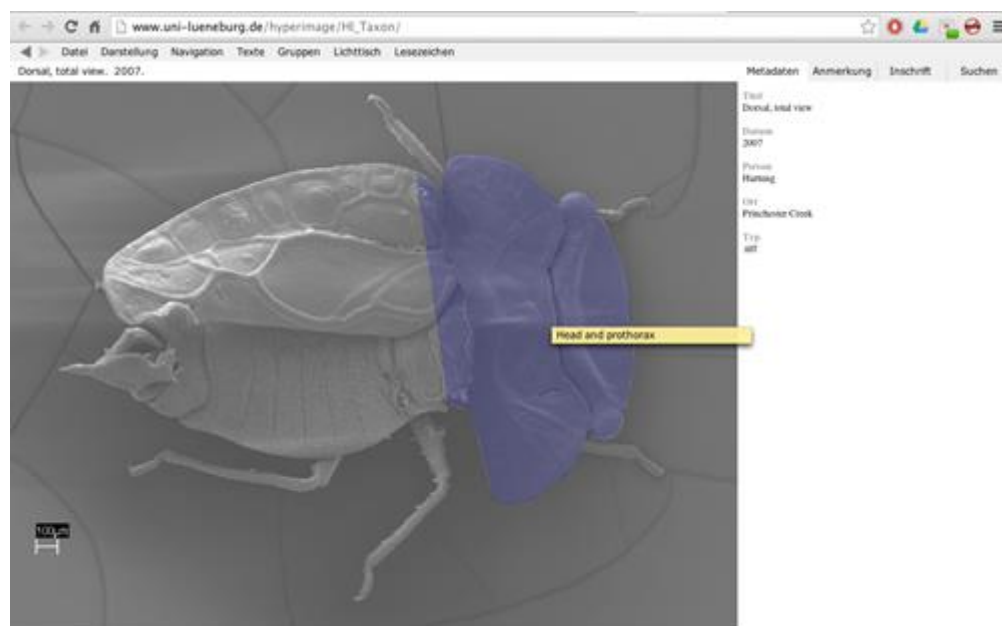
Figure 21. Screengrab of HyperImage Reader showing detail from HyperTaxon project

HyperGiotto[59] was an Art History project, which looked at the works of Giotto and his contemporaries from the Trecento period (14th Century). It investigated thematic relationships

---

[56] A full description of the other pilot projects associate with HyperImage can be found at: http://www.uni-lueneburg.de/hyperimage/hyperimage/beispiele_E.htm (Accessed: 6 January, 2014).

[57] http://www.uni-lueneburg.de/hyperimage/hyperimage/hyperTaxon_E.htm (Accessed: 6 January, 2014).

[58] http://www.uni-lueneburg.de/hyperimage/HI_Taxon/ (Accessed: 6 January, 2014).

[59] http://www.uni-lueneburg.de/hyperimage/hyperimage/hyperGiotto_E.htm (Accessed: 6 January, 2014).

among image data and metadata, using a combination of three tools: HyperImage, Mneme[60] and the Census of Antique Works and Arts Known in the Renaissance.[61] The project looked for similarities in elements within images of antiquities and picture that were then translated into sculpture in sarcophagi or friezes, or vice-versa. An image of a carving on a sarcophagus from AD 190 is repeated again in a sketch in the 14th century, as well as a painting with a different subject from the same era. The metadata from these images was drawn from the Census, as well as from databases such as *prometheus* and Mnemic (see Figure 22).
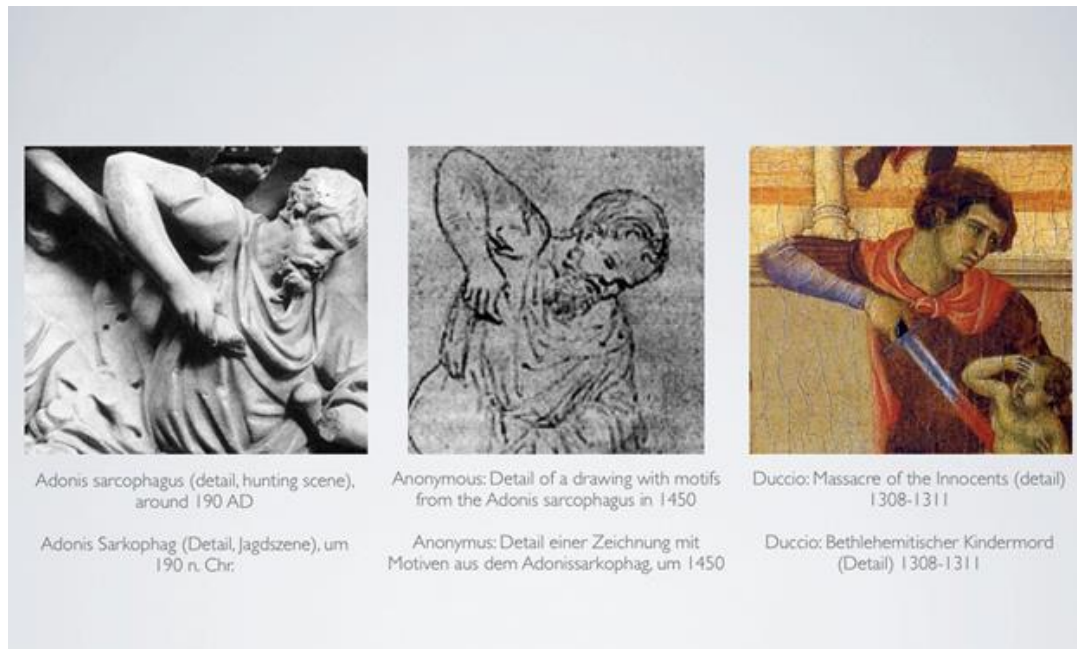
Figure 22. Images used in HyperGiotto showing similarities in limb alignment (picture: Jens-Martin Loebel)[62]

HyperMedia[63] was a project run by the University of Munich, which looked at cartoons and paintings created during and immediately after the French Revolution. The project managed to gather over a thousand items for annotation, and looked at how both commercial paintings and

---

[60] https://mneme.hu-berlin.de Mneme was the name that the Humboldt-Universität in Berlin used for a commercial asset management software developed by Zweitwerk. Humboldt-Universität is discontinuing this system as they now intend to use an open source alternative. (Accessed: 6 January, 2014).

[61] http://www.census.de/census/home/view?set_language=en (Accessed: 6 January, 2014).

[62] Screengrab taken from conference presentation by Jens-Martin Loebel http://hyperimage.ws/wp-content/uploads/imeji-Day-2013_Slides_Loebel_HyperImage.pdf (Accessed: 6 January, 2014).

[63] http://www.uni-lueneburg.de/hyperimage/hyperimage/hyperMedia_E.htm (Accessed: 7 January, 2014).

cartoon graphics were used in propaganda during this time, and explored the themes that were employed in those images to make their argument.[64] This included the parody of nations through maps, symbolism incorporated into more serious artworks, and motifs as a running theme throughout a series of images (see Figure 23).
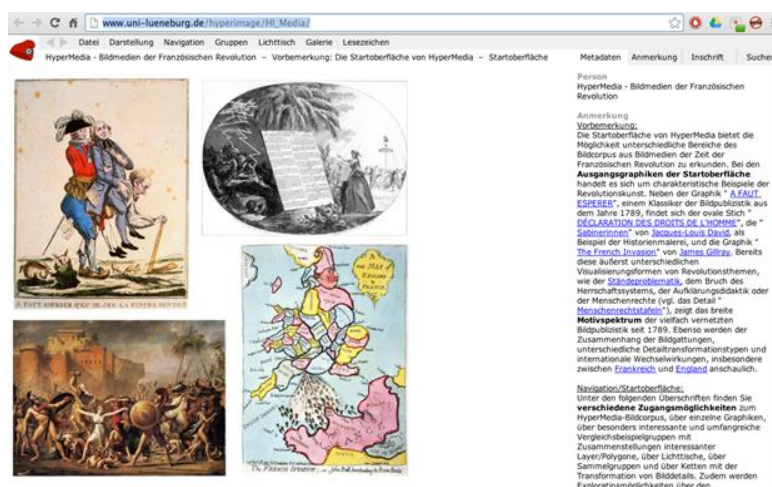
Figure 23. Screengrab of the HyperMedia project publication on HyperImage Reader

prometheus and HyperImage

In the second phase of the project, the HyperImage team worked alongside prometheus, Germany's largest art history database through the Meta-Image project[65]. HyperImage was employed as a means for users of prometheus to annotate images within the prometheus repository. It did not, however, allow users to upload their own images into the Meta-Image project. Instead, the project aimed to design a function whereby the user could upload images to their own prometheus account.

---

[64] http://www.uni-lueneburg.de/hyperimage/HI_Media/ (Accessed: 10 December, 2014).

[65] Dieckmann, Kliemann, and Warnke. "Meta-Image–a collaborative environment for image discourse." In *Proceedings of the 2010 International conference on Electronic Visualisation and the Arts*, British Computer Society, 2010, pp. 190-198.

Mirror of Kāśī

A team in Heidelberg University used HyperImage to reproduce and update a previous project on "A new Mirror of Kāśī"[66]. The project had originally been published in 2001, but they used HyperImage to update the outputs of the project to allow for newer features that HyperImage can provide. This included grouping elements and better searchability between the images (see Figure 24).
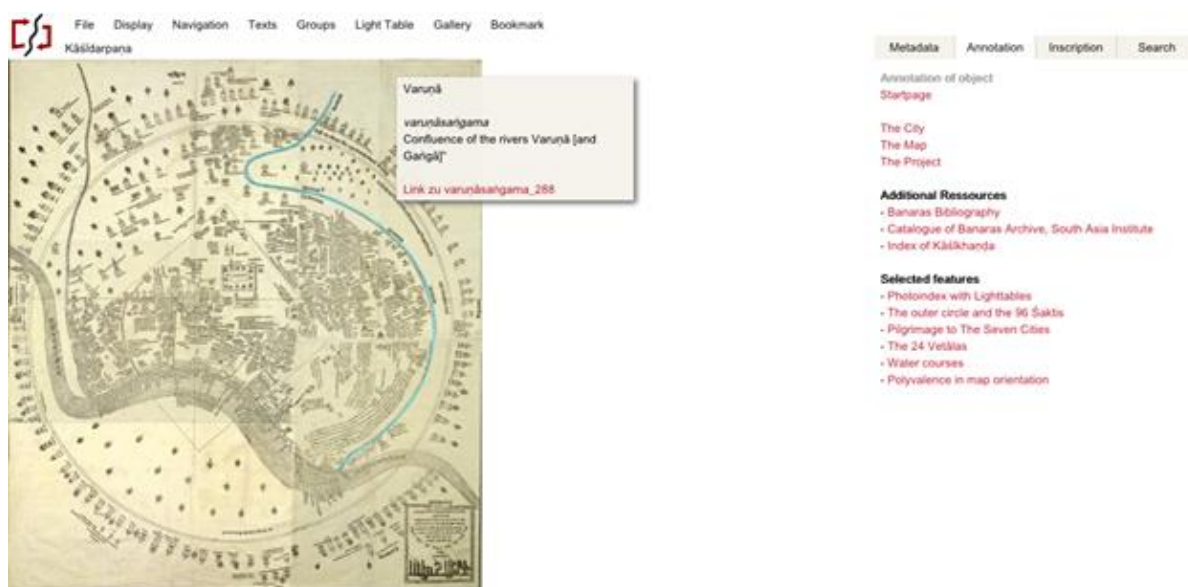
Figure 24. Screengrab showing annotation from the Mirror of Kāśī project

HyperImage's strengths and limitations

Strengths:

Compatibility

According to the HyperImage website, it is compatible with most operating systems – Mac, Windows and Linux.

Interface

The interface, according to Badzmierowska, is very easy to use – although the display is more functional than graphically pleasing. Indeed, once set up in the program, the interface is

---

[66] Gengnagel and HyperEvaluation Team (M Arnold, P Gietz, E Decker, G Niebler, W-T Wu) "Asia and Europe", Cluster of Excellence, Karl Jaspers Center, Heidelberg University 2011 http://kjc-fs2.kjc.uni-heidelberg.de/evaluation/Gengnagel04/PeTAL/index.html#T47593/ (Accessed: 7 January, 2014).

reasonably intuitive, although broken up into several windows, depending on whether you are working on a group of images, or an individual image. Images can be put into groups, either thematically, chronologically, by existing collection or by any other categorisation means, which makes it easy to find items for annotation. Labels for groups are free text, so this makes creating categories very straightforward.

HyperImage Reader

Once items have been annotated, they can then be previewed in the HyperImage Reader. This makes it very easy to see your project as the public would see it, thereby allowing the researcher to see any problems immediately and update or amend them accordingly.

Layers and Polygon Tool

Once you are familiar with the interface set-up, creating layers and polygons is fairly straightforward. The Polygon tool allows for quite a high level of accuracy in outlining the element you wish to annotate, and also allows the user to modify the colour and transparency of the annotation so that it is either invisible to the viewer until they hover over it, or is fully visible.

Documentation and support

The HyperImage website is very comprehensive. Currently, most of the information about the tool is available in both German and English, although there are still some documents that are only available in German. For background information on the tool, there is little printed information. Members of the HyperImage team have produced journal and conference papers on its development. These can be found in Appendix II – Documentation of HyperImage. Badzmierowska found the development team to be the best resource for support, however. The team will provide a training session via Skype if requested, which is very useful for giving basic information about how the tool works, creating links between objects, and links to external websites.

Limitations:

Compatibility

The reader and editor are not compatible with Google Chrome. Furthermore, HyperImage requires Java 7 to run. This is difficult for Mac users working on an OS older than 10.6 as Java 7 is not supported by previous operating system version.

Interface

The HyperImage team recognises that the current interface on HyperImage 3.0 is not an attractive one. They are currently working on improving this for HyperImage 4.0. The multiple windows within the tool make it difficult to keep track of work.

Pop-up windows that open when saving or creating a URL for linking purposes remain on screen after the link has been saved. The save function does not confirm that a change has been saved, making it difficult to know if it is safe to close the pop-up or not.

Future direction for HyperImage

HyperImage aims to expand its use across major EU projects. They have been in discussion with TextGrid with a view to establishing the tool for use on TextGrid Laboratory and TextGrid Repository. We have already seen how HyperImage teamed up with prometheus to enable annotation of repository items. It is the hope of HyperImage to build on this experience and make further reaches into working with trusted repositories.

HyperImage also aims to expand its scope to include annotation functions for text, video, audio files and 3D objects. They also plan to investigate how to integrate semantic web technologies and to increase the collaborative aspect to the tool to allow greater functionality within group projects, particularly in the classroom. One of the key goals for HyperImage is to increase its use in teaching environments.

In HyperImage 3.0 and 4.0, the developers hope to integrate the Reader into the Editor platform in order to reduce the number of programs needed to run the application fully. This should make HyperImage more streamlined and allow for instant publication of a project.

The current versions, HyperImage 2.0 and 3.0, do not allow for statistical data to be taken from the annotations. However, the team is working on HyperImage 4.0 in parallel, and hopes to introduce a means to draw statistical data from the annotations that can be used in further quantitative research.[67] It is envisaged that this statistical data will be in the form of .xls or .csv, although they are open to suggestions for what might be most useful to researchers. A spreadsheet-ready format might make most sense, though, as these are most easily fed into common statistical analysis tools such as SPSS[68].

---

[67] Personal Correspondence with Jens-Martin Loebel, Humboldt-Universität in Berlin, on 14th December 2013.

[68] Personal Correspondence with Jens-Martin Loebel, Humboldt-Universität in Berlin, on 27th December 2013.

How could image annotation software be used with Europeana content?

Europeana is currently rich with metadata for images, be they photographs, paintings, maps or diagrams. Images such as these could be used in any discipline: a linguist may wish to use an older map of a specific region they are studying; a sociologist may wish to look at photographs of groups of people, or street images that show everyday life. The same can be said for historians, art historians, geographers, and botanists to name but a few. In nearly all cases, the research requires cohesive links between themes within the images. Therefore the match between Europeana content and image annotation software such as HyperImage is an obvious and natural one.

Exporting images from Europeana

In the first instance, exporting images may be a simple case of adding an option to the Europeana interface that allows the user to export either individual or multiple images to an image-annotation tool. If multiple images are required for annotation, this export function could also be built into the My Europeana portal where links and Europeana items are stored. Multiple Europeana content items, along with their associated metadata, could then be directly exported from My Europeana into a user's image-annotation software (see Figure 25). To save the user from having to specify which items within their My Europeana file they want to export through clicking and opening each individual item, the option could also be available whereby they could right-click on an item and "add to export list". This export list could then be available from My Europeana, which they could then export to tools such as HyperImage, or indeed any of the other tools we have discussed within this report.
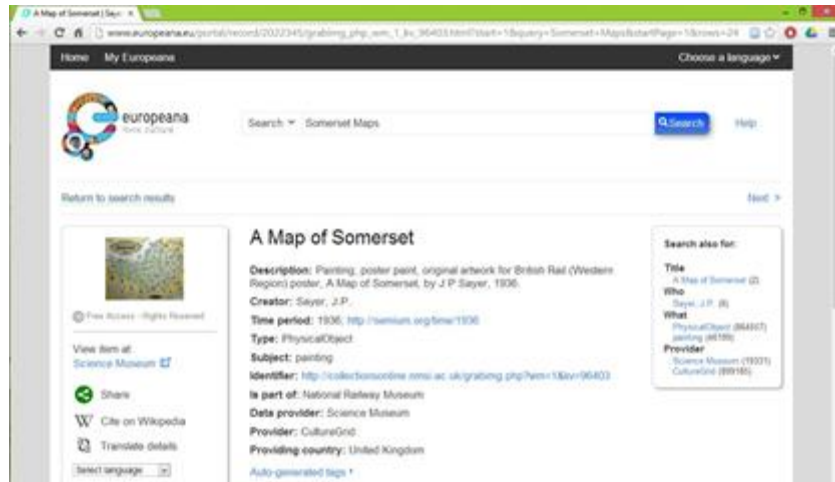
Figure25. Screengrab of Europeana item. Option to export to HyperImage / "Add to Export list" could appear under "Share" and 'Cite on Wikipedia' on the left hand side

As Europeana aims to deal with text, video, audio and 3D as well as images, these content types would match perfectly with HyperImage's planned expansion. In the case of HyperImage, it has shown that it is capable of working with repositories such as prometheus, although these may be on a smaller scale to Europeana. Caution might therefore be exercised in making any large-scale plans initially.

4.4 Case Study 3: Sociology and NodeXL

NodeXL (Network Overview, Discovery and Exploration add-in for Microsoft Excel) is a network analysis and visualisation tool that adds features and functions to the Excel spreadsheet. It is an open-source software package developed by the Social Media Research Foundation and can be downloaded directly from the homepage http://nodexl.codeplex.com/. Here one can also find the tool's source code and executables. As NodeXL relies on a spreadsheet, it works best when analysing small to medium size networks. While NodeXL has been most extensively used for the analysis of social media networks, the tool is relevant for all kinds of networks and can be broadly applied.[69]

This case deals with how NodeXL has been used by two sociologists: Dr Bernie Hogan and Marisa Ponti. Hogan is a research fellow at the Oxford Internet Institute. His research interests

---

[69] Smith et al., "Analysing (Social Media) Networks with NodeXL"*, Proceedings of the fourth international conference on Communities and technologies*, June 25-27, 2009.

include social networks, human-computer interaction, and social informatics. Hogan has used NodeXL for analysing social media networks with a special interest in analyzing relationships among Facebook users. Ponti holds a postdoc at the University of Oslo in which she is studying online crowdsource evaluation through social network analysis.

Functionality and features

Data handling

Before carrying out any analyses, it is necessary to gather and insert the data into the spreadsheet since this is what generates the network graph. There are essentially three ways to get data into the spreadsheet. Relationship data can be entered manually, or can be imported from an already existing source. NodeXL can import files from other major social network analysis (SNA) software, such as Pajek and UCINET, as well as from Excel. NodeXL can also import networks that are stored in several common formats: CSV, GEXF, GraphML, and Netdraw VNA among others. A third way is to import network data via plug-ins directly from social media sites, such as Facebook, Twitter, YouTube, and Flickr.[70] Likewise, NodeXL holds flexible export functionalities, being compatible with other SNA tools, for instance GraphML, Pajek and UCINET.

Functions of the program

NodeXL's building blocks are its worksheets that contain all relevant data for creating the network graphs. Regardless of how that data is inserted, the Edges worksheet constitutes the most essential cornerstone of the analysis, as this is where pairs of vertices, also referred to as nodes, are specified. The pairs of vertices define the relationships among the entities included in the network. These relationships can be specified by the researcher as undirected (default), bidirectional, or unidirectional. In the adjacent columns, visual properties and labels can be specified. (See Figure 26 for an overview of NodeXL's interface).

In the next step, NodeXL displays a network graph of the specified vertices. There are several graph layouts to choose from, which take into consideration different types of networks characteristics. Different layout types can be tried in order to reveal useful patterns and

---

[70] For a specification of available plug-ins, see: http://nodexl.codeplex.com/wikipage?title=Third-Party%20NodeXL%20Graph%20Data%20Importers. (Accessed: 17 Jan 2014).

relationships. The user can customise the graph manually through dragging, colouring, labelling, resizing, and filtering. Another functioning of NodeXL, especially useful when analysing more complex networks, is the possibility to break down the whole network into sub-networks, or "clusters" (groups of densely connected vertices that are not extensively connected to other groups or vertices). This can be done automatically through a built-in algorithm or manually.[71] If the researcher, prior to the analysis, knows which clusters to analyse and still select automatic clustering, there is a risk that NodeXL identifies more or fewer clusters than intended (which, on the other hand, can contribute to unexpected perspectives that otherwise would have remained hidden).
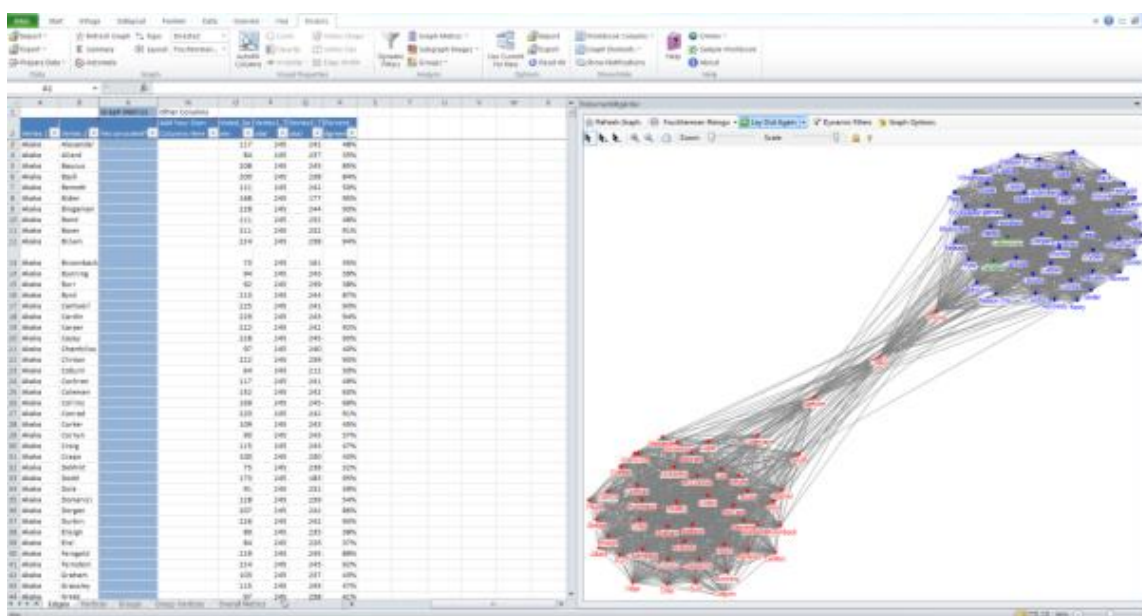
Figure 26. The main interface of NodeXL. At the top is the ribbon bar, situated under a specific tag for NodeXL's functions. The worksheet is to the left in which the pairs of vertices are specified. The network is shown to the right. The dataset comes from NodeXL's teaching resources.[72]

NodeXL also contains several network graph metrics for describing more accurately the network or its subgroups[73]. Some of these fall into the category named "degree" – the number of unique

---

[71] For the specific algorithm, see Wakita, and Tsurumi, "Finding Community Structure in Mega-scale Social Networks" in *Proceedings of the 16th International Conference on World Wide Web*, Banff, Alberta, Canada, May 08-12, 2007.

[72] http://nodexl.codeplex.com/wikipage?title=NodeXL%20Teaching%20Resources (Accessed: 20 February, 2014).

[73] An extensive review of Node XL's metrics can be found in Hogan, "Calculating and Visualizing Network Metrics" in: Hansen, Smith and Shneiderman (Eds.), *Analysing Social Media Networks with NodeXL*, Morgan Kaufman: New York, NY (2010).

edges that are connected to a particular vertex. If the directions of the pairs of vertices have been specified, two additional degree metrics can be calculated: "in-degree" and "out-degree". The former reveals how many edges point toward the vertex in question (for instance, the number of people that have sent an email to someone) while the latter shows the number of edges that the vertex points toward (for instance, the number of people that someone has sent an email to).

Other metrics have to do with centrality characteristics of the network. "Betweenness centrality" measures how central a node is for the network in the sense of providing bridges to different parts of the network. Second, "closeness centrality" is a measure of the average shortest distance from each vertex to each other vertex. Higher values thus indicate that a person is directly connected to more of the network's other vertices than are people with a lower value. A third measure of centrality is called "eigenvector centrality", which builds on the degree measure but takes into consideration not only a vertex's degree but also the degree of the vertices that the vertex in question is connected to. When dealing with an extensive set of vertices, the "clustering coefficient" can be useful for measuring the number of edges connecting a vertex's neighbours in relation to the total number of possible edges between those neighbours.

User experience of NodeXL

Bernie Hogan

Hogan has applied social network analysis in several research projects covering diverse themes such as socioeconomic status, local energy governance, and travel behavior. When it comes to NodeXL, however, his main interest has been the analysis of social media networks with a special focus on Facebook networks. He is the also author of chapter 5: "Visualizing and Interpreting Facebook Networks" in Hansen, Smith & Shneiderman, *Analysing Social Media Networks with Node XL* (2010).

Generally, Facebook network analyses are fairly large and complex with hundreds of nodes and thousands of edges. Manually specifying these relationships would be extremely time-consuming. Therefore, it is crucial to first download and install the Social Network Importer.[74] This graph data provider offers the possibility to automatically import various kinds of networks

---

[74] http://socialnetimporter.codeplex.com/ (Accessed: 17 Jan 2014).

from Facebook based on features such as co-likes and co-comments. Having imported the dataset, the graph is only one click away without any need for further data manipulation. Even though large networks easily can be imported, they are not necessarily easily interpreted. Hogan stresses that bigger data sets are not always better, especially when it comes to network analysis, as the graph has to be comprehensible. In a first step, it is useful to experiment with the different graph layouts offered by NodeXL. If the graph still remains difficult to comprehend, it would be appropriate to carry out clustering analyses within that network.

According to Hogan, the wide potential use of network analysis in combination with the rich and easy import possibilities of NodeXL may attract researchers to carry out network analysis on almost everything. This is somewhat problematic, Hogan explains, as you cannot necessarily do meaningful social network analysis on something even if relationships between entities can be specified. Working with NodeXL, it is therefore essential to reflect upon the reasons for using a particular dataset. On the one hand, NodeXL and similar tools certainly have influenced the research practices. Especially, researchers are now able to test their hypotheses much more quickly than before. On the other hand, it is important to bear in mind that the empirical researcher ought to work within a theoretical framework instead of testing infinite hypotheses at random. In this regard, digital tools offer no guidance.

Hogan has used NodeXL in several stages of the research process, such as data collection, data visualisation, and data analysis. He considers NodeXL to be most useful to get a quick first overview of what the data looks like by taking advantage of the tool's rich customisation possibilities. For more in-depth data analyses, other tools may be used to complement the investigation. One such tool is Gephi, which also has the advantage of not being restricted to Microsoft Windows.

Marisa Ponti

Ponti is interested in how people use and interact with technology. More specifically, she uses NodeXL for investigating crowdsource evaluation among a study group on the web site openstudy.com, a peer learning portal in which the users help one another by posting and answering questions in discussion threads. Essentially, Ponti analyses what kinds of networks exist among users in a particular study group. To do that, she uses both network visualisation and network metrics. The visualisation features allow her to customise and filter the sociographs in several different ways so that they show the relevant relationships in a simple and

comprehensible manner. She emphatically concurred with Hogan's point, that it is necessary to focus on clusters within the network or it will become too complex for a meaningful analysis.

NodeXL's strengths and weaknesses

A recurrent expression throughout the interview with Ponti was that NodeXL features an exceptionally user-friendly interface, especially for users already familiar with working with Excel-spreadsheets. No or little prior experience with SNA seems to be required in order to quickly learn NodeXL's main functions for analysing social networks. That circumstance, Ponti explained, was crucial in the process when she was about to choose among the many available SNA tools. This was of particular importance since she is working alone and have had to learn how to use the tool on her own. In a similar vein, Hogan described the user interface as "handy". However, he stressed that previous experience with Excel was a prerequisite to perceive NodeXL as a user-friendly tool. In fact, using NodeXL in an appropriate way may require the researcher to be highly skillful in Excel rather than just being an average user.

When starting to work with a new piece of software, problems will arise. That was true for Ponti as well as for us in our own evaluation of NodeXL. A couple of other advantages of NodeXL helped at this stage, though. As pointed out by Ponti, there are extensive documentation resources for NodeXL. This documentation takes the form of books[75], articles[76][77], an online forum[78], and videos. The discussion forum had been of great value for Ponti during her learning process, for instance.

As noted by Smith et al., network analysis has become increasingly utilised by researchers without a high-level experience from computational and algorithmical disciplines. Therefore it is all the more important to develop SNA tools that require no or little knowledge of programming language. NodeXL represents one such alternative together with other software like Gephi, Pajek, and UCINET.

---

[75] Hansen, Smith and Shneiderman, 2010.

[76] Smith et al., 2009.

[77] Aldhous. *NodeXL for Network analysis*. NICAR 2012, St Louis, February 24, 2012.

[78] http://nodexl.codeplex.com/discussions (Accessed: 17 Jan 2014).

NodeXL offers uncomplicated and clear visualisation features. The application affords a flexible layout of the network graphs in order to make them as useful as possible. When working with complex networks, there are a number of useful functions which enable the researcher to focus on smaller clusters within network as a whole.

One limitation of NodeXL, although not a problem for Ponti, is its limited capacity to analyse large networks. More precisely, NodeXL is not optimised for networks with more than a few thousand edges. Furthermore, NodeXL's graph metrics functions enable a number of simple yet essential analyses. It is, however, not possible to carry out more sophisticated analyses such as inferential statistics. As Hogan put it, NodeXL is most useful for "practitioners" of network analysis rather than experts as it lacks advanced data analysis functionality.

NodeXL comes with import functions which make it possible to automatically gather data from major social media sites such as Facebook, Twitter, YouTube, and Flickr. NodeXL is also compatible with other social network analysis programs (e.g. Pajek, UCINET), and several popular data formats (e.g. CSV, Netdraw VNA, GraphML).

A potential weak spot of NodeXL is its inherent link with a commercial product, namely Microsoft Excel. Thus, even if the software itself is free, usage implies a cost. This may also seem to implicate that it is necessary to work with the Windows operating system. Using another operating system, for example OS X as in the case of Ponti, it is possible to install and use NodeXL through the use of a virtual machine, even though this adds some complexity as it requires a higher degree of technical knowledge. Besides, one cannot get around the fact that a copy of Windows will be needed.

NodeXL is a free and open source software package developed in close cooperation with researchers, designed for non-programmers opening up network analysis to a potentially wide population of users.[79] The open source access to the NodeXL code ensures that anyone with the technical know-how can fix existing bugs or extend the code as they want to, without having to wait for new releases.

---

[79] This is not to say, however, that advanced social network analysts will not be interested in using NodeXL since it is perfectly possible to customize NodeXL through any .NET language. The statement rather refers to the main purpose behind NodeXL; "to create a tool that avoids the use of a programming language for the simplest forms of data manipulation and visualization" (Smith et al., 2009, p. 2).

NodeXL's further potential and future use with Europeana material

Analytic functionality

One point that Hogan made in the interview is that the program is at its best when it comes to data collection and data visualisation. Once the data has been collected, however, he thinks there are tools that are better suited for advanced analyses as NodeXL metric functions enable only elementary computational investigations. This seems to involve a two-step process where data are first gathered in NodeXL and then exported to another application where the serious analysis is being conducted. The analytic functions in NodeXL are not pointless, however, but are quite useful to get an overview of the data in the first stages of the analysis.

The Europeana network

As in the other cases, metadata enhancement seems to be an important part of improving Europeana usefulness in connection with network analysis. A possibility to export network-related metadata from NodeXL into Europeana would clearly add to the value of Europeana content as research data.

Visualisation of network connections

Once such relational data have been added to items in Europeana, it might prove useful to be able to visualise such connections within the Europeana platform. Connections between pieces of data that are based on previous research may very well be relevant criteria of selection for other researchers. They can also provide insight into the value of certain pieces of data and into previous research on those data from a variety of disciplines.

Visualising Europeana

Network visualisation functionality may also be useful within Europeana. For registered users of Europeana Research, a network of objects of interest could be created, which shows connections between researcher, research teams and the data they have taken an interest in. In this way, popular pieces of data can be identified and newcomers can get information on "what's hot" in their research field. At the other end of the spectrum, this would also show which pieces of data have not drawn any interest to them, a fact that may be equally interesting and relevant.

Linkurious

Linkurious is mentioned by Hogan as an interesting application and worth keeping an eye on. It is a web-based network analysis tool developed in France, and it allows its users to analyse

information networks directly in any web browser.[80] This functionality is likely to become more and more important in the future since it allows users to work on their data from anywhere, with fewer limitations connected to choice of and continuity in the use of operating system. Even though there are still some technical limitations (browser-based software usually demands a fairly modern browser), this increase in flexibility will probably become a central driving force in future software development. If NodeXL is to stay relevant to the users of Europeana, development toward a web-based service will likely be required.

Satellite cases of alternative tools

Education and alternative digital tools

The multidimensional character of the field of Education is interestingly imprinted on the various research paths and the consequent application of different digital tools, something that can be seen from the two interviews conducted for the satellite cases. These interviews demonstrate how the variety of digital tools employed in this field mirror the wide variations of research aims. For example, while Dr. Dimitris Koutsogiannis, Assistant Professor of Linguistics at the Aristotle University of Thessaloniki, focuses mostly on the collaboration among Education researchers by developing tools that support this function, Dr. Lampros Polkas, school teacher at the First Experimental Lyceum of Athens, explores the area of textual analysis and the use of visual representations of ancient Greek texts in the classroom.

Education and NVivo

Initiating the discussion on the alternative tools explored by Education researchers, NVivo is presented here and compared with Transana. NVivo, a digital tool used by Dr. Koutsogiannis, is software developed by the Qualitative Research Software Developer QSR International that supports qualitative and mixed methods research. Today, NVivo can be hardly considered an innovative tool. It was first launched in 1999 and was developed from earlier versions of NUD*IST (dating back to 1981), and several revised and updated NVivo versions have been launched (the most recent one in 2012). It allows the user to collect, organize and analyse content from interviews, focus group discussions, surveys, audio and recently social media data, YouTube videos and web pages.[81] Having numerous qualitative data, Koutsogiannis

---

[80] http://linkurio.us/ (Accessed: 17 January, 2014).

[81] http://www.qsrinternational.com/products_nvivo.aspx (Accessed: 10 January, 2014).

adopted this tool for his research project. His aim was to analyse and manage the data and to achieve quantitative results. More specifically, his project included the analysis of interviews and teaching scenarios for schoolteachers, collected in the context of a digital platform dealing with digital communities of Education researchers. Interested in building digital spaces of collaboration and communication among schoolteachers, on which they would be able to upload and/or retrieve content, as well as discuss and exchange ideas on learning scenarios, Koutsogiannis and his colleagues needed a digital tool to encode the resulting data. NVivo was chosen, as it was considered to provide numerous ways in which data could be analysed while it was also widely cited in research papers of the field. The characteristics attributed to this tool by Koutsogiannis and his team were: trustworthiness, easy data input to the program, and a wide range of methods for data analysis.

In order to evaluate Koutsogiannis's observations of NVivo, personal experience of the tool was necessary. Having no previous experience with this tool, the first contact with it was through tutorial videos provided on NVivo's web page. These tutorials offered an overview of the program's functionality. The NVivo interface could be described as intuitive for users experienced with digital tools, but fairly difficult to grasp for new users. As noted by the interviewee, uploading content is easy for the user; however, this data input focuses mainly on text data. Regarding audiovisual material, NVivo provides the possibility to edit and analyse audio and video, but this is rather time-consuming in comparison to other digital tools. In comparison to Transana, using NVivo for analysing audiovisual material can be quite time consuming. It requires the material to be transformed in order for it to be accepted by the tool. Transana, on the other hand, is particularly suited to the analysis of audio and video material and allows the material to be edited, analysed, and explored in a variety of manners[82]. As they currently stand, these two tools are thus complements in that they specialize on different types of material. Even without suggesting that these tools uniquely complement each other in their use by researchers (in fact, there is some overlap in functionality), it must be concluded that the coexistence of various tools, such as NVivo and Transana, suggests that a future Europeana will need to take into account the great variety of tools and research activities that researchers will want to use with the available Europeana content.

---

[82] http://transana.org/support/faq.htm#TransanaandNVivo (Accessed: 10 January, 2014).

Exploring NVivo more in depth, various strengths and weaknesses were revealed. As a new user of this tool, it was interesting to be able to easily uncover connections in the text, tag words, and get statistical results regarding word frequency in a text, for example. However, in order to be able to use this tool's feature, the user must have the necessary knowledge to create his/her own modelling. As Koutsogiannis noted, "the user must be careful when using [NVivo] in order to get true results, due to the relativity lying in qualitative data". Apart from that, the user does not have access to an overall view of his tags on the content as NVivo lacks visualisation features. On the other hand, it promotes collaboration, as it is possible for several users to analyse the same text and construct simple, straightforward summaries.

Education and Voyant

This visualisation shortcomings of NVivo is the key topic of discussion in Polkas' interview. Polkas is interested in the visual aspect of ancient Greek texts and he employs several digital tools in his research, within and outside the classroom, in order to allow him to create such visualisations. The tool he most commonly uses, and considers to be very satisfying for his research, is Voyant.[83] Voyant belongs to the family of Voyeur, a web-based text analysis environment. It is part of Hermeneuti.ca, a collaborative project aiming to develop and theorize text analysis tools and text analysis rhetoric. This tool reveals a different kind of connection between text and image, rather than within text (as in the case of NVivo). Polkas uses the tool for his own research in order to embrace digital tools in ancient Greek texts, as well as to find inspiration for teaching ancient literature. Simple, with helpful guidelines, powerful, colourful results and with a user-friendly interface, Voyant is easy to use even by an inexperienced user. Polkas pointed out two disadvantages, however: firstly, there is need for more specialized presentations of multitonic text; and secondly, and more significantly, there is a lack of multilingualism. Voyant, like several other digital tools that he employs in his research, accepts, analyses and presents English texts exclusively. This lack becomes critical in addressing younger users, for instance students, who will be hesitate to explore a digital visualisation tool for a text in a foreign language. Therefore, such useful tools inevitably constrain their audience to academics and other researchers, excluding the wider community of students, for example.

---

[83] http://voyant-tools.org/tool/Cirrus/; http://hermeneuti.ca/voyeur (Accessed: 10 January, 2014).

Matching Europeana and alternative digital tools

Examining these two tools with regards to Europeana, some observations can be made. Firstly, in terms of Voyant, a basic requirement of using this tool is having digital text available for creating its visualisation. Europeana, although having a wide and impressive collection of text, does not currently allow any analysis of texts apart from sharing or saving as an image. Therefore, researchers cannot search within a text; observe word frequency either visually, using a digital tool such as Voyant, or numerically, using NVivo; tag content or cross-search among texts. These possibilities would allow new results, actual analysis of the available content and, most possibly, wider engagement of users with Europeana's content. Developing the possibility to download machine-readable texts would be crucial for allowing the use with tools such as NVivo and Voyant.

Discussing the current state of Europeana, Polkas and Koutsogiannis seem to be skeptical to use it in their research. On the contrary, they both stress how they expect a useful database to offer content that can be easily exported for use in research rather than using a database which is perceived as a "virtual museum", as Polkas noted. Interpreting their comments in the context of scholarly primitives, they also identify a need for larger emphasis on cross-cutting primitives, such as monitoring, note-taking, translating, and on collaboration, suggesting thus that these activities are important parts of future research, regardless of which particular tools are employed to carry those activities out.

Discussion: How can the case studies inform future development of Europeana

Humanities and Social Sciences Case Studies concluded in a discussion on how the tools relate to Europeana and how the cases can inform Europeana's future development adopting a conceptual and structural framework. Such a framework is offered by Carole Palmer et al.'s division of primitives over broader "scholarly information activities" which comprise a number of "scholarly primitives", basic activities common to scholarly activity across disciplines as originally introduced by John Unsworth.

Palmer et al. present five scholarly activities, functions in the research process supposed to be common to scholarly activity across disciplines. The five scholarly activities they list are: searching, collecting, reading, writing and collaborating. To these, a sixth category of "cross-cutting primitives" is then added. Each category comprises a number of primitives. As Palmer et

al. focus on information activities, they have not included "analysing", a scholarly activity relevant to our discussion here.

Below, we look at those activities and primitives that are applicable to our case studies, drawing conclusions that are relevant to the future development of Europeana. The structure largely follows Palmer et al.'s 2009 presentation.

Collecting > Gathering

Within the perspective of Europeana Cloud and the case studies which form the basis of this discussion the primitive called "gathering" most accurately applies to the (repetitive) act of importing Europeana content into analytic tools such as those focused on in the case studies. All three tools support import of a range of different formats.

On the metadata side, Transana is fairly restrictive as its metadata are stored within an XML-based database structure. Importing ready-to-use metadata into Transana is only possible if they were in turn created in Transana and properly exported. NodeXL, on the other hand, stores its data and metadata in a spreadsheet, which makes them easily importable and exportable. HyperImage does not import metadata, as far as this case study can determine. It allows for the user to input their own metadata, however, using the annotation panel. HyperImage allows for metadata to be exported in XML format.

Import and export functionality and usefulness are closely connected to data formats and therefore format-related considerations certainly should be a key aspect of future Europeana Research development. It would seem reasonable to try to keep the selection of formats as wide as possible in order to be able to cater to the requirements of as many different tools as possible while also ensuring that long-term usability is maintained so unexpected future developments can be met.

Collecting > Organising

The way of organising a dataset is similar in the three tools: it basically means adding metadata to the data. This is central to all three tools examined in the cases and also in a way constitutes the first step of another important scholarly activity, that of analysing (see 7.6   Analysing) since an analysis must build on a structured, organised, set of data and the way the data are organised shapes the way the analysis proceeds.

The implication for Europeana Research is that metadata management is just as important as the question of formats. Keeping metadata together with the data is an essential part of making Europeana content interesting to researchers; since metadata is where previous researchers' (partial) analyses or ways of organising the data are stored (be it the compiler's or another secondary investigator's), and having access to previous analytic data is a necessary basis for getting synergy effects. Also, the more organisational metadata there are, the easier it gets for other researchers to analyse the data within their theoretic framework of choice. A dataset that is already organised leaves the secondary investigator more time to conduct the actual analysis.

Collecting > Archiving

Archiving is not one of Palmer et al.'s primitives but is considered here a potentially important activity. Both HyperImage and Transana enable the researcher to organise large collections of data by means of adding metadata. These metadata are essential for the analysis of the data but they can also form the basis for archiving as they provide a means of accessing the data in an organised fashion. Functionality that enables the researcher to organise Europeana content into personal, group, and public collections within the future Europeana portal is a sensible goal.

Writing > Assembling

The process of putting together one's thought on paper (or on a screen) is referred to as assembling by Palmer et al. This activity "is practiced in concert with searching, as well as reading, and extends well into the writing phases of a research project". Although assembling cannot be considered a core function of the pieces of software examined here, they comprise some activities that fall under this primitive. The tools under examination facilitate the assembling phase by giving the researcher the opportunity to put down analytic decisions, categorisations, and observations within the program and then have these systematised, aggregated, and visualised. A similar function in the Europeana portal that enables the researcher to select, organise and export analytic data to a text document or an in-portal text editing tool may prove useful.

Writing > Co-authoring

The process described in the next section (7.4.1) for Transana might be argued to constitute a first step in a co-authoring process. The remarks made there also apply to this primitive.

Collaborating > Co-analysing and consulting

The two primitives co-analysing and consulting are both addressed by Transana in its multi-user version, which allows several researchers to work on the same dataset at the same time. The same can be said for HyperImage and Nvivo. This function is important and appreciated since it spawns synergies. Most Europeana content lends itself to this kind of multi-faceted analysis since it is relevant to more than one scientific discipline. A potential feature of Europeana Research might be a set of collaborative functions aimed at groups of researchers, enabling them to send data between different analytic tools or at least share "play-lists" of Europeana records, thereby creating synergies.

Cross-cutting primitives > Note-taking

Transana offers the researcher an opportunity to document all research-related decisions within the program's notes function. Depending on how the future Europeana portal turns out, it may be relevant to provide a similar function for documenting decisions made on the Europeana side of the workflow.

Cross-cutting primitives > Translating

Transana demonstrates that it is possible for digital tools to facilitate translation of data and metadata. On the other hand, both tools examined in the satellite cases revealed a significant lack of multilingualism, which was strongly criticised by researchers in the interviews.

As Europeana is a pan-European project, multilingual cooperation ought to be a key point for the project. The future Europeana Research platform should provide tools that enable researchers to translate metadata, and, if possible, some forms of primary data, in order to make the material more broadly accessible to research communities across the European Union.

Cross-cutting primitives > Annotating

Annotating is the primary way of creating metadata in Transana, NVivo and HyperImage. One important point for Europeana to be aware of is the annotation functions of different tools, and especially the formats in which annotations are saved and exported in order for annotations to be migrated in a useful fashion. Moreover, it would be highly useful if it were possible to manage and process annotations within the Europeana Research platform.

Analysing > Visualising

Visualisation is central to the analytical process in all the tools examined in this report. Europeana Research ought to provide at least basic visualisation functions for different kinds of data. Such visualisations, e.g. displaying annotation layers (such as those in HyperImage) on top of Europeana original images, can provide an overview of a selection of data, something that will facilitate the process of assembling a dataset for a research project.

Analysing > Calculating statistics

NodeXL also offers the possibility to compute graph metrics in order to understand the data in more detail. Essentially, these operations quantify patterns in the data and work in tandem with the visualizing functions of social network analysis software. There is information about Europeana records missing about for instance the completeness of the material, which makes some statistical operations problematic, some basic descriptive statistics functions should be included in Europeana Research for the same reasons that motivate the visualisation functions.

Web-based research tools

This is neither a scholarly activity nor a research primitive but it is nevertheless an important point to be taken from the case studies.

As two of the interviewed researchers have demonstrated, web-based data analysis tools are increasingly interesting to serious researchers. Tools such as Dedoose, Linkurious, IFTTT and GoogleDocs enable researchers to both partly automatise data gathering and to conduct scientific analyses using cross-platform tools that are not bound to specific locations, laboratories etc.

Europeana is of course already a part of this trend, seeing as the repository is available online. Nevertheless we want to stress here the importance of keeping up-to-date with the latest developments in the area of web-based data analysis. It is likely that a lot is going to happen here in the next few years and Europeana needs to be able to interact neatly with the tools that exist, as well as with future tools.

Concluding remarks from Humanities and Social Sciences Case Studies

This case study report complements other tasks within Work Package 1, providing insights into how the Europeana collection could be utilised by researchers within the humanities and social

sciences. More precisely, we examined the functionality and application of some innovative digital tools for research in the fields of Sociology, Education Studies, and Art History. Tools similar to these could potentially be effectively applied to the content types of Europeana. For this to be achieved, however, several considerations need to be made with regard to the development of Europeana Research. Despite its different methodological approach, this task results in some conclusions that are similar to those of previous tasks within Work Package 1. However, this report also offers a few novel suggestions regarding the nature of a future Europeana.

In order to be a useful source of research material, Europeana needs to provide a broad selection of data and metadata formats to ensure applicability in a wide variety of analytic tools and long-term stability. Unlike previous reports (see for instance the Expert Forum reports of Deliverable 1.5), enriched metadata was not stressed as important for a future Europeana by the interviewees in this study. Developing Europeana for researchers was not among the subjects of the interviews, however, so the lack of discussions about metadata enrichment should not be taken to imply that such functionality would not, in fact, be appreciated by the research community. What metadata are present (whether added by users or the contributing institutions) should be easily accessible in terms of format as well as comprehensibility. A researcher about to write a publication based on analyses of Europeana content should be able to export analytic information, e.g. transcripts of clips created by another researcher or illustrations of the analysis of a certain network, into a text or graphics editing application, either third-party software or within My Europeana. Those pieces of analytic data could then be the points of departure for scholarly analysis.

Basic visualisation and statistical functions within Europeana Research would provide the researcher with a means of evaluating the relevance of a certain selection of data for a specific purpose. Thus, the researcher can avoid downloading and ingesting large data quantities only to discover that they hold no interest to the project at hand. Another useful functionality within Europeana Research would be a note-taking tool in which the researcher might document all steps of the research process. This feature would allow researchers to go back and trace analytical decisions made in their own previous research as well as reconstructing other researchers' methods and decisions, and would be invaluable in the documentation process of research data management (RDM).

My Europeana should feature functions that enable users to organise Europeana content into collections. Differentiation of collections should be possible, allowing personal collections to be kept separate from group collections (accessible to members of a user-defined group, e.g. a research team) and from publicly available collections. Europeana Research should consider incorporating a set of collaborative functions connected to the collections function. These might include the possibility of sharing analytic data and decisions within Europeana Research, or collaboratively annotate data and metadata.

Being a pan-European project, Europeana Research needs to provide multilingual access to the data. More specifically, a function for users to translate metadata and some types of research data into other languages would be highly useful to facilitate access to a wider array of researchers.

The case study provides the following key recommendations for Europeana Research:

- Provide import and export functionality for a wide variety of formats (awareness of different formats and the problems they bring with them is important).
- Functionality that enables addition, import, export, and collaboration metadata for Europeana records should be flexible and easy to use.
- Functions for mark-up, organization, and collaboration in My Europeana would be highly useful.
- Provide basic analytic functionality and the possibility of documenting research decisions.
- Support for (the creation of) multilingual data and metadata would greatly increase the material's usefulness.

5. APIs in Humanities and Social Sciences Research

5.1. Introduction

In March 2014, the authors of this report were charged with the task of investigating best practice in the use of data of Application Programming Interfaces (APIs) by researchers and for researching data from the Cultural Heritage sector. This research would inform and underpin the significant developments already underway at Europeana in the development of a specialised access point to the Europeana API for researchers, to be known as Europeana Research.

This clear research goal required nuancing in the delivery, however, because "researchers" don't often use "APIs." There are plenty of researchers using cultural data, and it may well be that in some cases their systems of data capture do make use, wholly or in part, of an API service. But most of the researchers we were able to identify really only cared about the data, and had no specific opinions about how that data was accessed.

The term API seems therefore to be *a priori* restricted to the use of and by developers. Our research therefore required a slightly different approach of looking not only at the creators and developer/users of APIs, but also at support services within the data-intensive humanities and social sciences research lifecycle as well as those humanists or social scientists reusing data themselves, with a specific focus on how they would acquire data and what they would want to do with it (that is, in most cases, what structure they would apply after download).

The report that follows is comprised of three main sections: first, an overview of API provision by cultural heritage institutions, focusing on a number of high-profile best practice exemplars. The second section looks at researcher behaviours, including those related directly to API use as well as those that might act as predictors of possible API use, such and programming skills or collaborative tendencies. The final section will provide a synthesis of these two approaches, extrapolating from the trajectories of API development and research practices a series of recommendations for increasing the attractiveness of a API for research purposes.

5.2 Section 1: Current trends and best practice examples in API Creation and Usage

It seems that developing an API has become a new 'must have accessory' for digital cultural heritage collections. Although not all DH projects follow this policy, the clear distinction between an API and User-Interface is considered significant here. The following examples illustrate how varied the practice of developing these tools can be, and point toward possible definitions of a 'successful' API implementation.

Who is developing APIs? What do they offer and how are they used?

Trove

Trove is a free search service provided by the National Library of Australia[84]. The service allows the user to search through metadata of digitised content from newspapers and magazines as well as academic articles, journals and theses. Similar to Europeana, this is aggregated from external resources both within Australia (such as the Australian National Bibliographic database, and the Australian Parliamentary Library) and overseas resources (such as the HathiTrust - discussed below, PubMed Central and ProjectMUSE).

Trove provides an API for users to enable searches within their metadata. The documentation for the API is available on the Trove website[85]. They also showcase an example of how an API has been used with trove data - with the 'Beverage Wars' infograph[86] from Wragge Labs[87], a company set up by Trove's Tim Sherratt. The graph in the figure below shows the different mentions of 'tea' versus 'coffee' in the articles available via Trove, which he was able to deduce using an 'unofficial' API, that Sherratt built himself, in conjunction with the TroveTools Scraper. The graph is 'clickable' and allows the user to click on any point on the graph to reveal a list of articles on the right-hand side of the interface which mention either 'tea' or 'coffee', depending on which set of data you have clicked on (see Figure 27).

---

[84] http://trove.nla.gov.au/?q= (Accessed: 14th April 2014).

[85] http://trove.nla.gov.au/general/api (Accessed: 14th April 2014).

[86] http://wraggelabs.com/shed/trove/graphs/coffee_tea.html (Accessed: 14th April 2014).

[87] http://wraggelabs.com/ (Accessed: 14th April 2014).
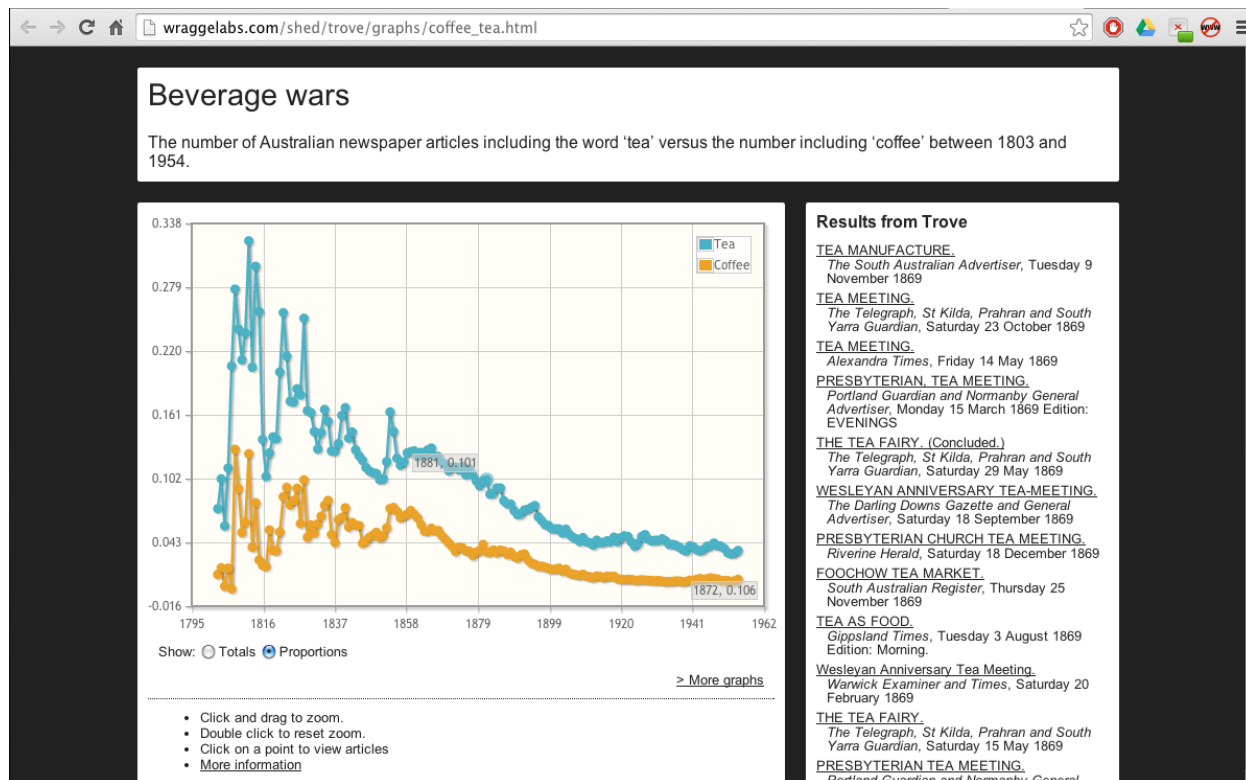
Figure 27: "Beverage Wars" from WraggeLabs, using data from Trove

Since Tim Sherratt's "WraggeLabs" exercise, Trove has built an accessible API, which is supported by a dedicated website with detailed information about the API, including its purpose, and examples of blogs with information on building on the Trove API[88]. It also provides examples and experiments that have been carried out by the Trove staff in order to showcase what can be done[89]. This will be discussed further on this document.

HathiTrust

The HathiTrust[90] is a collaborative research centre run between The University of Illinois, and Indiana University, and is the joint effort of "academic & research institutions, offering a

---

[88] http://help.nla.gov.au/trove/building-with-trove/examples (Accessed: 4th June 2014).

[89] http://help.nla.gov.au/trove/building-with-trove/experiments (Accessed: 4th June 2014).

[90] http://www.hathitrust.org/ (Accessed: 14th April 2014).

collection of millions of titles digitized from libraries around the world."[91] The HathiTrust Digital Library allows users to browse collections online, read books from collection online, and (if they have the required access) download certain texts and create collections.

Information about their API[92] use is less easy to find than on the Trove website. There are two APIs made available to potential developers, though. The 'Bibliography API'[93] offers metadata about bibliographic, copyright and volume information.   The 'Data API'[94] offers images of webpages, OCR text and associated metadata. The content that can be provided via the Data API is determined by which 'dataset' it falls into. The HathiTrust allows researchers access to all public domain works within its corpus. However, it divides this public domain work into two main categories - those that have been digitised by Google, and those that have not. The non-Google-digitised works are freely available, but the Google digitised works are only available through an agreement with Google[95]. The Non-Google Digitised corpus comprises approximately 350,000 volumes, and represents documents in the English Language published prior to 1923. The Google-Digitised sets contain around 2.8million public domain volumes in multiple languages. The restrictions on this latter dataset are such that they can only be used for scholarly activity, and not commercially. Customised datasets can be created in agreement with the HathiTrust.

At the JADH 2013 conference in Kyoto, Prof. J. Stephen Downie and Prof. David Bainbridge outlined their development of an affiliated Hathi Trust Research Centre (HTRC) which combines their API with other tools to allow enhanced access to content for scholarly purposes (Downie and Bainbridge, 2013). The tools in the example shown were the Greenstone[96] digital library

---

[91] From the same website as above.

[92] HathiTrust Digital Library "Data Availability and APIs.". http://www.hathitrust.org/data .

(Accessed: 14th April 2014)

[93] http://www.hathitrust.org/bib_api (Accessed: 14th April 2014)

[94] http://www.hathitrust.org/data_api (Accessed: 14th April 2014).

[95] http://www.hathitrust.org/datasets (Accessed: 15th April 2014).

[96] http://www.greenstone.org/ (Accessed: 10th June 2014).

system and the Meandre[97] workflow analytic environment. Greenstone empowers individuals and smaller groups to build their own digital libraries, which can then be shared online. Meandre "provides the machinery for assembling and executing data flows". Combined and working alongside the HT's bibliographic and data APIs, they created a tool that enables scholars to create their own subsets of data from the HT corpus which they can then download. In some cases, the associated pages and images within a specific work can also be downloaded automatically when that item is in the public domain.

Digital Public Library of America

The DPLA (Digital Public Library of America) works on a model very similar to that of Europeana. It draws in digital data and content from libraries, museums and archives from across the United States of America, and makes them available for public use via their portal and platform infrastructure. The portal is available for searches, whereas the platform allows users to make calls to the API.

In developing their current set up, they devised personas of the people they envisaged using the DPLA site, and short stories around each one that show how they would use it. They developed 12 in all, ranging from a teenager trying to write a paper in a hurry for school homework to a lady in her 70s who wants to share her pictures of her home-town with the world. None of the profiles quite matches the technology-loving humanist who wants to produce a ground-breaking piece of research using data extracted via the API, but there are examples of academics and members of the public who want to upload items to the DPLA site, which would require an API, but this is not explicitly mentioned.

---

[97] http://www.seasr.org/meandre/ (Accessed: 10th June 2014).
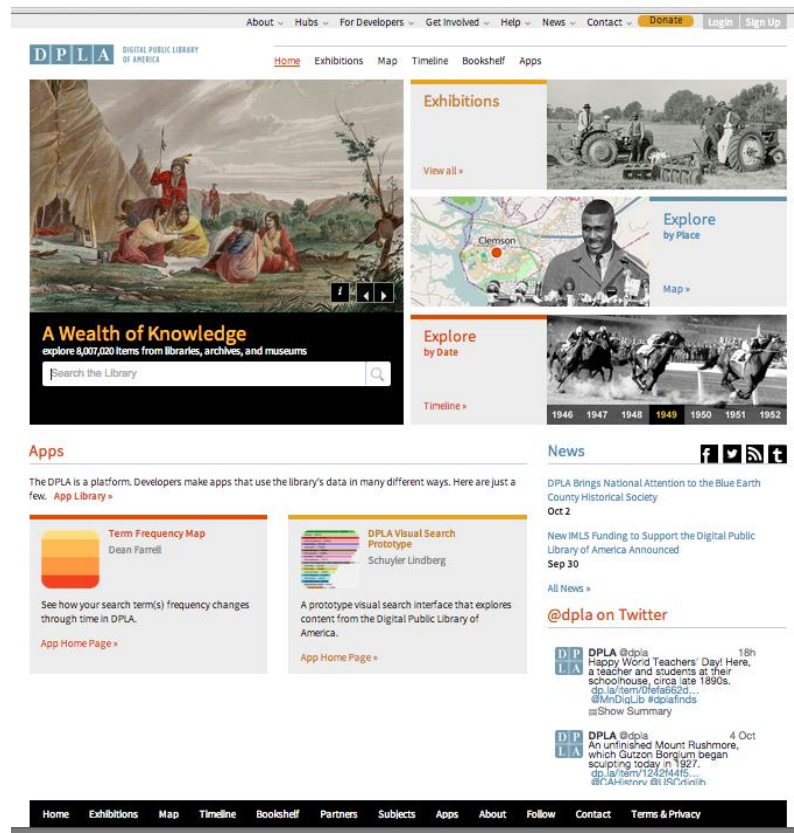
Figure 28: Homepage of the DPLA website, showing potential uses and apps clearly

The DPLA's main draw, however, to many researchers and curious members of the public is the vast array of apps associated with the resource. The home-page[98] alone allows several different ways in which the content can be searched (see Figure 28). On the right-hand side, visitors can search by place, or by date, while also being able to make use of the apps lower down on the page for frequently used search terms, or a visual search engine. The larger Apps Library[99] on the website showcases the various apps and widgets that have been built using the DPLA API, and that can be used to access content within the portal. These apps can be used for academic purposes (for example the Library Observatory app from Harvard[100] which allows users to search through many of the collections by institution across the USA, and then drill down into the content within certain categories) or simply to amuse (such as the 'Historical Cats'

---

[98] http://dp.la/ (Accessed: 6th October 2014).

[99] http://dp.la/apps (Accessed: 6th October 2014).

[100] DPLA App - Library Observatory http://www.libraryobservatory.org/ (Accessed: 6th October 2014).

Twitterbot [101] which selects random items from the DPLA resource and tweets them out via the 'Historical Cats' twitter account, usually accompanied by some feline related exclamation).

Europeana

Europeana is a European-wide resource, which provides open access to cultural heritage data from museums, archives and cultural organizations across the continent. Data is 'aggregated' by various partners. Main public access is via the Europeana portal[102], which allows users to search for items using keywords. Thematic collections have also been curated that respond to a certain topic, for example the Europeana 1914-1918 project[103] draws on items from its collections that relate to the First World War. These thematic projects are also a way for Europeana to ingest data and content, as is currently underway with the Europeana 1989[104] project, which is calling for contributions of images and memorabilia that relate to the fall of the Iron Curtain, not only from Europeana's aggregators and affiliate network members, but also from the general public.

Europeana currently offers two APIs for use by researchers. The first is a REST-API, which can be used to retrieve items from Europeana, much like a search engine, or indeed the Europeana Portal in its present form. The second is, as the Europeana Labs website puts it, 'more experimental' allowing for entire metadata sets to be retrieved in SPARQL-query language.[105] So far, around 1863 API-keys have been issued for access to the Europeana API, 850 of which were issued in 2014. Nearly 5.2 million requests were made to the API in August 2014, with peak days reaching 200,000 requests. Of the API keys issued, around 100 are used regularly on a monthly basis[106].

---

[101] Twitter Profile - Historical Cats https://twitter.com/HistoricalCats (Accessed: 6th October 2014).

[102] Europeana Portal - europeana.eu

[103] Europeana 1914-1918 project website - http://europeana1914-1918.eu/en (Accessed: 30th October 2014).

[104] Europeana1989 project website - http://www.europeana1989.eu/en/ (Accessed: 30th October 2014).

[105] http://labs.europeana.eu/api/ (Accessed: 9th June 2014).

[106] Data from James Morley - Europeana Labs - personal correspondence 30th Oct 2014

In order to promote re-use of the data and content available via Europeana, several 'hackathons' have been hosted over the past couple of years, allowing researchers across Europe the opportunity to make use of the data and possibly create something marketable using the APIs. This is discussed in more detail below.

Europeana Labs is Europeana's space for developers (and keen digital humanists/social scientists) to create interesting and innovative tools or apps using the Europeana API. Europeana Labs is currently in the process of evaluating the Europeana APIs, and in particular the kinds of people that are using it. In doing so, they have devised 'personas' for users, however they seem to be mainly of individuals who already have a good deal of experience. In order to devise these personas, they approached media artists and people working in creative industries, and asked them a series of questions to establish what tools and websites they already use in their work, and what is good / bad about them.

Additional Exemplars

In addition to these four large projects, we also looked at several smaller, project or institutionally-based API implementations, each of which had certain interesting characteristics.

Victoria and Albert Museum

The Victoria and Albert Museum (V&A) is the world's largest museum of decorative art and design. Founded in 1852, it has long housed collections of cultural significance, and has recently undergone large-scale digitisation of much of its content through the 'Ground Floor' project[107]. In 2009, this digitised collection contained images and descriptions for approximately 50,000 objects, and a further 70,000 images[108]. The V&A has 5 main curatorial departments: Asia; Furniture, Textiles and Fashion; Sculpture, Metalwork, Ceramics and Glass; Theatre, and; Word and Image. Access to the Word and Image Department is available both onsite at one of

---

[107] http://oclc.org/content/dam/research/activities/lamsurvey/2009-05-02e.pdf - slide 24 - presentation by Douglas Dodds, May 2009 (Accessed: 15th April 20140.

[108] Douglas Dodds - same presentation as above.

the Museum campuses, or online via the search tool. The V&A API is, according to their website, "designed as a RESTful interface" to their collections[109].

The V&A API also provides supporting documentation, which is intended for the lay-person. This is discussed in more detail in the next section.

The British Library - British National Bibliography

The British Library currently hosts access to two different types of content. Access is given to metadata contained within the British National Bibliography about published books and serial publications[110]. Rather than hosting the API on the British Library website much of the development of tools and apps is done via the British Library Labs project, which encourages researchers and developers to use and develop tools with the British Library Labs team to answer interesting research questions[111]. The British Library is instead 'leveraging' the Flickr API[112] to allow users access to the available digitised content[113].

The British Library has now ceased using the term 'API' as a means to access their metadata, instead providing access to such metadata on the BNB via their Flint SPARQL-Editor[114].

---

[109] Victoria and Albert Museum "Victoria and Albert Museum API Documentation" http://www.vam.ac.uk/api/ - (Accessed: 15th April 2014).

[110] http://bnb.data.bl.uk/faqs - British National Bibliography FAQs (Accessed: 30th October 2014).

[111] http://labs.bl.uk/ - British Library Labs wiki (Accessed: 30th October 2014).

[112] https://www.flickr.com/services/api/ - The App Garden, Flickr (Accessed: 30th October 2014).

[113] http://blog.flickr.net/en/2013/12/16/welcome-the-british-library-to-the-commons/ (Accessed: 30th October 2014).

[114] http://bnb.data.bl.uk/flint-sparql - British National Bibliography's FLINT SPARQL Editor (Accessed: 30th October 2014).
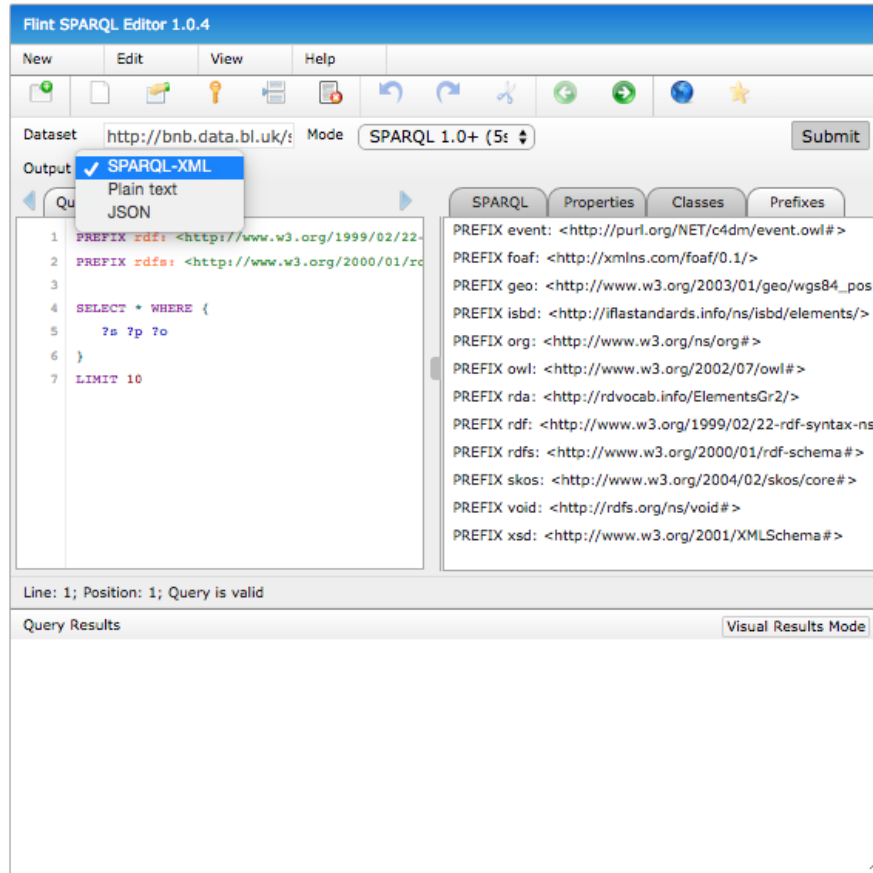
Figure 29. Screengrab of the British National Bibliography Flint SPARQL Editor 1.0.4

Users of the BnB Flint SPARQ Editor can also select the output they require, be it in SPARQL-XML, Plain Text or JSON. This comes with ample documentation and examples of queries that can be made using this editor, in much the same way that others such as the V&A provide documentation to their API. In addition to this editor, users can also access metadata in the BNB using a '/sparql' endpoint remotely[115].

---

[115] http://bnb.data.bl.uk/ - British National Bibliography home page (Accessed: 30th October 2014).

Finnish National Gallery[116]

The Finnish National Gallery comprises three art galleries, the Ateneum Art Museum, the Museum of Contemporary Art Kiasma, and Sinebrychoff Art Museum. The documentation to support users of the API is quite extensive, similar to the V&A museum.

EDINA - AddressingHistory project[117]

The EDINA 'AddressingHistory' project is a website and API that takes information about addresses from the Scottish Post Office Directories and combines it with modern maps of Scottish cities, such as Edinburgh, Glasgow and Aberdeen. It is aimed at researchers, local historians and genealogists, and the API allows users access to the raw data via multiple output formats, e.g. geo-tagged maps (MacDonald, 2011). The project followed JISC's 'Good Practice Guidelines'[118] for those providing APIs, as well as those using APIs, which were put together by UKOLN in 2009.

The AddressingHistory API also allows for 'write' as well as 'read' access, but hosts its user-generated content ('UGC') separately from its existing content in order to conduct cross-checks on the accuracy and relevance to the AddressingHistory project, and to avoid 'contaminating' the data they already have.

How are these data APIs being promoted and made accessible to non-technical users?

Providing thorough and simple documentation

In the best cases, the API providers are giving clear instructions and support to developers on their websites. Trove has a group of pages outlining how best to use their API[119].

The Victoria and Albert Museum has a one-page layperson's guide to their API for researchers and developers who wish to access the collections via more customised means. The

---

[116] http://kokoelmat.fng.fi/api/v2support/docs/#/overview

[117] http://addressinghistory.edina.ac.uk/

[118] JISC Good Practice Guidelines for API Providers http://ie-repository.jisc.ac.uk/344/

[119] Trove Help Centre - http://trove.nla.gov.au/general/api-technical#intro (Accessed: 2nd April 2014).

documentation talks the potential developers through the API, beginning with the simple API Query Builder, explaining how each part of the resulting url will produce a different result, for example in which format you want the results to appear (e.g XML, JSON, etc.). This then moves on to more information about keyword searching, geospatial searches and searching by date.

Using Social Media to build awareness

The EDINA *AddressingHistory* project made use of social media such as Facebook and Twitter, as well as running a blog[120] and creating videos, which were uploaded to YouTube and Vimeo.

We have also seen how the DPLA have made use of the Historical Cats Twitter profile.

Hosting Hackathons to directly encourage development that can support non-technical users

The British Library regularly hosts competitions, similar in a way to how Europeana ran their Hack4Europe competition to encourage creative ways to use the API and datasets held within the British Library's collection.[121] The results of these competitions are showcased at half-yearly British Library Labs Symposiums.

Europeana also hosts Hackathons as a means of encouraging use of its collections. Starting in 2011, Europeana began running a series of one - two day workshops under the label of 'Hack4Europe', many running as a kind of roadshow throughout 9 countries in Europe[122]. The aim of these hackathons was similar to the British Library, in that they were intended to encourage use of the Europeana API, and the data available within the collections. Run as competitions in 2011 and 2012, the finalists for the Hack4Europe events were recognised at the Digital Agenda Assembly in Brussels[123]. The 2012 Hack4Europe launched Europeana's

---

[120] EDINA - AddressingHistory website http://addressinghistory.blogs.edina.ac.uk/ - (Accessed: 2nd April 2014).

[121] http://labs.bl.uk/BL+Digital+Collections+Interlinking+to+Europeana

[122] http://pro.europeana.eu/hackathons (Accessed: 29th October 2014).

[123] http://pro.europeana.eu/web/guest/hackathon-prototypes (Accessed: 29th October 2014).

initiate of 'Connecting Society to Culture'[124] at the Europeana Awareness event "Culture for digital innovation" on 9 May 2012 in Brussels.

Providing Training

The British National Bibliography (BNB), part of the British Library, ran a series of workshops, 'Introduction to APIs' for curators and librarians within the British Library who have little to no experience of APIs in 2013 (McGregor, 2013). The workshop and tasks assigned in it were made available online[125]. However, the British Library have since changed their platform, and the tasks assigned online are no longer valid (as is evident from the comments section). One of the organisers of this workshop, who also maintained the online version on his blog, believed the workshop was a success - although it seems that they did not have sufficient time in the one day to complete both tasks assigned. The pace of each workshop was determined by the level of expertise of the participants. However, as the end goal wasn't to have major projects coming out of the data held within the British Library, it was merely to educate the Library staff as to the possibilities. The aim, he told us, was to give curators of collections and librarians the knowledge to be able to point library users to new ways of re-using the data. Educating the library staff, then, seems to have been one of the British Library's approaches to promoting the API[126].

The Maryland Institute of Technology in the Humanities (MITH) also delivered a workshop for digital humanists in February 2011[127]. Its aim was to discuss "ways that existing and future APIs could be leveraged for digital humanities projects". The backgrounds and needs of the 50 participants, a mix of digital humanists and developers, were varied. Some were looking for a more hands-on approach to using APIs, whereas others were just interested in what could be done with APIs. The schedule[128] for the day included presentations on basics ('APIs: What are

---

[124] http://pro.europeana.eu/hack4europe-2012 (Accessed: 9th June 2014).

[125] http://www.meanboyfriend.com/overdue_ideas/2013/02/introduction-to-apis/ (Accessed: 31st March 2014).

[126] Personal correspondence - 7th July 2014

[127] API Workshop - http://mith.umd.edu/apiworkshop/ (Accessed: 7th April 2014).

[128] http://mith.umd.edu/apiworkshop/schedule/ (Accessed: 7th April 2014).

they and why do they matter to digital humanists?'[129]) as well as more practical workshops and 'Mashup' sessions. Participant in the workshop, Fred Gibbs, Assistant Professor of History at the University of New Mexico, comments "In terms of bridging the humanist/technology gap, isn't it easier to slide the APIs a bit closer to the humanists than the other way around?"[130] on his blog. His over-riding question coming out of the workshop was whether or not there should be a separation of APIs and User-Interface.

Creating and showcasing inspirational examples of usage

In some of the cases we have seen from the API providers, there is a data evangelist associated with the Museum or Library who regularly produces work based on data from the collections. This work not only highlights the use of the API, but also shows what can be done with the data.

WraggeLabs is the work of Tim Sherratt[131], an Historian and Archivist based at the National Library of Australia, (Trove). One could reasonably call him someone who is already invested in the API provision of Trove, as he works for the API provider, however the fact that he is also an Historian and Archivist makes him interesting within the context of this research. An exemplar of his work with the Trove API can be seen above with the 'Beverage Wars' site. Tim has also conducted research using an API he built himself for use with Trove, to look at Australian newspapers[132]

In some cases, the digital project or CHI works closely with an academic, such as the HathiTrust working with Ted Underwood[133]. This is often mutually beneficial as the humanist has supported access to the data, and the CHI gets to use the outputs as promotional materials. In many cases, however, these relationships have a particular status, and often involve working together

---

[129] http://labs.dataunbound.com/doc/2011/02/mith_intro_talk.html (Accessed: 7th April 2014).

[130] Gibbs, F. "The UI / API divide", *Fredgibbs.net.* February 28th 2011. http://fredgibbs.net/posts/post/data/ui-api-divide/ (Accessed: 7th April 2014).

[131] http://trove.nla.gov.au/people/479364?c=people (Accessed: 14th April 2014).

[132] http://wraggelabs.appspot.com/api/newspapers/ (Accessed: 29th May 2014).

[133] Ted Underwood, "Getting Everything You Want From HathiTrust". *The Stone and the Shell* Blogsite http://tedunderwood.com/2012/07/27/getting-everything-you-want-from-hathitrust/

on data sets in a way that could not be facilitated only through an API. John Bradley, at King's College London, used the V&A API to obtain images for annotation as part of a wider project investigating digital annotation of different media, and how annotation might be further developed. He does this within the context of the Pliny project[134], a digital annotation tool that Bradley and his team built.

Jake MacMullin from StripySock[135] has worked with the National Library of Australia (NLA) to produce the Forte app[136] for the iPad. This app makes use of the NLA's API, and provides access to up to 13,000 files of digitised sheet music (as well as the cover artwork) held within the NLA's archives. The app had come about following the NLA's 'LibraryHack' competition. The competition itself was not a success for the NLA, as the dataset they had provided was not used significantly in any of the entries. However 12 months later, Jake approached the library with his idea for the app. His intention had been to produce an app that would give the musician the experience of 'flipping through a physical score" which had been lacking in previous digitised sheet-music renditions.

According to Paul Hagon of the NLA:

"To me there were three really important lessons to come out of this:

1. Make your data available.
2. Don't expect things to happen immediately
3. If you can, guide the developer to help them fine tune their product & give them insight into the intricacies of the data."

At a smaller scale and within Europe, Janne Heinonen worked with the Finnish National Gallery's API, to find and visualise all the items in the gallery that are 'untitled'[137]. The graph

---

[134] http://pliny.cch.kcl.ac.uk/ (Accessed: 8th October 2014).

[135] http://stripysock.com.au/ Jake MacMullin on StripySock

[136] http://www.paulhagon.com/2013/03/26/forte-the-national-library-of-australias-sheet-music-collection-ipad-app/ (Accessed: 8th April 2014).

[137] 'Untitled / Unititled" http://www.lily.fi/blogit/saattaa-sisaltaa-nykytaidetta/nimetonuntitled 'Lily' blog, written on 19th January 2014

below shows the number of 'untitled' pieces and the years in which they were created by decade.
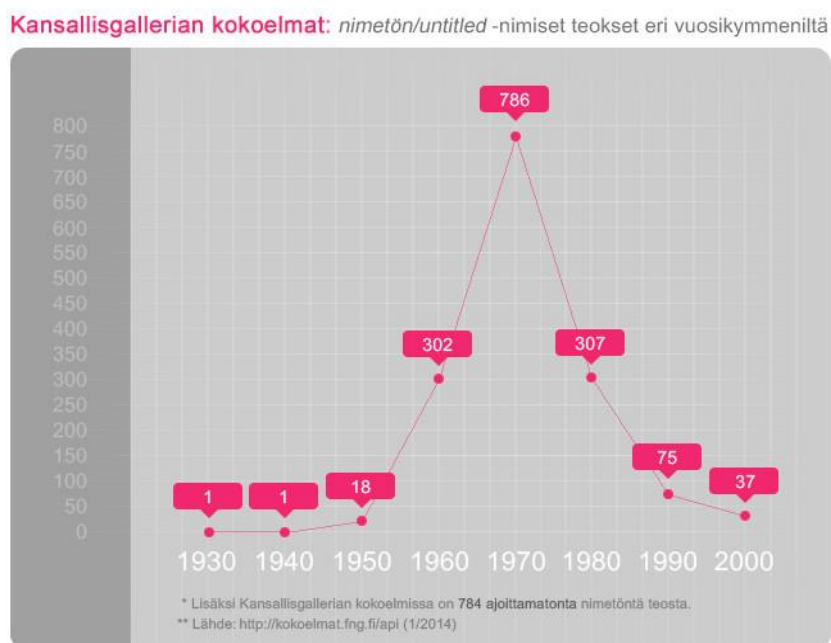
Figure 30. Untitled artworks in the Finnish National Gallery by Decade

In another public project from the Finnish National Gallery, a Flickr user name 'Hugovk' took the metadata regarding the dimensions of its artworks to create multi-layered images dimensional outlines[138]:

---

[138] User - Hugovk Flickr page - https://www.flickr.com/photos/hugovk/sets/72157644045546102/
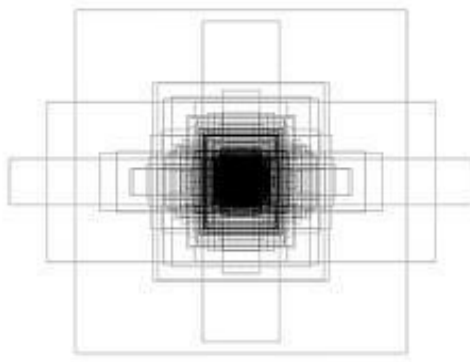
Figure 31. The size of 23,557 artworks in the Finnish National Gallery

139

Collaborations between Developers and Cultural Researchers

Although it is not always directly instigated by the digital projects providing the API, many inspirational examples of API usage result from active collaborations between researchers and technical developers. Sarah Kenderdine, based at the National Institute for Experimental Arts at the University of New South Wales[140], used the Europeana API to develop ECLOUD - a visual representation of artifacts held within Europeana that surround the First World War.[141]

---

[139] "The size of 23,557 artworks in the Finnish National Gallery" from Hugovk flickr page https://flic.kr/p/naiid2

[140] http://www.niea.unsw.edu.au/people/professor-sarah-kenderdine (Accessed: 8th April 2014).

[141] http://alive.scm.cityu.edu.hk/projects/alive/ecloud-2012/ (Accessed: 8th April 2014).

Figure 32. Sarah Kenderdine and her ECLOUD Experience in action. Photos courtesy of Sarah Kenderdine.

This was done in conjunction with ALiVE and ACIM. In order to obtain the data she needed, she worked with a developer to create the call to the Europeana API. This perhaps suggests a mid-way compromise for humanists and social scientists who are aware of the potential of data within large collections, and how it might be obtained, but don't have the specific coding language and expertise to bring it to full fruition.

These examples illustrate the reality that the majority of users of cultural heritage APIs are still developers and computer scientists, although many of the most exciting applications are coming from clearly 'dual-skilled' individuals or organisations, such as Wragge Labs or Stripy Sock. However, this is perhaps inevitable given that by definition an API is a programming interface. There is also a small group of Humanities and Social Science researchers who are re-using the data they can obtain through a web-service or API. Who is doing the extracting of that data, however makes the difference. In the case of Sarah Kenderdine and some others, they are making use of developers to make the calls to the APIs and obtain the data they need. They have the expertise to know what they can do with the data when they get it, but they don't necessarily have the time or specific skills to write the calls to the API themselves. Identifying

and supporting both of these modes of usage is a key task for institutions looking to build usage of their APIs, as they show not just what the API can deliver, but what it can do.

One thing API-providers tend not to do: gather user and usage data

APIs developed by the large international information companies, like Google and Twitter, have a large enough reach and user base that they do not need to actively promote their APIs to potential users. This is not necessarily the case for digital cultural heritage projects, however. The strategies described above have developed to address this challenge, with widely varied success. Tim Sherratt at Trove/NLA comments that "It would be great to see cultural institutions doing more to watch, understand and support what people are doing with collections in their own spaces — following them as they pursue their passions, rather than thinking of ways to motivate them." But few cultural heritage institutions that provide APIs do in fact monitor their usage.

However, when approaching other API providers such as HathiTrust and the Finnish National Gallery, it appears others have not paid such close attention. This seems an important and yet oddly overlooked point. One of the principles of any product design is an evaluation process to assess impact of the product, who is using it, and what its strengths and weaknesses are.

Yet, while it would be ideal for institutions to be able to monitor use of their platforms and APIs, it is not always possible. The reasons why this data is not collected are varied. In some cases, this could be due to policy decisions or other national legal restrictions predating the development of the API made earlier ensure privacy in the process of allowing access, or legal restrictions within that country. The Finnish National Gallery, for example, does not currently monitor details of those using the API: having taken the policy decision to allow anonymous use.

"When a person applies for an API key, we ask for an email address but not a name or other personal information. We have wanted to make it possible to make the key request anonymously."

Sana Hirvonen[142], Finnish National Gallery.

The British Library, as shown above, is beginning to pay more attention to who is using their Labs services, and gathering this information through competitions and through working with the VISIBILITY tool developer Peter Balman[143] to see what kinds of things are being produced and by whom, as well as what developments might increase and encourage this use.

Where the digital humanist / social scientist can take advantage of greater access to digitised records is where development of APIs can come in.

5.3 Section 2: Researcher Behaviours With Regards to APIs and Data Generally

Research Methodology and Process

As discussed in the introduction to this report, it was initially not clear where we should draw the boundaries of a research project investigating API usage by researchers in the humanities and social sciences. Had we followed this most narrow definition, it would have been a short report indeed, as API usage is simply not a characteristic that researchers use when describing their projects, even in the (already relatively rare) occasions when it does occur.

We therefore decided to pursue a series of research questions, which we considered to be slightly 'upstream' from the central concern of API use itself. According to this approach, we believed that we would capture not just API usage happening below the visible level, but also behaviours and requirements that would predict the likelihood of API use, should an ideal tool and environment be available. We chose as well to complement this direct contact with researchers and the results of research projects with perspectives from developers (to develop our understanding of what could be done, were technical skills not a barrier) and library/infrastructure professionals (to understand their roles in supporting technical

---

[142] Sann Hirvonen, Educational Curator, Museum of Contemporary Art Kiasma. Personal Correspondence 6th June 2014.

[143] http://cargocollective.com/peterbalman/About-Peter-Balman

methodologies and in developing/promoting APIs). Through this approach, we hoped to be able to gain a full picture of what an ideal environment for API use by researchers might look like.

Our initial desk research indicated that very little published literature would be available to contribute to our work. We used what we did find as a baseline, incorporating it with the results of the Europeana Cloud Project Expert Fora. To elicit specific information, however, we conducted a series of 11 full interviews supplemented by 4 additional communications with key individuals not available for interview (see further information on this process below).

Key Baseline Findings from the Europeana Cloud Expert Fora

The Europeana Cloud Expert Forums were conducted from June to November 2013. There were 3 Forums in total, each covering a particular theme. The first Expert Forum was held in Dublin in June 2013, and brought together social scientists, humanists, librarians and computer scientists. The participants were selected in this forum due to their experience in using digital data, and were joined by members of the Europeana Cloud consortium who might not necessarily have had the same level of experience. Therefore, the groups were of mixed ability, but each had members who were very familiar with the kinds of things that can be achieved using metadata and web-services.

The participants were asked to evaluate Europeana in its current state, and provide a 'wishlist' of tools and services Europeana could provide. This 'wishlist' then assisted us to develop 'case studies' (scenarios) that could showcase the potential direction for Europeana. Many of the items on the wishlist included an improvement to the content (specifically the metadata), visual representations of the data (timelines, maps) and a means by which the data could be filtered.

During the second and third Expert Forums in which Humanists and Social Scientists were asked separately about their requirements from content and tools within Europeana, broad themes were highlighted. These groups were a little different, though, as they did not have so high a level of experience using digital tools. Many were familiar with qualitative analysis tools such as NVivo, but few retrieved the digital data themselves. Both groups of researchers were keen to see the same visual representation (timelines and maps) of the search results.

These fora showed that researchers in the Humanities and Social Sciences are interested in visual representations of search items, and information about how else the data is being used.

The researchers showed a particular interest in geo-mapping the data to reveal such geographical information as where the item in question is currently held, where it was created, where it relates to (e.g. is it a manuscript about the Crimean War, in which case, is it an item written on the battlefield, or a memo written in London, Paris or Moscow regarding strategy?).

Equally, timelines were strongly supported among the researchers, with similar questions being asked about the temporal information of an item or dataset.

Most importantly, however, they were eager to realise the requirement known in the information management community as: *"Simple search, advanced results."[144]* They were keen to have filters as part of the interface when the results of a search were presented. This filtering function should allow researchers to 'zoom in' to much more relevant results. This approach is often used by online shopping websites as a way to guide customers to products that would be most appropriate to the search they have conducted, without excluding products that may also be attractive to them. Amazon and eBay are go-to examples of this kind of adaptive hypermedia.

Terras et al investigated the potentials for High Performance Computing (HPC) in 2009, looking at whether Humanists could make use of e-Science technologies.[145] They found similar results when asking researchers what kinds of technologies they would put on a 'wish-list' for their studies. Filtering, visual representations and information on how the data was being used all featured prominently.

This focus from the expert forum on advanced search and visualisation highlight areas that could indeed be implemented through an API. But a specific requirement for an API never emerged, leaving open the question of whether non-technical researchers were aware of the potential for APIs in their research, whether the gap was simply too big between their skills and what an API could do, or whether other barriers existed to a wider uptake by researchers of APIs.

---

[144] Lehane, 2012.
[145] Terras et al, 2009.

Interviews

The interviews were conducted after the main desk research had been carried out. This enabled us to focus on the areas for particular investigation, drilling down into key lacunae we could identify in the literature, but still leaving us open to new discussion points as the interviews revealed them.

The list of interviewees was always intended to be representative, rather than comprehensive, across the profiles we had identified. Interviewees were drawn either from our own networks or those of other Europeana Cloud colleagues (including the ERCG and WP1), identified from exemplar projects and individuals we had come across in our desk research, referred to us by other interviewees or recruited through open information sources (Humanist blog, LSE blog). The interviewees were first selected. We also incorporated people that we had previously worked with, or had knowledge of use with APIs. We were then able to put these people into 3 main categories:

- Humanist and Social Science researchers working with big data
- Developers and Data Evangelists using Humanities data
- API Providers (e.g. librarians, curators)

Although we had different levels of success in identifying appropriate individuals from each category, we did manage to provide sufficient coverage in each category to achieve a compelling convergence among the interviewees on key points.

Interview set-up

Nominated interviewees were approached by email in the first instance. If they were unknown to us (or indeed, us to them) we introduced ourselves within the context of Europeana, and explained the aim of our research. We then invited them to take part in a Skype call at a time convenient to them. In some cases, where interviewees were not available for a full scale interview, we were still able to ask them a few questions by email (these are noted with a * below).

The final list of Interviews is as follows:

Profile of Interviewee	Number of People Interviewed	Interviewee References
Data intensive Humanist and Social Science researchers	4	Interviewees R1, R2, R3 (Humanities) & R4* (Social Science)
Developers using Humanities data, including 'data evangelists' and participants in API-developed hackathons	4	Interviewees D1, D2, D3 & D4
Librarians, infrastructure specialists and curators involved in supporting or promoting APIs	4	Interviewees L1, L2, L3 & L4

Where possible, we recorded Skype calls for future referral, but it was made clear that those recordings would not be made public. Only snippets from the transcripts will appear in this report.

Interview structure

The initial topic of conversation was determined by the background of the interviewee, and their experiences with APIs, data-centric humanities research or other Web-Services to date. In the case of API Providers, we would discuss their API initially, and what the aim of making it available was to them. We would then discuss if and how they participated in the research being carried out using the API, and if they knew of any projects that had used it recently. In the most part, the answer to this was 'no', for reasons outlined above.

For Humanists using APIs, we would ask them about their topic, and how the API or web-service had helped them to investigate questions that wouldn't be possible otherwise. If they had used a particular API or web-service, or had a relationship with a particular project or provider, we asked them the strengths and weaknesses of that service.

Over the course of the interviews, a key set of questions on key topics emerged: these questions are highlighted in the bordered sections below.

Interview Findings on Researcher Needs and Wants for an API

Over the course of the interviews, three main topics emerged as key drivers of data reuse across APIs or indeed across any services. These topics were as follows:

- Data: what researchers want it to be and what they want to do with it
- Technical Expertise: How to develop it or get access to it
- Environments: social and technical preconditions to data reuse

Each of these issues will be discussed in turn below.

Data

- How much data should an API provide access to?
- What are the key characteristics of this data (what is acceptable/undesirable)

What researchers want is data, and several researchers commented that what they really want from a digital data service is that it 'just give them everything.' Completeness of data is therefore a key requirement of what the API delivers, which has implications both for what the API encompasses, but also places a high bar for the quality of the metadata underpinning any filtering that occurs. The consistency and quality of metadata can therefore make or break a project's usability.

From a user-perspective, R3 reflected his frustration at the inconsistency in metadata in some of his previous projects. "Many projects have their own descriptions of authors of texts and manuscripts. This was not just redundant, it was incoherent!". The multiple standards for metadata can make it confusing and frustrating for anyone working with multi-institutional data collections. Europeana has tried to get over this hurdle by allowing metadata fields from all metadata standards to be included and completed. But this has run the risk of meaning that there is confusion over what certain fields refer to; for example the field 'date' means many things to different people: date of creation, date of digitisation, date of acquisition, etc.

For D3, from the Decipher[146] project, the quality of the metadata became an issue when trying to develop on top of Europeana. "...people would get the wrong stuff because the queries were getting large hits. But really, when they started drilling down into the results, they ended up with just a thumbnail". The promise of good quality metadata can be almost as frustrating as none at all.

This is something that Europeana recognises and is working on. A recent study conducted on Europeana use among researchers in the Humanities revealed further issues in the metadata and lack of actual content that the Europeana Cloud project hopes to ultimately tackle. The Europeana Cloud Content Strategy[147] acknowledges that the content in Europeana was not subject to a particular content strategy from its inception. The content was gathered from CHIs across Europe who were uploading their metadata under very lenient circumstances. This was of course an advantage at the time as it meant that Europeana had a rapidly expanding library of data available via its search portal. But the hangover from this means that the content is very difficult to search, as search terms are more or less useless on metadata that is inconsistent and/or incomplete. This will be a significant barrier to the development of research via the Europeana API.

But this problem isn't limited to Europeana. D2 from CLARIN also describes their frustrations at setting up reliable metadata. This is most prevalent for them in their Virtual Language Observatory (VLO)[148]. "We have a metadata taskforce which is creating tools to benchmark quality. This can be very labour intensive, and requires a lot of interaction with content providers. It's a continuous process that requires some effort. But on the other hand if you have unified easy-to-use tool, people get confronted with the poor quality of the metadata." CLARIN often has to field complaints about the quality of the metadata, but as our CLARIN contact explains, the ironic thing is that often the individual or organisation complaining about the metadata is responsible for supplying the metadata in the first place. This mainly occurs within the transcription metadata, with data being either outdated or no longer linked to online resources.

---

[146] Decipher project - http://decipher-research.eu/ - accessed 15th October 2014

[147] Europeana Content Strategy, Europeana Cloud Deliverable 1.2.

[148] http://www.clarin.eu/content/virtual-language-observatory (Accessed: 16th July 2014).

In October 2013, the Digital Repository of Ireland (DRI) released its 'Metadata Factsheet'[149] outlining its overall approach to gathering metadata, in which they listed 5 metadata standards that they would support: Dublin CORE (Simple and Qualified), MODS[150], EAD[151], and MARCXML[152]. Certain fields are to be compulsory, and all fields required in these metadata standards will be available to encourage richer metadata ingestion. This does still raise the issue of competing fields with similar names, and how to handle that.

Data should therefore be correct and complete, and properly structured so as to allow a rich experience of search and retrieval. But what about the source of the data? Digital libraries must select and curate what they deliver as well, actively or tacitly, which includes in the current age the decision of whether or not to include crowdsourced information. Whether or not to do this is the first question: researchers will be suspect of data they do not know the provenance of, and if the difference between authoritative data (such as, for example, institutional metadata records) and crowdsourced data cannot be made clear, then the use of such data presents a serious risk to the credibility and usability of a resource. Apart from that, there is also the long discussion of the ethical dimension to not making publicly available information generated by the public which unfortunately extends beyond the scope of this report.

But working responsibly with crowdsourced data poses more questions that the binary of whether to use and a UI challenge of presenting it. L1 at the Bodleian Library, Oxford reported on an occasion when they investigated crowdsourcing as a means to create and store data. The Library partnered with Zooniverse for its "What's the Score" project[153], asking members of the public to complete information about certain items. According to L1, however, this was less than successful. "We were too librarian about it, and tried to say that for things to be fully described, participants had to fill in 35 fields, which isn't really what Zooniverse users go in

---

[149] Digital Repository of Ireland, (2013) *Metadata and the DRI* - http://www.dri.ie/sites/default/files/files/dri_factsheet1_ver2.pdf

[150] Library of Congress - Metadata Object Description Schema - http://www.loc.gov/standards/mods/ (Accessed: 15th October 2014).

[151] Library of Congress - Encoded Archival Description - http://www.loc.gov/ead/ (Accessed: 15th October 2014).

[152] Library of Congress - MARC in XML - http://www.loc.gov/standards/marcxml// (Accessed: 15th October 2014).

[153] Zooniverse / Bodleian Library "What's the Score" project - https://www.zooniverse.org/lab/wts

for![154]". The Bodleian ended up having to re-evaluate its metadata fields and reduce them in number in order to make it more user-friendly, but according to L1 it was still too many.

Data Re-Use

It must be remembered that reuse of data via an API is a subsidiary question to the larger issue in the research community of the reuse of any data. Data produced by digital projects is not further developed by other, for reasons that have little or nothing to do with how it is made available for access. What we have seen from the responses, and from our desk research is that building an API doesn't guarantee use of the API, or re-use of the data. R2, for example, admits his pessimism around this, having seen many projects invest in a great resource, only to see it used almost exclusively by the person who built it. He suggests that the research question is the thing that fuels the investigation, and it is rarely the other way around. In other words, researchers will be interested in a subject, and formulate a question before going near any silos to look at data. They might even resist data and tools that are being 'pushed' towards them, and data providers can therefore find themselves pushing against a closed door.

This 'pushing' motion seems to be echoing what Fred Gibbs says when he comments that it is easier to 'slide' these products towards Humanists than for it to be the other way around. So while these products might be built, all the providers can do is let people know it's there, let them know what it can do and let the researchers figure out for themselves how they want to use it. This perhaps is reflected in the 'data evangelist' approach. By showcasing what can be done with a particular API attached to a collection, it might prompt or inspire researchers to think about what's possible.

Technical Expertise

What support is available for researchers who don't have a high level of technical expertise? Should training be made available to researchers to give them the means to access and re-use this data themselves in a manner best fitted to their research methodology? If so, training in what? To what level?

---

[154] R2 Interviewee - personal correspondence - 8th August 2014

Many of the humanities and social science researchers were very adept at using languages like Python (R1), R (R2) or Perl (R4) to write software code. Of all of the researchers we spoke to, however, only 1 (R4, a social scientist who developed a web crawler to pull articles from Google News and the RTE[155]) used coding to pull data from the internet. Instead, these researchers took their datasets, usually acquired through a download or data transfer from a trusted human partner, and structured and manipulated them off-line.

The question of whether social science and humanities researchers need to code seems to be a divisive issue. Those working on the software development side consider a researcher with a little knowledge to potentially be more likely to break something or otherwise misunderstand a system developed by software professionals. Researchers with coding skills generally take the opposite perspective: for example R2 told us. "I've come to the belief over the last 20 years that we really have to get over the idea that programming is a foreign occupation for Humanists. It's ridiculous that we discuss if someone needs to be a programmer or not" he says. This perspective has led to innovations in some of the basic building blocks of humanistic training, such as example reported to us by this same interviewee of one American university changing its requirements for the Doctoral degree to allow one of the traditional foreign language requirements to be fulfilled with a computer language.

But these polarised positions do not assist us with the very practical question of what skills and what level of ability does a humanist or social scientist need to allow them to use code as a part of their methodology, to experiment enough to understand the implications of a resource like an API for their work?

L2 described an example of a Bioinformatics researcher who was already reasonably familiar with coding, and taught himself how to use an API to archive his own workflow. Equally, D1 relays how, 3 years ago, two very traditional humanists of his acquaintance at Huygens who took it upon themselves to learn how to use digital tools for network analysis and transcription purposes. They are now highly specialised in using these digital tools.

---

[155] *Radio Telefis Eireann -* the Irish national broadcaster

These examples show what can be achieved with a little support and a compelling research application, but as D1 says, a positive attitude towards learning a new skill is required. L2 commented that, in his experience, the scientists at the younger end of the community tended to be more willing to learn these new digital skills than the older humanities researcher. This may not be typical, but D1 also commented on how PhD students are perfectly willing to learn coding in languages such as Python or using XML and HTML if it can be directly applied to their research.

Training and support for researchers

If technical expertise is a barrier to the application of advanced data methods by researchers, then promoting an API could potentially involve providing training. This was an element of both the V&A site and EDINA's. But those resources do not generally begin with the most basic rudiments of a programming language: that would potentially be too detailed, and not necessary in the context of the wider environment. The self-motivated researcher does have a wealth of resources available to them, if they have the time and inclination to do so. Software Carpentry[156] is an online community run by volunteers providing open access lessons to researchers (currently mostly in the sciences) who want to learn how to code. Their goal is "to make scientists more productive, and their work more reliable, by teaching them basic computing skills"[157]. As well as taking free lessons, more experienced users in the community can create lessons to help their fellow scientists. As they explain the rationale of the movement on their website:

"The problem is, most scientists are never taught how to do this. While their undergraduate programs may include a generic introduction to programming or a statistics or numerical methods course (in which they're often expected to pick up programming on their own), they are almost never told that version control exists, and rarely if ever shown how to design a maintainable program in a systematic way, or how to turn the last twenty commands they typed into a re-usable script. As a result, they routinely spend hours doing things that could be done in

---

[156] http://software-carpentry.org/index.html (Accessed: 16th July 2014).

[157] http://software-carpentry.org/faq.html#what-is-swc (Accessed: 24th September 2014).

*minutes, or **don't do things at all because they don't know where to start**." [158] (our emphasis)

Equally, Codecademy[159] has a series of online bite-sized lessons available for free which are written by volunteers. In the case of Codecademy, however, the frustration comes when the 'student' becomes stuck in a particular task, and is unable to get the support they need to find the solution in a timely manner. Forums are available, but again are run by volunteers.

The self-taught route, however, has its own barriers to contend with. Finding the time, and indeed motivation can be tricky enough, but an additional lack of support available in that instance can also be demoralising.

Summer Schools are available for early stage researchers, such as that offered by CLARIN-DE. These 'PhD-Days' offer specific training in skills. In March 2013, PhD students were able to learn skills such as corpus annotation using their tools (Weblicht)[160] and how to statistically interrogate the data they obtain from corpora. So they cover the 'what', the 'how' and the 'now what' of data collection and re-use. DARIAH also offers summer schools in areas relevant to digital humanities to early stage researchers. For example in 2014 the DARIAH summer school covered digital text analysis and constructing and building 3D worlds[161]. Summer Schools such as these are clearly important for developing skills for researchers, but the focus on more digital means of obtaining publicly available data indicates the growth in both recognition of the importance of using digital data and the gap between the skills of researchers and the potential of the data.

The British Library has put together a programme for its curator and librarians to open up the possibilities for them to use digital means of research, such as APIs. This was more of a staff training initiative rather than a workshop for potential users of British National Bibliography

---

[158] http://software-carpentry.org/blog/2013/06/lessons-learned.html (Accessed: 24th September 2014).

[159] http://www.codecademy.com/ (Accessed: 16th July 2014).

[160] The *Clarin-D Doktorandentage - Corpora -* University of Saarland, March 2013 - http://fr46.uni-saarland.de/lsteich/ClarindDS2013/program.php (Accessed: 15th October 2014).

[161] http://www.gcdh.de/en/teaching/2014-dariah-international-dh-summer-school (Accessed: 29th September 2014).

(BNB) data, but it highlights the importance of library staff keeping ahead of the curve. The structure of the workshop could potentially work as a model for our own workshops in Europeana. Following this example might provide a gateway for Europeana to showcase the potential for the Europeana Research Platform.

Working with Developers

Of course, the alternative available to a researcher to acquiring the technical skills required to develop data themselves is to work with a developer. This approach has worked very well for researchers like Sarah Kenderdine, though it is perhaps its own skill set to understand how to to construct and manage these relationships productively. This potential is further supported by Paul Hagon of the NLA, who discussed: "guid[ing] the developer to help them fine tune their product & give them insight into the intricacies of the data" as one of the stages for API promotion and use. There is an implicit assumption that a developer is there doing the technical work for them. These relationships are expensive and therefore exclusive, however, and require the researcher to conceptualise in a language they don't themselves know, a perhaps more difficult task than learning Python or Ruby. Many would question whether the result is better, but if lowering the barrier to researchers is the ultimate goal, assuming access to technical developers will not achieve the goal (unless the institution that 'owns' the API can somehow facilitate this, as in the British Library Labs Competition structure).

Environments

Where is this kind of research being done, and is this in a collaborative or individual manner? Does the environment allow for collaborative research? Should libraries become this environment? Or does the expertise of the library play a limiting factor or barrier to the research?

Should support for users take the form of an intermediary (a developer, or indeed a library) to assist them in building a means to retrieve the data?

Should an API provider also provide tools?

Are there certain data sets that would be more attractive than others as an API (news feeds, change rapidly)

How data is created, stored and made accessible constitutes an environment which, viewed at a macro scale, includes not just technical elements, but human and infrastructural/institutional elements as well. What are the optimal characteristics of this environment to supportive to research or scholarly activity, in particular activities which challenge the scholars and their accustomed methodological approaches?

Libraries

L2 believes that this can be determined through studying the workflow, and identifying the point at which software and web services (in this case, the archival tool, Dataverse[162]) becomes most useful. He describes an example where he brought together PhD students from various disciplines within the Humanities and Social Sciences to work on an informal project that collected data from different sources. The group collected and cleaned the data. While the PhD students conducted this work, L2 and his team at the Library documented the workflow of the group. This then provided not only a data-set of the collection created by the students, but also of the workflows and practices they undertook to reach completion of the project. In this instance, the library almost became the research while also facilitating the research on the collections it held.

L1 agrees, to an extent. As a Humanities librarian, she generally doesn't hear about it, when someone is using APIs for their research, as they can look after themselves. But she believes this is as it should be. "Libraries shouldn't be the gatekeeper. We shouldn't be making judgement calls, and we shouldn't be doing the work for people. These are the methods and workflows for the research - we can't be making decisions as to how the data should be used". That said, L1 has, on occasion, partnered up with researchers to do some of the more technical work required to retrieve data sets, but she stresses that this is very much led by an understanding of the researcher's question. The library's involvement is more to suggest a methodology, leaving the final decision to the researcher.

As the nature of the data the researcher is using and creating changes, one would expect that the role of the library would change as well. L2 seemed to feel that this could involve a greater

---

[162] http://thedata.org/ (Accessed: 15th July 2014).

support by libraries in the future for methodologies: whatever they are. But in terms of having the skills or human resources to provide that, we are still very far from this potential future.

Tools and Interfaces

The rise of new model digital infrastructures alongside the old ones offers different options for addressing the environmental issues of data reuse and API uptake. According to D2, the work at CLARIN-DE suggests that more remote means might be the answer.

CLARIN-DE provides access to data via different web-servers from centres across its network. There are online web services available via the CLARIN site that allow users to build and tailor their own tool-chains for data collection and re-use. CLARIN lists these tools, which can be deployed from the CLARIN graphical interface. This provides a visual way of creating tools to access data, for example Weblicht[163] which can make use of tools in a customisable 'chain'. This is, the outputs of one tool in Weblicht can then be fed into another tool available on Weblicht (and so on). According to D2, this user-interface will hopefully break down perceived barriers to tools and data for those who are not so specialised in using such means of research.

Once a user is logged in to Weblicht (using either a Shibboleth or CLARIN username), they have access to the tool, which allows these chains to be created. This can be used in "Easy Mode", "Features Mode" or "Advanced Mode". Currently, the 'easy mode' is only available in German.

---

[163] http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page (Accessed: 15th July 2014).
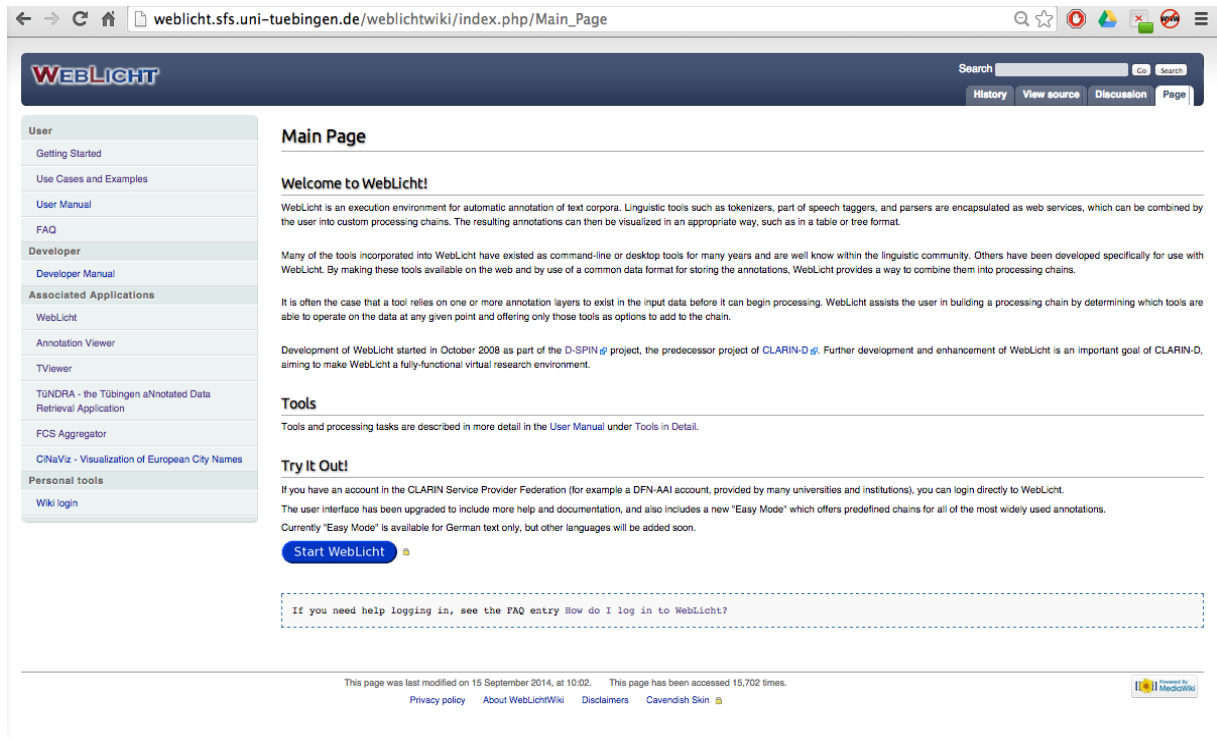
Figure 33. Weblicht home page (after login)
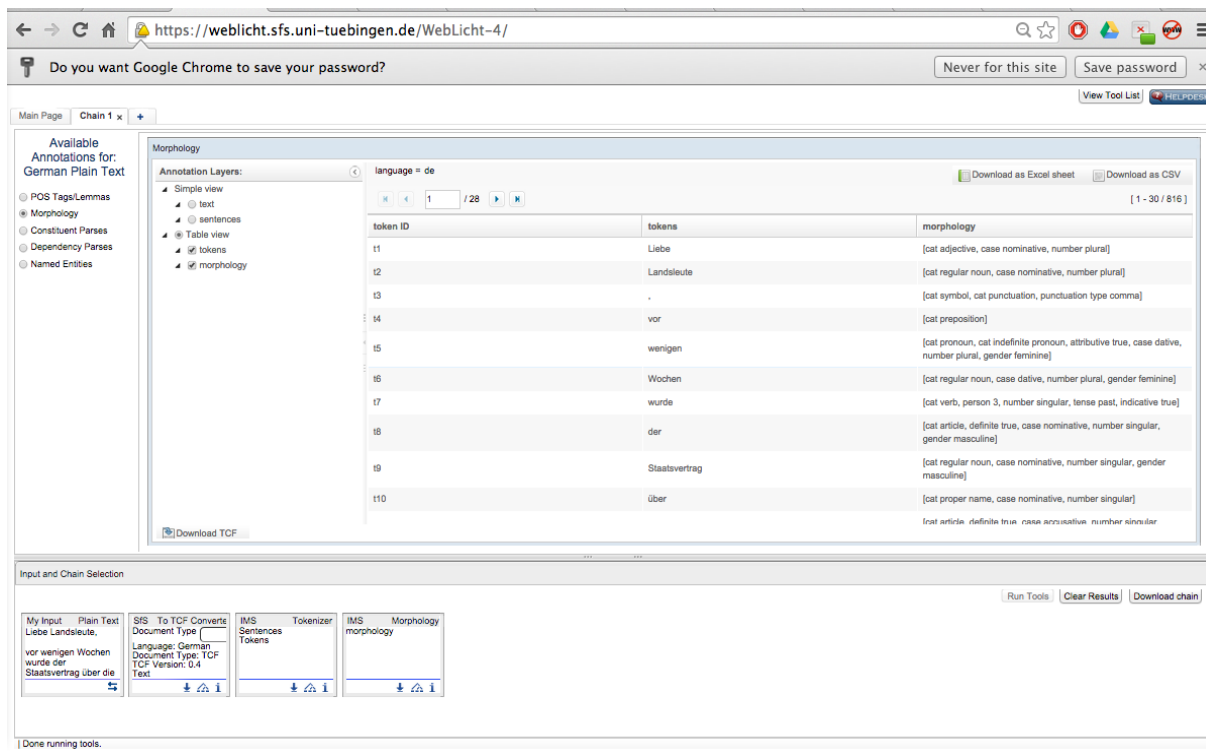
Figure 34. Weblicht's "easy mode" tool in use, annotating input text by morphological elements.

The CLARIN tool chain, like the Hathi Trust Research Centre Meandre tools, offers opportunities to manage and manipulate data, opportunities that might not be within their grasp of a researcher otherwise. The DPLA hosts another model, where 'Apps' are developed externally and hosted on the site. But lack of user data makes a deep analysis of the utility of such tools difficult.

R2, however, is not such a fan of 'Humanist friendly interfaces', as he believes that ultimately they don't work. He feels instead that a Humanists needs to understand the modules in systems in order to properly get to grips with the data. A Humanist needs coding expertise to be able to understand this. Graphical interfaces are, he says, good for smaller jobs, but will not be able to handle big data as the builders of such interfaces can't predict all the different ways in which to manipulate data; "the tweaking is where the innovation comes in".

Other interviewees are more measured in their assessment of the value of tools. D1 expressed a hesitation about tools that did too much, or operated at the wrong level. But certain tools, able

to batch process otherwise time intensive or tedious processes, could certainly bring value, but he, like R2, did not necessarily feel that most tools gave more than they took away.

5.4 Section 3: Recommendations and Best Practices Europeana might Apply

Recommendations regarding Data, Metadata and Content:

- To be considered a truly useful resource, Europeana will need to improve its metadata, either through funded or a carefully constructed crowdsourcing project
- It should also if in any way possible provide access to the underlying sources

For a repository, platform or portal to be successful, the one thing it supplies must be the thing it's best at. For Europeana Research to be success in its execution the metadata must be in good working form.

Other areas within the Europeana Cloud project have already begun looking extensively at the content within Europeana, and how this might be built upon and improved. Much of the current failings within the metadata have been acknowledge both within the Content Strategy (as discussed) and in various Expert Forums conducted as part of Work Package 1 (T1.5). We also acknowledge the responses from various researchers who have used Europeana and found it lacking. This was also the response during interviews conducted by Dean Birkett at Europeana in September 2014[164].

As we've seen, the promise of content can lead to frustration when actual content isn't available. This has become an issue when a researcher is using Europeana for anything other than metadata, but also when results don't match user expectations. Links to the original items will need to be monitored more closely to ensure that a link hasn't become obsolete, by using

---

[164] Europeana Publication (forthcoming).

persistent identifiers if possible. The Europeana Cloud Content Strategy report revealed that many of the items currently held within Europeana no longer exist - at least not at that location - and therefore need updating.

We have also seen how other organisations have responded to this need for concise and clear metadata. The Bodleian Library at Oxford tried to use a crowdsourcing approach to enriching its metadata, but found the approach to be ultimately 'too librarian' and therefore off-putting to the average Zooniverse contributor. We would expect that Europeana would face different primary challenges, for example recruiting participants to enhance/correct metadata and ensuring that crowdsourced and official data did not become mingled and confused. But the first of these can be managed via clever user interaction (gamification, rewards, development of bespoke user communities, building on the success of earlier Europeana crowdsourcing initiatives) and the second by clever UI design (Amazon, for example, makes the difference between catalogue data and user feedback very clear on their site).

The Digital Repository of Ireland found that while they wanted to maintain a trusted repository for digital data that was robust enough to support reusable metadata yet not apply any restrictions on the ingestion process for its 'clients', they still had to overcome the issue of conflicting fields. Their guidelines are forthcoming, but their interim solution has been to adopt 5 metadata standards and integrate them into their ingestion form.

The issue of maintaining good quality metadata and access to as much underlying data as possible is therefore the baseline for success for Europeana Research. On one hand, Europeana is nothing without its aggregators uploading and enriching metadata, but it will be next to useless to researchers without trustworthy data to download.

Recommendations regarding technical skills and support for researchers:

- Make sure the API is well documented, at multiple levels and through multiple media
- Offer tutorials with clear technical prerequisites, and pointers toward other sources of technical training
- Host periodic training 'workshops' (either online or in person) to allow those keen to learn new digital techniques alongside Europeana content.
- Host periodic 'workshops' for Europeana affiliated institutions

- Contribute to the formation of a 'software carpentry' interest group for cultural data
- Offer a marketplace for linking developers with researchers
- Make sure that the platform features a few actual worked use cases, showing both how the research question was developed and the results delivered (and by whom)
- Make sure use cases are made visible and eye-catching, both on the Europeana website and via other channels

Many of the sites we reviewed carried information about their API, with helpful notes as to how to use them. We have referenced many of these further on up in this document so it is not necessary to reproduce that here. The ease with which the API can be used depends on how well written this documentation is. To a complete novice embarking on their first forays into API use, this documentation can still be slightly baffling as they try to navigate the technical terms and envisage just what it is the API is going to do for them, and what the output will look like. The DPLA website is particularly good at guiding 'newbies' through the process of making their first call to the API[165].

If Europeana wants to make itself useful to novice API users from the Humanities and Social Science, it will need to also follow this guiding process for the API. This, coupled with workshops either in the form of physical meetings, or YouTube videos, it might allow the novice digital humanists the opportunity to 'catch up' and make use of the API on offer.

We have also seen examples of in-house training in the likes of Maryland Institute of Technology in the Humanities, and the British Library, the aim of which was to educate the staff as to the potentials of data re-use within their own collections, so that they may be able to make recommendations to their customers. Europeana, as an entirely online resource, has not at the moment such specialized 'staff' to assist. However, the aggregators are more typically physical buildings, which do have staff onsite to assist. Training options may therefore be useful to educate those staff members as to the potentials of Europeana.

---

[165] DPLA, "API Codex" http://dp.la/info/developers/codex/ (Accessed: October 8th 2014).

Finally, Europeana will need to inspire people to work with their data and give them the confidence they will achieve results. The Australians seem to have mastered this, with Tim Sherratt acting as the data evangelist par excellence, and tools like *Forte* proving the use of the data. But even the DPLA's humourous Twitter feed contributes to the overall awareness of their content, and Europeana will have to ensure it catches researchers' attention and imagination before it can assist them in changing their research.

Recommendations regarding the Research Environment:

- Offer tools and the potential to build tool chains, but don't offer too many at the top level and make sure that each provides a unique service. Others the community developed could perhaps be registered for use via a Wordpress-like widget directory
- Build content and use cases from the bottom up by incentivising the formation of user communities

Research Communities

There have been discussions about an online community that could also support the Europeana API. Inspiration could be taken from the Software Carpentry model whereby experienced users of software create lessons for other community users. Admittedly this is mainly within the sciences at present, but the model itself presents an opportunity for the Europeana audience, which from the content strongly suggests Humanists and Social Scientists working in areas such as Social History, Sociology, Anthropology and Art History (among others).

The idea of a forum has not proved popular, and it would of course create the need for increased staffing resources at Europeana to continually monitor. However, there is potential within Europeana Labs or Europeana Research to allow researchers to build their own communities around collections. These communities could be user-driven and monitored, similar to how Facebook operates groups and pages.

Tools and chains

Obviously, there are split opinions about the place of tools in the humanist's data workflow. But it is beyond doubt that certain tools would certainly be of value, so long as they do not get lost in a morass of applications too specialised or under developed to be an asset to the collection. In

thinking about tool chains, we also explored the free service Yahoo Pipes, as an exemplar for how data sources and manipulation tasks might be combined for non-specialist users. Yahoo! Pipes has been available online for some time now. The service provides a mean to 'mash' or aggregate web feeds and other data from various online sources, and creates an output using that data. Between the input and output phases, the data can be filtered or 'unionised' to custom the output. The most basic example is creating an RSS newsfeed from a website. The resulting output can then be given in a variety of coding languages (RSS, JSON) or formats (MyYahoo, .csv) and put on a user's website. For example, if someone blogging on a particular topic wants another way to demonstrate what is happening within their field, they might create an RSS-feed based on Pipes, and include it on their blog as a 'widget'.

To find out how they are being used - Matthew Dinmore and C. Curtis Boylis conducted a study of more than 30,000 Pipes compositions to determine end-user behaviours (EUBs)(Dinmore and Boylls, 2010). The Pipes platform is such that it is more typically used by end-users, as it is a very visual platform. Dinmore and Boylls wanted to see how these end-users might compose solutions to problems that are very specific to their domain. They investigated a total of 38713 'pipes' published by 22285 unique accounts, or 'users' (although the authors were at pains to point out that they cannot categorically say that all the unique accounts were indeed 'real humans'). They did not characterise the level of expertise each user had of programming, but they did acknowledge that there would be differences in approach between a professional developer and a hobbying humanist. They then separated the 'sub-pipes' from the 'pipes' as sub-pipes "take on the persona of a Pipes data-source module..". Modules are the elements that a user brings into their own workspace to create their 'pathway'. There is no limit to how many of these modules a user can employ, so the composition can get as big as they would like it. Despite this, however, Dinmore and Boylis found that the average number of modules per pipe was 4, and an entire composition was only averaging at around 40 or fewer modules.

They also found that when it came to choosing the 'wiring' of their modules, whether the data-flow is 'split' or 'merged' throughout the composition, or even put into a 'loop'. However, from the data they analysed, they found that users tended to keep this quite simple, with the number of modules usually mirroring the number of wires, keeping the number to a minimum. This indicates that users preferred to create a Pipe that took a very linear form, perhaps designed to answer a very specific question.

Their overall conclusions from their data seemed to suggest that users of Yahoo! Pipes only made use of a small number of the features available to them, preferring linear straight designs to the multi-branch pipelines that were possible. Dinmore and Boylis suggest that rather than looking to make an elaborate data-flow that could be re-applied in future work, they tended to go for as simple a model as possible that answered the question before them. As they put it "... users set design-time parameters - as few as possible - to achieve an objective, again asking the question 'am I done?'" (Dinmore and Boylis, 2010).

Limitations in APIs and other Web-Services

As we have seen, technological know-how can have an impact on a researcher's willingness to use APIs and other linked-data techniques. Many researchers still prefer a less-technical, more 'analogue' (non-digital) approach to their research, and don't see the need for more digital means. And this is fine. Not all research *should* be digital. However, there is increasing use of digital items in research, and indeed digital approaches open up some very interesting questions for future research.

The lack of take up of web-services among humanists is that perhaps there is a mis-match in what they want, and what developers think they want. D1 raised this point when he was describing how few opportunities there are for researchers to get to the data. He suggests that the leap to allow API access to data is often the developers' perspective to answering a question. In other words, the limitations of APIs are often set by the lack of understanding developers might have with researchers' needs, and appreciation for the challenges their research questions pose. On the other hand, resolutely non-technical humanists may not even have the capacity to imagine what such an approach might mean for their research, much less have a mastery over the vocabulary required to describe their desired research process to a developer. In deciding how and indeed whether to create a researcher platform for Europeana, these macro-level considerations must also play a role. Rolling out an API for researchers linked to Europeana will require fundamental change not only in Europeana itself, but in its users. If viewed as a long-term investment, such a development could be instrumental for cultural research in the digital age. If viewed only in the short term as a technical development only, our research indicates it will almost certainly struggle to find its place.

6. Research Themes for Aggregating Digital Content

At a time when digital research infrastructures across Europe attempt to support digital research in the Humanities through meeting documented user needs of the so-called research communities we considered that it is not particularly agreeable to view the Humanities and the Social Sciences as two solid disciplines, or, in fact, any particular disciplines within them as self-contained when almost all empirical studies on user requirements in the context of digital research infrastructures have revealed numerous underlying variations in the ways researchers carry out research.

Work conducted in the context of other EU infrastructures (ie EHRI, ARIADNE et.al.) has convincingly showcased that, similarities notwithstanding, research methods, activities, resources, research goals, schools of thought as well as interaction with data, both digital and analogue, vary between sub-disciplines in the Humanities and the Social Sciences. Such an approach, we believe, informs and affects the concept and design of effective user-centred digital research infrastructures. Needless to mention, research sub-communities are neither autonomous nor self-sufficient. Only those digital research infrastructures which address this granularity in sub-communities of practice and their respective user needs are to be meaningful to researchers. As Ian Hodder put it recently, "one of the things I thought was very important during the 1980s was the idea that culture is meaningfully constituted. I still think that is right. But now, I put the emphasis on the "meaningful constitution" rather than the "cultural" bit. (...) I prefer to break it down and talk about the various processes that constitute it"[166].

To this end, and within the course of our work within this Work Package we soon realised that a focus on particular research communities, in the manner in which WP3 built its scenarios, would provide much better insight to both the User Requirements, tools use and Content Strategy recommendations (see chapter 7 below) than a generic approach to what is called the Humanities and Social Sciences researchers. Based on the respective areas of expertise of the members of this Work Package, but also of other, external colleagues we decided to concentrate the final piece of our empirical research on particular research areas in the broad

---

[166] To The Trowel's Edge: An Interview with Ian Hodder,
http://ucexchange.uchicago.edu/interviews/hodder.html (consulted on 07.10.2015)

fields of the Humanities and the Social Sciences. Thus, after consultation with the Research Communities Advisory Board, we focused on the following areas: Parliamentary Papers, Population Displacement, Political Propaganda, Children's Literature, School History, and The Ancient Mediterranean. In each of these areas, we conducted both background as well as empirical research, through discussions with researchers as well as workshops, in order to further deepen and also consolidate our understanding of the ways researchers make use of data, digital and non, and what needs and expectations they have of existing tools and services they use in their respective areas of research. These individuals and groups of researchers were selected in accordance with their specific area of expertise (as research topics were quite particular per se), as well as their familiarity (though not necessarily extent of use) with Europeana. Certainly there might have been more - or, for that matter, more appropriate researchers – we could have contacted, but the ones chosen were considered absolutely fit for this purpose.

6.1 Parliamentary Papers in Europe

What is your research topic?

The emphasis of this research topic is on the state of digitized Parliamentary Papers of European Union member states. The primary focus lies on their availability, but also investigation into how the data is used by professional research has been undertaken. In a final step the research potential of digitized Parliamentary Papers is examined as well.

Can you give a brief description of the topic, including the chronological and geographic span?

Parliamentary Papers are a rich and diverse source that is utilized by different disciplines in the Arts and Humanities - not only as a source for straightforward political history. The main emphasis lies on transcriptions of parliamentary debates in the main debating chambers. However, this is but one subcategory of a very diverse field. Parliamentary Papers can also include session papers, committee reports, petitions, etc.

Geographically speaking we concentrate on the European Union member states and their main debating chambers (national assemblies or upper and lower houses, depending on the administrative structure). This means, however, that regional parliaments, which are existing in numerous states, are not taken into account. The reason for this is that due to the diverse

parliamentary architecture within Europe a wealth of local knowledge is needed to yield high-quality results.

Many of the European parliaments commenced during the nineteenth century to produce Parliamentary Papers in a more structured manner as many of them introduced constitutions. Thus, the oldest digitized sources date from this era. But European nation states have very diverse histories. Thus, the time frame for each individual country rests of course on its administrative history.

It is worth adding that the focus of this work has been on the historic material. Many European have a great deal of material published from the last ten or fifteen years so. The arrival of the Internet has made the publication of all kinds of parliamentary outputs feasible. For instance, the Estonian Parliament has, amongst other collections, the *Riigi Teataja: State Gazette : legal acts in Estonian and English*, *Minutes of the sessions of the Maanõukogu* and *Asutaw Kogu, Minutes of the sessions of the Riigikogu* and *Videos of the Parliamentary Sessions*. However, this piece of work has not explored the availability of this contemporary material

Can you list a few possible research questions emerging from this topic?

{See below}

What are the academic disciplines corresponding to this topic?

These two questions are answered together, by stating possible research questions and their corresponding scientific field. It needs to be noted that these are only examples from a large range of possible questions.

History: By applying a comparative angle it could be asked how the development of parliaments differed within Europe. How differed the debating cultures within Europe? How were historical moments (e.g. crises, wars, etc) debated in parliaments?

Sociology: Who were the actors in European parliaments? Do they have similar social backgrounds in European parliaments? This topic has been tackled in a more traditional manner. One example is the study, published by the Austrian Academy of Sciences on the Austrian parliament's members between 1848 and 1918. Having digitized data ready would lift these studies to a new level by offering the possibility of comparative research.

Economics/Economic History: How were taxes, duties and other economic measures discussed? Was there an increased attention to economic arguments within debates? Which aspects received special attention? Which parties backed which budget cuts or increased spendings?

Political Science: How did certain groups (e.g. the extreme left or right) act in Parliament? What were the topics they were covering in their speeches?

Linguistics: How did the language employed in public debates change? What kind of vocabulary was employed in order to bolster an argument.

History, Political Science, Sociology are three of the most obvious fields of research, but the resources are conceivably of interest to anyone in the social sciences and large chunks of the humanities. For example, anyone studying literary, musical or artistic history may well find attitudes towards the fine arts and particular artists, cited in the Parliamentary Papers.

For work that is already being done in the realm of Digital Humanities - which is of course influencing possible research questions - some indication is given in question 9.

Can you identify any Europeana datasets which could support research on the topic?

Europeana has a mass of data related to parliamentary processes around Europe. However, initial exploration indicates much of the data is not the Parliamentary Papers themselves, but related documentation. There are two key reasons for the limited usability of Europeana for this topic. Firstly, Europeana presents many documents from the same series as single, independent documents. This means in turn that a large number of hits are returned when searching a specific term relating to Parliamentary Papers. Secondly, as full-text search is unfortunately not available for documents in Europeana, queries need to be based on keywords likely to be found in title or abstract of a document. However, such keywords might return a large number of hits as they also appear in documents not related to Parliamentary Papers. For example a search for Hansard - a term used for transcripts of parliamentary debates - returns 28,010 hits. This number is quite simply overwhelming and an in depth analysis would be very time consuming. Thus, due to the language barriers and the diverse labels used for this material, it is rather cumbersome to uncover these sources in Europeana.

The following links represent examples of Parliamentary Papers within Europeana.

- British hansards in Europeana:

  http://www.europeana.eu/portal/search.html?query=title%3Ahansard&rows=24&qf=PRO
  VIDER%3A%22The+European+Library%22&qt=false

- Documents relating to the German Democratic Republic in Europeana:

  http://www.europeana.eu/portal/search.html?start=457&rows=24&query=sed&qt=false&
  qf=PROVIDER%3A%22Archives+Portal+Europe%22

- Documents relating to the transitional time in Estonia

  http://www.europeana.eu/portal/search.html?query=stenogrammid&rows=96

Can you identify any other datasets available on the topic?

{See below}

> Do you have any indication of who is responsible for these datasets? How are they licensed? Would it be easy to aggregate this content into Europeana?

We have tried to identify the state of digitized Parliamentary Papers for every European Union member state. The results can be found by following this link: https://docs.google.com/spreadsheets/d/1LLj-pPmo2WG1YWqA18U5hSAv9TpnZX4_wRa5bcPgvFo/edit?usp=sharing

A fruitful approach is to differentiate between historic and recent Parliamentary Papers, the latter describing digitized material of the last twenty or so years. Our research shows that almost all countries offer access to recent material. Only Cyprus has not any digitized material available[167]. Other national parliaments offer recent material only. This is the case for Luxembourg, Romania and Malta. Their digital holdings start in 2002, 1996 and 1991, respectively.

Although perhaps surprising at first glance a fair number of states offer historic, digitized material on their parliamentary proceedings to the public. For ten countries the digital collection even starts in the nineteenth century, while some states (Ireland, Estonia) that gained their

---

[167] Unfortunately we were not able to confirm if Bulgaria and Croatia have digitized Parliamentary Papers available online.

independence in the early twentieth century start their holdings in that period. Additionally, we have identified a digitized set of material on Yugoslavia. This shows that the focus is not solely on recent material but that a lot of effort is being undertaken to make historic Parliamentary Papers available to the public.

How is this data made available?

A key question for researchers in the realm of Digital Humanities is if material is available in full-text. If material is solely available as scans, its availability does of course help researchers but the source itself is approached in a traditional way. Independence from travelling or opening hours is gained, but the sources still need to be read page by page. If full-text is available, however, larger and different analyses can be made.

Not all countries have their Parliamentary Papers available in full-text. Of the researched countries Austria, Greece, Lithuania as well as older material from Belgium and Germany is only offered as scans. But this case is almost the exception because most of the other countries offer their material as full-text. PDF is by far the prevalent download option. This format is used by any of the researched countries that allow for downloading (which is not the case for Ireland and Lithuania). A handful of states also offer other formats - html, epub or txt files (France, Italy, Slovenia, Sweden). An notable exception are the Netherlands, since the whole data set is offered for download.

Diversity is also found in the way the material is approachable. Most parliaments link to their respective websites and a search and browse portal makes the data retrievable. But there are also examples found of countries offering elaborate portals that allow searching their historic material. Notably are the Joint Czech and Slovak Parliamentary Library[168] or Debates Parlamentares[169] from Portugal.

What does the diversity in availability of digitized Parliamentary Papers mean?

Our findings show that for most European member states at least scans are available. But this is only of limited use, both to the general public but particularly for researchers.

---

[168] The Joint Czech and Slovak Parliamentary Library, http://www.nrsr.sk/dl/?lang=en

[169] Debates Parlamentares, http://debates.parlamento.pt/?pid=mc

Most countries offer download as pdf, however not always as full-text pdf. This means that users and researchers alike can use the material but cannot search within the files. In short great variety is found both in term of availability and usability. This is where a centralized thematic portal could greatly support research in this area.

Licensing and Re-Use of the Data

In nearly all cases, there is little clear information about the license for re-use of the parliamentary material. Most resources do not explicitly refer to it. The website may have general copyright statement, but that applies to the website as a whole and one can assume that there has been no thinking applied to the copyright status of the historic material, much of which is likely to be out of copyright. One notable exception are the Netherlands who offer a digital dataset containing all material older than 1995 as a dataset licensed CC0 / PDM[170]. On the other side, the UK's parliamentary debates are licensed to a commercial publisher, ProQuest, who claims copyright over their House of Commons Parliamentary Papers site.

It is also worth to highlight the magnificent efforts of the Italian government in opening up its data. Its contemporary data sets for its Senate and Chamber of Deputies include downloadable datasets and SPARQL endpoints, all on data licensed as CC-BY.

Elsewhere, the lack of consideration for licensing also relates to the inability to access this data any other way but through the search and browse portals developed. For example, there is no way, on any of the sites surveyed, for a bulk download of metadata or full text, or any kind of API to permit programmatic access. Given the nature of the institutions (with plenty of other day to day concerns, and dealing with many other stakeholders before digital humanities developers and researchers) this is not surprising. However, until it is addressed a significant method for the better comprehension and analysis of Europe's shared political history is curtailed.

In conclusion we can say that a lot of effort has been made to digitize material, also historical papers, and offer them to the public. However these initiatives are very de-centralized. Europeana, or indeed other continent-wide actors, could fill the gap by offering centralized access to the data.

---

[170] https://www.kb.nl/bronnen-zoekwijzers/dataservices-en-apis/staten-generaal-digitaal

What other sources (digital or non) would you consult?

The scope of the work on this topic has a focus on parliamentary debates at a national level. It excludes an extraordinary amount of other documentation created by national parliaments (e.g. session papers, committee reports, petitions, details of politicians' lives) etc. It also has excluded regional parliaments - federal nations or those with some elements of decentralisation will have parliamentary documentation outside the capital.

Which other researchers' work would you monitor for this particular topic? In what ways?

There are several research projects working with these sources that can help display the potential of this topic.

Maybe the most elaborate is Digging into Linked Parliamentary Data (DiLiPaD)[171]. The project focuses on three countries: the United Kingdom (UK Hansard), Canada (Canadian Hansard) and The Netherlands (parliamentary proceedings of The Netherlands) and explicitly fosters comparative research. The project's research blog offers interesting insights in possible challenges but also opportunities of working with the mentioned data sets. The project also generated scientific articles (e.g. Richard Gartner, A metadata infrastructure for the analysis of parliamentary proceedings[172]) that might help guiding future initiatives.

DiLiPaD itself is based on an earlier project called LIPARM project that ended in 2013[173]. There a controlled vocabulary and XML schema (the so called Parliamentary Metadata Language - PML) was developed which is now used in the project mentioned above.

Another scientific work based on the digitized parliamentary proceedings is Hinke Piersma, Ismee Tames, Lars Buitinck, Johan van Doornik and Maarten Marx, War in Parliament: What a Digital Approach Can Add to the Study of Parliamentary History published in 2014[174]. The 'War

---

[171] Digging into Linked Parliamentary Data, http://dilipad.history.ac.uk/.

[172] Gartner, "A metadata infrastructure for the analysis of parliamentary proceedings," in Big Data (Big Data), 2014 IEEE International Conference, pp.47-50, 27-30 Oct. 2014.

[173] Linking Parliamentary Records through Metadata, http://www.liparm.ac.uk/?page_id=61.

[174] Piersma, Tames, Buitnick, Doornik, Marx, "War in Parliament: What a Digital Approach Can Add to the Study of Parliamentary History", in Digital Humanities Quarterly, Vol. 8, 2014.

in Parliament'[175] project, developed a search engine, which in turn was used to test if certain terms were associated with a right-wing party in the Dutch parliament. The paper shows two important elements: It investigates how the availability of full-text can take research further, by focussing on interdisciplinary approaches. But it also shows that digital scholarship is still text-based.

That Parliamentary Papers are of great interest both for the general public but also for researchers working on very specific topics is shown by the SPHERE project, which aimed to make the Stormont Parliamentary Hansards available online. The SPHERE team undertook the challenge to examine different measurements for the impact of their resource and their findings are quite interesting. The team concludes that "The resource is widely used for public engagement, is well known internationally, and usage is generally high." However, they also found that "an integrated corpus of Irish Parliamentary Papers would have far greater value than Hansard alone." The latter point is bolstered by interviews taken with researchers in that field. This highlights the potential of a resource or portal that brings together various types of Parliamentary Papers from a wide array of countries.

An highly interesting project both from its field of research but also from its administrative structure is the German project "Postdemokratie und Neoliberalismus." The project follows an explicit interdisciplinary approach in which researchers from the humanities, in this case Political Science, collaborate with researchers in the field of Linguistics. The goal is to investigate lines of argumentation within German debates. Although the project is mainly based on newspaper articles its methodology could also be applied to Parliamentary Papers.

Although there are only a few examples, these projects and their outcomes can give us few clues about the potential about this topic.

Firstly, digitizing Parliamentary Papers is only the first step. There are numerous initiatives by nearly every European Union member state works in this direction. But as Piersma et al. wrote: "Unfortunately, digitized sources are generally used in much the same way as the originals. The advantage of digitization is then limited to easy access (no more trips to the university library) and to finding what you are looking for (a particular debate or speaker) more quickly." Therefore

---

[175] WIP: War In Parliament, http://www.clarin.nl/node/410.

real innovation comes from taking it a step further. Having enriched full-text available for a growing number of states would enable researchers to pose new, transnational questions, that have the potential of yielding new results. Having such a resource available would as give strong incentive to collaborative, interdisciplinary research.

Which digital tools/services would you consider to be most useful when processing such data?

The projects cited in the section above give some sense of how the Digital Humanities might take the analysis of Parliamentary Papers beyond the search and browse portals offered by the existing parliaments.

Much of the focus of these projects (as with other Digital Humanities projects working with full text) is on refining the metadata and understanding the richness of the full text to allow for more fine-grained analysis. In the case of Digging into Linked Parliamentary Data , the richness comes from searching across multiple datasets.

In theory, the richness of the full text material could allow its reuse in myriad ways. Tools for linguistic or semantic annotation, topic modelling, analysis, visualisation etc could all be applied. However, this is entirely dependent on the full text not only being licensed for such use, but for it being available in the first place. Of those parliaments that had digitized historical Parliamentary Papers, many had produced the resultant images only as jpegs or pdf without automated transcription into full text. This fully restricts its potential for further research.

Are there any infrastructures (national and/or European) relating to this research area?

The European Information and Research Network on Parliamentary History (short name EUPARL.NET[176]) is the obvious candidate. This is a combined group that brings together many national interests. Initial contact was made with some members though this could be extended to pursue formal collaboration.

Summary of available, digitized Parliamentary Papers in Europe

---

[176] http://euparl.net/9353213/v/.

6.2 Population Displacement as a result of Conflict in the 20th Century

Population Displacement due to the exchange of Greek and Turkish Populations in the 1920s, the Revolution in Hungary in 1956, and the continuing displacement due to the Yugoslav Wars in the 1990s.

What is your research topic/area?

This desk research looks at migration, specifically population displacement due to conflict in 20th Century Europe. Vicky Garnett at TCD, and Norman Rodger at the University of Edinburgh addressed this topic, focussing mainly on conflicts that took place in Eastern Europe in the 20th century. Norman investigated the outcomes of the Greek/Turkish conflict in the 1920s, Vicky looked primarily at the Hungarian Revolution and the impact this had on population displacement, as well as the conflict in former Yugoslavia in the early 90s, often referred to as the Yugoslav Civil War, or Yugoslav Wars.

In both the Greek/Turkish conflict and the Hungarian revolution, small amounts of content exist within Europeana, although these have been difficult to find. Information about the Yugoslav Wars, however, is less readily available still, all for reasons we will discuss later in this document.

Background to the conflicts

Each of the conflicts has an extremely complicated background, however to provide some context, we provide a (very) brief summary of the main points within each conflict and the circumstances that led to population displacement.

Greek/Turkish conflict

At the end of the First World War and with the belief that they had the full backing of the former Allied Powers, Greece began an invasion of Turkey, with the eventual aim of achieving "The Big Idea," capturing of Constantinople and restoring it as a Christian capital. The Anatolian area on the west of Turkey had a large Greek speaking population and it was the Greeks aim to incorporate them within an expanded Greek state.

The armed conflict started with the Greek occupation of Smyrna, (contemporary Izmir) in May 1919, Greek forces then moved inland but their advance was halted checked at the Battle of Sakarya in 1921. The Greek front collapsed with the Turkish counter-attack in August 1922 and the defeated army began the long retreat west, along with a wave of Anatolian Greek and Armenian refugees, all pursued by Turkish forces. The war effectively ended with the re-capture of Smyrna by the Turks and the forced evacuation of troops and civilians as the city was razed to the ground.

The Treaty of Lausanne, signed in 1923 recognized the independence of the Republic of Turkey and its sovereignty over Asia Minor and Constantinople and led to the formal population exchange between Turkish communities living in Greece and vice versa.

Hungarian Revolution of 1956

On 23rd October 1956, a student demonstration in Budapest against the Soviet oppression of life in Hungary since the end of the Second World War gained huge public approval and was spontaneously joined by members of the general public. A popular saying joked that there were three types of classes in Hungary: those who had been in prison[177], those who were in prison, and those who were going to prison. People in Hungary were desperately poor following a failed economic plan imposed by the Communist Party. The students had 16 points (or demands) to see a peaceful transition of power from Soviet to Hungarian leadership. They tried to enter the radio building in Budapest in order to broadcast those demands. While a group of the students were inside the radio building, the AVH (secret police) began firing into the crowd outside, killing many. Sympathetic Hungarian soldiers donated arms to the protesters to enable them to fight back. For days, it looked as though the Soviet regime in Hungary would be defeated by the civilian street fighters. But while Soviet forces had left Budapest, they hadn't entirely left Hungary, and the newly appointed Hungarian Prime Minister Nagy threatened to take Hungary out of the Warsaw Pact if they didn't leave entirely. The Soviet forces didn't leave, and Nagy declared Hungary a neutral country, while simultaneously appealing to the West to help maintain their neutrality. Overseas aid, however, was slow to come. Many historians have blamed the distraction of the Suez Crisis for Britain and France for encouraging

---

[177] Taken from "Cry Hungary: A Revolution Remembered", a BBC documentary from 1986 (available on YouTube https://youtu.be/8FckhPmtE1A, 7mins 06 secs)

the Soviet Russians to head back into Hungary to crush the revolution on 4th November, whereas there are others who cite Nagy's declarations of intent to pull Hungary from the Warsaw Pact and make the country neutral as the reason behind Soviet Russia's forceful return to Budapest. 200,000 Hungarians fled the country to escape the violence and indiscriminate killings over the following weeks, most to France, Britain and the USA.

Yugoslav Wars

The Yugoslav wars were a series of conflicts within the confederated republics of Yugoslavia. The region had seen continued levels of conflict throughout the 20th century, and a growing disparity between the wealthier republics (Serbia, Croatia, and Slovenia) and the "unproductiveness of the South" saw resentment growing. In particular, the predominantly Albanian population in Serbian occupied Kosovo (which had been given autonomous status in a constitutional amendment in 1974, but not actually in practice) were increasing their claim for independence. By the early 1990s, with Yugoslavia facing insurmountable debt to the IMF, and a growing sense of nationalism within each of the republics (particularly Serbia), came to a head when Serbian leader Slobodan Milosevic used force against striking Albanian miners in Kosovo, he claimed in order to protect the Serbs living within the Kosovo region of Serbia (as it was at the time).

By 1995, multiple wars were waging in Yugoslavia. Bosnia had declared independence in 1992, but Bosnian-Serbs rejected the state of independence, and fought against it. This escalated quickly and over the next three years, starvation and gun-fights within Bosnia had killed many on both sides. The UN set up a peace centre just outside Srebrenica, and in July 1995, 25,000 Bosnian Muslims fled to the Dutch UN Peacekeepers for refuge. Overwhelmed, and unable to 'enforce' peace, the UN Peacekeepers allowed the Bosnian-Serb army to separate the refugees into women and children, and men and teenage boys. The result has since been described by The Hague as the genocide of Srebrenica, resulting in the death of over 7000 men and teenage boys. To date, however, Serbia does not officially recognise this as genocide, and 300,000 people remain displaced in 'temporary' accommodation, waiting to be returned to their home country in the region, or settled more permanently elsewhere. This therefore remains an incredibly sensitive topic for those in the region, and those who fled the country and sought refuge in the EU and further afield.

Can you give a brief description of the topic, including the chronological and geographic span?

This topic looks at population displacement due to conflict in Europe in the 20th Century. The 20th Century was a century of war as never seen before, with industrial weaponry and ideological conflicts across the world. Within Europe, this led to huge displacement of people, either fleeing as refugees or forced migration. This continued right up to the last decade of the century with the Balkans' War in former Yugoslavia.

While we could cover every conflict within this period of time, this study focuses on 3 major conflicts and the effects they had on populations in 3 key locations across Europe - the Greek-Turkish conflict of the 1920s; the Hungarian Revolution of 1956; and the Yugoslavian Wars of the 1990s. In each case, refugees left the conflict areas for regions in Europe or further afield that they considered safer. Many appealed to aid from the Western European countries, and the response to aid differed from conflict to conflict.

The conflict in Asia Minor in the early 1920s and the exchange of populations between Greece and Turkey formalised in the Treaty of Lausanne in 1923 led to extensive population displacement. Much of the Anatolian Greek population had already left or been forced out of Turkey before the treaty was signed but about half a million Greeks left Turkey about a million Turks left Greece as a result.

In the case of the Hungarian Revolution of 1956, Britain and France have often been criticised for responding with more conviction to the concurrent Suez Crisis, rather than assisting the Hungarians in their short-lived uprising against the Soviets. They were, however, open to refugees from Hungary, although there are those who believe this was more down to guilt at a lack of response to calls for aid during the return of Soviet troops in Hungary in November 1956[178].

Finally, the Yugoslav Wars were exceptionally complex, and still have repercussions for those involved today. 330,000 fled the region, and many of this number are still displaced 20 years

---

[178] Nemes, " The Welcome Refugees: Why the West opened its arms to fleeing Hungarians in 1956", Central European Review, Vol 1, No 19 (1999). Retrieved from http://www.ce-review.org/99/19/nemes19.html

later. In 2012, it was reported that the UNHCR had made settlement of 74,000 of the most vulnerable people from this displaced population one of its priorities[179]. Many refugees from former Yugoslavia are still living in 'temporary' shelters.

Can you list a few possible research questions emerging from this topic?

Research within this area has huge potential, mostly around social history, but also around social sciences and political history. The issue of population displacement due to conflict is heightened within the news due to the current refugee crisis in southern Europe. Free movement within the EU is praised by many, but challenged by some, and migration of refugees can often fuel this discontent.

Using materials currently existing in Europeana, focus can easily be applied to the patterns of migration, the causes of migration, the ease (or indeed difficulty) of migration, how the migrants were received in their destinations, and how did this depend on the sympathies of the destination countries with their plight? The involvement of the surrounding countries and how they may or may not have been able to help? In the case of the Hungarian Revolution, could UK assistance have helped the situation, for example? How have attitudes changed towards migrants from Eastern Europe within the EU since the Cold War ended? Looking at the Greek/Turkish conflict, This might help to answer questions on the perspectives/interests of foreign powers, the attitude of the local communities pre and post exchange and the impact on Greece and Turkey in the immediate and longer term.

Taking a broader view, how have attitudes towards conflict changed since the 20thC, and in particular how have the 20thC European approaches to conflict and resulting migration impacted on Europe's approach to the conflicts in the Middle East in the 21st century (Iraq and Afghanistan in the 00s compared with Syria in this decade).

What are the academic disciplines corresponding to this topic?

"Those who cannot remember the past are condemned to repeat it."

---

[179] Taken from "Suffering goes on for 330,000 refugees of the Yugoslav war", an article in the Independent newspaper, 6th April 2012, http://www.independent.co.uk/news/world/europe/suffering-goes-on-for-330000-refugees-of-the-yugoslav-war-7622108.html (Accessed: 6th October 2015).

George Santayana[180]

Scholars within the field of history will most likely benefit from a collection around this topic, in particular social and political history. However, given the current situation in Europe, it is highly likely that this collection could have huge potential for journalists, social scientists, sociologists and those working within peace/conflict resolution fields.

From a wider perspective, anyone within the literary or performing arts may find some inspiration from a collection focussing on this topic. An understanding of the personal stories from migrants who were directly impacted by these conflicts may provide an invaluable insight that may enhance the item that they are creating, be it a work of fiction or poetry based around these events, or for a play or dance piece.

Predominantly historical research but also on the social and cultural changes that arose from these events, e.g. the emergence of Rembetika music in Greece.

Can you identify any Europeana datasets which could support research on the topic?

Greek/Turkish Conflict

There is not a huge amount of material to be found in Europeana. Europeana Newspapers would appear to be a possible source but the limitations of its search functionality make it difficult to use without knowledge of several languages.

That said, the following collections may be of use:

The Question of Population Exchange Between Turkey and Greece in the Lausanne Treaty

http://www.europeana.eu/portal/record/9200418/BibliographicResource_3000125958743.html?start=1&query=what%3Agreece+turkey+population+exchange&startPage=1&qt=false&rows=24

The Place and Importance of Friendship Treaty in 1925 in Establishing of Political Relations Between Republic of Turkey and Bulgaria

http://www.europeana.eu/portal/record/9200418/BibliographicResource_3000125963956.html?start=4&query=treaty+of+lausanne&startPage=1&qt=false&rows=24

---

[180] Santayana, "The life of reason. Vol. 1, 1905", *Fixed-versus Random-Effects Models* (1998).

Europeana Newspapers – La Siecle, 16 November 1922
http://www.theeuropeanlibrary.org/tel4/newspapers/issue/3000113941465?hp=1&page=1&query=asia+minori+catastrophe+1922

Europeana Newspapers – La Figaro 30 July 1939
http://www.theeuropeanlibrary.org/tel4/newspapers/issue/3000113894186?hp=3&page=3&query=izmir+1922

Europeana Newspapers – La Siecle, 29 September 1922
http://www.theeuropeanlibrary.org/tel4/newspapers/issue/3000113941500?hp=1&page=1&query=venizelos+1922

Hungarian Revolution

The Radio Free Europe[181] transcripts offer first-hand accounts of those fleeing Hungary (and other Soviet states), which were then broadcast back into the USSR as propaganda from the US during the Cold War. Many of the typed documents reveal the propaganda that the USSR was also using (although of course as this was propaganda from the West, it might also be taken with a pinch of salt!). A good deal of the documents, however, are in the original Hungarian, and therefore difficult to read by someone who doesn't speak the language.

Many of these transcripts are part of the HOPE (Heritage Of People's Europe) project's collection within Europeana, a large collection of items relating to social history.

EuroPhoto[182] also offers photos from across Hungary (and the rest of Europe). The project aims to digitise historical pictures from the archives of 10 leading European News Agencies and is funded by a European Commission ICT grant[183]. Images also exist within Europeana from the National Library of Wales which show people preparing care packages for refugees, or making space in their homes.

Less specifically, making use of the Europeana API brings up several results, although these are perhaps less easy to filter down for research purposes.

---

[181] Radio Free Europe transcripts available in Europeana Collection 2022043.

[182] EuroPhoto available by searching for "europeana_collectionName:2023710*".

[183] Taken from http://site.project.europhoto.org/project/index.php (Accessed: 24th August 2015).

Yugoslavian Wars

There is little in Europeana around the wars in Former Yugoslavia. This is perhaps unsurprising for ethical reasons and institutional ones. As discussed, there are many who many people who may have given accounts who are still alive, and therefore their data is subject to embargos for ethical reasons. There are also inherent difficulties in obtaining testimonies or artifacts from the time as people may well want to forget the conflict, to move on. There are also those for whom it is still too painful a memory. And finally, there are institutional barriers, such as Official Secrecy Acts which may also hold an embargo over the releasing of information around this conflict. This war only finished 20 years ago, and therefore there may still be a long term yet to complete before official documents around this can be released.

Can you identify any other datasets available on the topic?

Hungarian Revolution

Hungarian Archival Portal (http://mlp.archivportal.hu/id-2035-introduction.html) (HAP)

Fortepan photographic archive hosts a number of images related to the Hungarian Revolution. The collection is of amateur photographs that have been submitted through crowdsourcing, so the metadata associated with each image is reliant on the information provided by the person submitting them. This could be a benefit as it might reveal interesting back stories to each image. On the other hand, information might not be provided for personal reasons.

www.rev.hu was a specific archive relating to stories from the 1956 revolution as part of the National Library Archives. Unfortunately this is now defunct and has been absorbed by the National Library Archives.

Columbia University Libraries Archival Collections -

Columbia Center for Oral Histories portal

https://clio.columbia.edu/catalog/4077605

(686 items)

Reminiscences of Lajos and Kati Piros: oral history, 1979

https://clio.columbia.edu/catalog/4074504

(114 pages)

Oral history from couple who were Hungarian Refugees

Greek/Turkish conflict 1920s

There are currently no datasets that can be identified regarding the Greek/Turkish conflict.

Do you have any indication of who is responsible for these datasets? How are they licensed? Would it be easy to aggregate this content into Europeana?

HAP (from their website: "Creation of the Hungarian archives portal was sponsored by the Ministry of Education and Culture from the budget for the Renaissance year of 2008. At present, the portal includes information related to local governmental archives, and to a lesser extent the Hungarian National Archives, but the editorial staff wishes to involve open private archives, professional archives and the archives of higher educational institutions in the circle of participating institutions in the future.") Therefore it's not clear if these items can be licensed easily.

The Fortepan[184] photo archives are part of the larger Open Society Archives from Hungary[185]. Items from the archive are available under CC-BY-SA-3.0. Open Society Archives (OSA) is already an aggregator for Europeana, however the Fortepan collection is not currently available through Europeana. This is perhaps because the collection's contents are not CC0, the current requirement of content going into Europeana.

Data from the Columbia University Center of Oral Histories has the following access information:

"Access: Open.

Copyright by The Trustees of Columbia University in the City of New York, 1984. Permission required to cite, quote, and reproduce. Contact repository for information."

---

[184] http://fortepan.hu/ (Accessed: 25th August 2015).

[185] http://www.osaarchivum.org/ (Accessed: 25th August 2015).

In some cases, the subject in the oral history interview is still alive, which can have an impact on the availability of the content. Currently content is only available through a username and password, held by members of the Columbia University Library.

What other sources (digital or non) would you consult?

Name of item	Provider / Publisher	Source type	Author	Access
Greek/ Turkish Conflict				
Parliamentary Papers				http://parlipapers.chadwyck.co.uk/fulltext/readview.do?area=hcpp&id=1923-026125&source=&pagenum=1&resultNum=241&entries=4395&queryId=..%2Fsession%2F1440499335_9947&backto=FULLREC&fromSearchHistory=&searchWithin=&showall=&enlarge=Y&gotoFirstHit=&SEARCHSTRING=
FORCED MIGRATION IN GREECE AND TURKEY IN THE 1920s	Migration Citizenship Education	Web archive		http://migrationeducation.de/14.0.html
The Exchange of Populations: Greece and Turkey	Pontos World	Website		
The Greek Turkish Population Exchange	Arizona	Powerpoint slides		http://cmes.arizona.edu/sites/cmes.arizona.edu/files/3.%20Case%20study%20-%20Greek%20Turkish%20Population%20Exchange.pdf
Smyrna 1992	Vimeo			https://vimeo.com/10069165
The Gennadius Library - Videocast. The 1923 Exchange of Populations: An Ongoing Debate	YouTube			http://www.ascsa.edu.gr/index.php/gennadius/newsDetails/george-mavrogordatos-and-ayhan-t.-aktar-the-1923-exchange-of-populations.-a
Population Exchange Between Greece And Turkey	YouTube			https://www.youtube.com/watch?v=AJqv1_XsPJI

– YouTube				
Greek invasion of Turkey and the aftermath of the events, 1919-1922	YouTube			https://www.youtube.com/watch?v=oRDBVmTlPaA
SMYRNA: THE DESTRUCTION OF A COSMOPOLITAN CITY, 1900-1922	YouTube			https://www.youtube.com/watch?v=c9XkFCBZtOs
				https://uk.pinterest.com/search/pins/?q=smyrna%201922
				https://uk.pinterest.com/greekgenocide/the-refugee-experience/
				http://levantineheritage.com/bornabat.htm
				https://uk.pinterest.com/greekgenocide/smyrna-izmir/
Hungarian Revolution 1956				
"Laying claim to Hungary's 1956 revolution"	BBC Online	Webpage	Victor Sebestyen	http://news.bbc.co.uk/2/hi/programmes/from_our_own_correspondent/4651808.stm
On This Day 23rd	BBC News	Webpage	anon	http://news.bbc.co.uk/onthisday/hi/dates/stories/octob

				er/23/newsid_3140000/3140400.stm
On this Day 26th October 1956	BBC News	Webpage	anon	http://news.bbc.co.uk/onthisday/hi/dates/stories/october/26/newsid_3200000/3200703.stm
On This Day 4th Nov 1956	BBC News	Webpage	anon	http://news.bbc.co.uk/onthisday/hi/dates/stories/november/4/newsid_2739000/2739039.stm
Special section on the BBC website was dedicated to the 50th anniversary of the the 1956 revolution.	News coverage from the BBC Online,	Webpage	anon	http://news.bbc.co.uk/2/hi/europe/6075886.stm
Yugoslav Wars				
Migration Citizenship Education website		Website		http://migrationeducation.de/16.0.html

Which other researchers' work would you monitor for this particular topic? In what ways?

Greek/Turkish Conflict

Dr. Antonis Klapsis is an Adjunct Lecturer at the Department of Political Science and International Relations, University of Peloponnese.

Hungarian Revolution

Dr. Balazs Apor in Trinity College Dublin has studied the Hungarian Revolution of 1956, and the '1956ers' who fled the country. While his interests have shifted now to propaganda, he did still have some interesting notes for the topic of population displacement. His research is mainly focussed on physical items within archives, in particular the archives in Budapest. The materials he works with are typically not digital.

Dr. Vera Sheridan at Dublin City University has done considerable work looking at refugees coming to Ireland. Her work has covered the aftermath of the Hungarian Revolution and the '1956ers' that came to Ireland[186].

---

[186] See references section for a brief bibliography of Dr. Sheridan's work.

Yugoslav Wars

Nela Milic is a Digital Artist who worked with the Kulturklammer founders to set about retaining oral histories of the recent Balkans War and the impact it had on the local community in Belgrade both during and after the conflict[187]. When asked if her oral histories revealed anything about those who had moved away from the area as a result of the conflict, she said that there was no such archive as the region had little money to maintain such records.

Which digital tools/services would you consider to be most useful when processing such data?

Accessing data from many of these resources could be most efficiently done via an API - however this would be considered quite advanced technical knowledge and might not be so common among the typical Humanist research community.

Pinterest would be useful if there were more powerful features. Story Map Journal has also been useful. Culture Collage has potential with this also, although it is not such a powerful tool at this stage.

Talking to Dr. Balazs Apor, he hasn't necessarily used any digital tools, but he commented that an image annotation tool such as HyperImage (for example) would be very useful in order to create links between images and keep his notes alongside the image he is analysing.

Dr. Apor also commented that some manner of Zotero-like tool would also be useful for cataloguing the content found and creating links between the content. If this could then be downloaded as an index of some sort that would enable him to find content and references more easily. By extension, Dr. Apor commented that in his research he generally didn't use any sort of reference software, and typed in the references by hand. However, there are many who would find this sort of software very handy indeed, particularly if they are writing a longer piece, such as a thesis, chapter or even an entire book. Being able to use reference software alongside Europeana content with all annotations made on the Europeana Research Platform also included would be highly beneficial to researchers in all sectors, not just in social history.

---

[187] http://www.kulturklammer.org/view/64

Are there any infrastructures (national and/or European) relating to this research area?

In the case of the Yugoslav Wars, this is still a very sensitive subject, as many of those living or even who participated in the conflict (on one side or the other) are alive. There is still controversy over even vocabulary, let alone the outcomes of the conflict. Only a small fraction of the bodies from mass graves all over former Yugoslavia have been identified, so this remains a raw and recent conflict for the region. Funding for such infrastructures is not a priority within the region which is still recovering 20 years later.

This sensitivity to local opinion is also a hindrance in the older conflicts studied. One of the problems within the Greek/Turkish Conflict subject is that there is an element of denial on the part of both Greece and Turkey over the events leading up to the exchange and its immediate aftermath, as well as within those countries with a vested interest at the time, notably Great Britain, France and the USA.

6.3 Political Propaganda

Political Propaganda: Power and Communication

In many cases, expressions of political propaganda provide exceptionally vivid examples of organized efforts to mobilize, stir, steer or influence public opinion on a wide variety of policies and debates. Aggressive or more benign, confrontational or more persuasive, overt or more manipulative, ever since the Classical Era propaganda has been part and parcel of both official and subaltern political communication[188]. Virtually every state, regime, government, political party or pressure group has used available media (art, film, radio, literature, theatre, music, posters, and more) as propaganda. In the late 19th and early 20th centuries, posters are often seen as "the quintessential form of propaganda," because they provided a cheap way to reach a wide audience, could be understood even by illiterate citizens, and could be changed quickly in response to political shifts.

Yet, questions on how to define and delineate the general notion of "political propaganda" abound. Whereas few would hesitate to label political campaigns of regimes such as Pol Pot's

---

[188] For recent studies of propaganda in Classical Rome and Greece, see Enenkel and Pfeijfer (Eds.), *The Manipulative Mode: Political Propaganda In Antiquity. A Collection Of Case Studies* (2005).

Khmer Rouge in Cambodia in that manner, many other officially sanctioned campaigns fall less clearly in this category. What are we to make of public health campaigns advocating the consumption of dairy products that help sustain organized agricultural lobbying organizations in Brussels, working to support a complex political-economic structure allocating the majority of European subsidies to industrialized farming? Or, to name only one other example, to what extent are Amnesty International campaigns to denounce political systems that routinely persecute, imprison and torture political opponents free of propagandistic calls for interventions in a (foreign) political culture or system?

In inventorying this theme as a potential organizational and scholarly instrument for future efforts by Europeana (Research), the WP1-team deployed a two-pronged approach. First, they conducted desk research in order to compile an overview of relevant (digitized or partially digitized) collections of library, archival and, to some extent, museum materials that could be classified as relevant to the broader theme. Second, it attempted to query a small set of specialized researchers, asking for their contributions to the inventory, their use of research tools, and their assistance in creating relevant subsections from the larger collections identified by the team. For these purposes, a 'Topic questionnaire' was sent out to 10 selected researchers.

Preliminary inventory

The team's work primarily yielded larger (national) collections of political propaganda materials. The inventory is preliminary by nature, and mostly restricted to more conventional expressions of political propaganda. Thus, it virtually excludes subjects and materials that cannot be classified as easily as pertaining to (national) propagandistic manifestations. By means of example, it should be noted that numerous researchers study artistic expressions and representations, from classical to contemporary art, for their interactions or entanglements with, support for or resistance against official propaganda[189]. Any elaborations on or additions to this inventory will by necessity have to be structured and driven by conceptual discussions of the central but complex issue: what is "political", and what is not?

---

[189] For a recent curatorial example, an art history project in Romania, see "From Political Propaganda to Baby Boom," analysing the relationship of Romanian artists to different political ideologies and social climates during the interwar 1918-1940, post-war and communist 1946-1989 periods.

Given its more conventional orientation, the following preliminary inventory ranges from wars to elections, and from (national) pamphlet collections to materials from various strands of the workers' emancipation movements. And whereas content modules can easily be imagined regarding earlier historical eras (i.e., the Dutch pamphlet collection on the nation's genesis and early development[190]), the collections listed here relate primarily late 19th- and 20th-century materials.

A natural and profitable starting point is to search the Europeana portal itself. A query for Spanish Civil War (1936-1939), and this term translated into six languages, yielded 16.158 results (as of Sept. 9, 2015)[191]. This might very well form the basic foundation for clustering and identifying such interrelated materials.

Listed by data provider, this translates into the following 20 repositories with 20 relevant items or more:

- National Library of Spain (6544)
- EuroPhoto EFE provider (4516)
- Archivo de la Imagen de Castilla La Mancha (2188)
- NIOD (593)
- Amsab-Institute of Social History (575)
- Fototeca de la Diputación de Huesca (297)
- Arbejdermuseet & ABA (292)
- TopFoto.co.uk (225)
- National Library of the Netherlands - Koninklijke Bibliotheek (156)
- Archiv der sozialen Demokratie (AdsD) (138)
- Complutense University Library of Madrid (113)
- CER.ES: Red Digital de Colecciones de museos de España (71)
- Österreichische Nationalbibliothek - Austrian National Library (47)

---

[190] '17th-century Dutch Pamphlets', part of the Knuttel Collection (1486-1853) at the National Library of the Netherlands. TEL-record: http://www.theeuropeanlibrary.org/tel4/collection/a0028 (access conditions have not been considered in the current inventory).

[191] For instance:

http://www.europeana.eu/portal/search.html?query=spaanse+burgeroorlog&rows=96&qf=-UGC%3Atrue

- Gredos (Universidad de Salamanca, Spain) (45)
- Nederlands Instituut voor Militaire Historie (37)
- TVC (28)
- Archivo de la Ciudad de Arganda del Rey (25)
- Deutsche Fotothek (21)
- Nationaal Archief (21)
- Hedatuz (20)

On the First (or Great) War, iIn addition, the query yielded another 62 repositories with anything up to 19 relevant items. Having available thematic sub-collections, such as this one on the Spanish Civil War, provides the basic requirement for designing similar queries. The results could then be labelled with an additional descriptor in the Europeana metadata schemes.

A similar search, but in the module Europeana Collections 1914-1918, applying the considerably more generic term propaganda, naturally produces a larger and more inclusive 'collection', but again yields fertile grounds for organizing and labelling relevant (sub)sections for the broader theme[192]. One finds, for instance, a lesser-known collection of 28 pamphlets written in Antwerp by Jozef Buerbaum in the years 1915-1918. Using the pen name Janus Droogstoppel, Buerbaum's epistles are powerful expressions of rejection of the German occupation of his country – and the rhetoric used to justify it.

Returning to the subject of the Spanish Civil War, and to the broader subject of 20th-century war propaganda in Europe, it becomes obvious that a number of researchers have recently focussed their attention to the internecine warfare on the Iberian peninsula. Some of them have assembled significant collections on the subject, developed catalogues and/or have published low-resolution images of relevant materials on the internet. For example, Alexander Vergara presents his poster collection from the conflict on a website entitled 'The Visual Front'. His 96 posters, all from the Republican (Left) side in the conflict, constitute an important corpus, as only 1,500-2,000 posters were produced in total (in editions of 3-5,000 copies). His catalogue is based on a collection in the University of California San Diego's Southworth Collection. He

---

[192] http://www.europeana1914-
1918.eu/en/collection/explore/keywords/Propaganda?qf%5Bindex%5D%5B%5D=a

provides excellent contextual and scholarly information[193]. It should be noted that some of the items are already included in Europeana, thus providing an excellent steppingstone for further content acquisition. In a somewhat similar fashion, Rod Oakland has collected and described a range of leaflets, disseminated by both sides in the Spanish Civil War. Relatively little has been written about them in English, and many researchers are unaware of the existence and impact of these items that were usually airdropped[194]. A collection from a similar vein that could be approached with sufficient editorial / scholarly guidance, presents a surprising number of Republican posters, and a smaller set of Nationalist ones[195].

In addition to the materials already listed in Europeana (Europeana Collections 1914-1918), it should be noted that an important collection on the subject that will become available in the (near) future is stewarded by the German National Library. The Deutsche Nationalbibliothek's Plakatsammlung consists of more than 14,000 posters and placards and is based on two collections that were started in 1914 and 1939. At least 14,000 are directly related to the First World War; an additional 8,000 relate specifically to the Leftist November Revolution of 1918. A digitization programme is being developed since many of the items are too fragile to be made available for researchers – or the wider public[196]. Similarly, the National Library of Portugal, World War I collection, is to be added to Europeana Collections 1914-1918 in 2018 (see http://purl.pt/index/IGuerraMundial/PT/index.html).

One of the prime collections of (traditional) expressions of political propaganda deriving to large-scale conflicts in Europe is at the Wolfsonian Institute at the Florida International University, Miami Beach. It also contains material on US involvement in wars, like the Spanish-America War and WWI and WWII. It holds an impressive collection of war propaganda from almost all the combatant countries in both World Wars along with a number of smaller wars. Items in this category range widely from posters, postcards, sheet music, and medallions to paintings and sculpture.

---

[193] http://libraries.ucsd.edu/speccoll/visfront/intro.html

[194] http://www.psywar.org/spanishcivilwar.php

[195] http://digitalpostercollection.com/1936-1939-spanish-civil-war/

[196]

http://www.dnb.de/DE/Wir/Sondersammlungen/plakatsammlung/plakatsammlung_node.html;jsessionid=5
553B7BB62B1FEB668B90CFE426238C8.prod-worker2

The collection also contains holdings related to the far-reaching propaganda campaigns waged during the interwar period, especially by the totalitarian governments of Italy, Germany, and the Soviet Union. Vivid representations of the personality cults constructed around the dictators of these countries are also prevalent[197].

The output of propaganda materials attained far greater volumes during the Second World War and the Cold War. Several broad collections are listed here.

One of the main Dutch national portals for cultural heritage, Memory of the Netherlands, contains an assemblage of German war posters (or from political organizations affiliated with the Nazi regime). Totalling almost 5,000 pieces, it has been integrated from the holdings of the National Library of the Netherlands and the NIOD-KNAW institute for the study of the War, Holocaust and Genocide. Subjects include: German propaganda campaigns, calls for the Waffen SS, instructions for protecting the population against air attacks, leisure activities and German death sentences[198].

From the same portal, pamphlets and brochures against the German occupation in Netherlands. This collection of the Memory of the Netherlands includes some 1,000 pamphlets (about 1,750 pages) and 300 illegal brochures (about 5,000 pages). Most of these pamphlets were produced and circulated by the same underground groups that made and distributed illegal newspapers, e.g. Trouw (Faith), Vrij Nederland (Free Netherlands), Het Parool (The Motto) and De Waarheid (The Truth). Some of these pamphlets are special bulletins issued by these underground papers, but the contents are mostly unrelated. Examples of topics covered by the illegal pamphlets are: forced labour in Germany; the persecution of the Jews; the Royal Family and the black market. Like the pamphlets, the brochures were used to make a strong stand against the German occupation. Ultimately, the brochures had one common goal: to undermine the occupier's power and influence. The more than 6,000 items derive from the same repositories as listed above and, in addition, University of Leiden's Library[199].

---

[197] http://www.wolfsonian.org/explore/collections/collection-themes/war-and-political-propaganda

[198] http://www.geheugenvannederland.nl/nl/collecties/oorlogsaffiches_1940-1945

[199] http://www.geheugenvannederland.nl/?/nl/collecties/verzetsliteratuur

The German National Library's poster and placard collection (see above) contains ca. 5,000 leaflets and proclamations from the Second World War.

Considering the propaganda materials that accompanied the Cold War and the colonial projects of European countries and the USA, the WP1-team had insufficient time at its disposition to assemble a reasoned overview. It is expected that initial consultation with researchers will produce adequate listings of holdings and repositories[200]. (See also: Discussion/recommendations, below.)

More generic collections

For election-related political pamphlets and party political writings, some interesting pointers can be provided.

National Library of Denmark, party programmes and party political writings from Denmark

This collection of digital items brings together party programmes and party political writings from 1872 until 2001. It consists of Danish political parties' programmes, election pamphlets, brochures and similar information material which the library has digitised and made available online. Altogether the collection contains 1,600 items[201].

From Gallica, to be selected: pamphlets politique

Gallica is a very large and diverse collection, which also contains political pamphlets. These should be culled from the general collection, and subsequently classified. Unfortunately, searching exactly for the phrase 'pamphlet politique' is not possible, it generates noise[202].

Kulturmanufaktur and Östereichische Nationalbibliothek, Posters relating to parliamentary elections between 1919 and 1930.

---

[200] The Wolfsonian Institute's collection 'Colonialism' contains many official representations endorsing the colonial European (and US and Japan) projects, see
http://www.wolfsonian.org/explore/collections/collection-themes/colonialism.
[201]

https://www.google.com/url?q=http://www.theeuropeanlibrary.org/tel4/collection/a0434&sa=D&usg=ALhdy29QUiv3Ol7GBm5CyGSNwaYc0kfZKQ
[202] http://gallica.bnf.fr/?lang=EN

Over 500 posters relating to parliamentary elections held in 1919, 1920, 1923, 1927 and 1930. Accumulated from all Viennese poster collections[203].

British Library, Northern Ireland Political Pamphlets

The Northern Ireland Political Pamphlets Collection comprises approximately 1,000 items published by a range of political and cultural organisations based in Northern Ireland from the middle of the 1960s to the beginning of the 21st century. The collection complements the books and periodicals received by the British Library through Legal Deposit. The collection was acquired from the Linen Hall Library, Belfast, which has retained a duplicate of each item[204].

A separate category?

The earlier mentioned Dutch cultural heritage portal, Memory of the Netherlands, has a search possibility that leads to the identification of a somewhat separate category of political propaganda. Its collection of posters amounts to 52,000 items (1870-now), and the portal offers an option to limit one's search results[205]. Selecting the posters of the International Institute for Social History (IISH, in Amsterdam) still yields 22,000 posters, but adding the selectors "politics" and a time period allows for results to be limited to posters generated by specific Leftist movements that advocated the emancipation and empowerment of the workers in European factories and offices during the late 19th and 20th centuries.

The IISH stewards one of Europe's most extensive, international collections of publications, organizations' archives, leaders' personal archives and political writings from the Left / progressive political movements, ranging from radical to social-democratic movements and parties. The IISG is part of a closely intertwined network of similar sister institutes throughout Europe. Their cumulative coverage of the workers movement ideologies, organizations, parties is impressive – and, as the example of the posters on the Dutch portal indicates, much could be

---

[203] http://www.oeaw.ac.at/cmc/kds/index.php

[204]

http://www.bl.uk/reshelp/findhelpregion/europe/northernireland/northernirelandpoliticalpamphlets/nirepolpam.html

[205] http://www.geheugenvannederland.nl/?/nl/collecties/affiches. The website states that "all materials are owned by public institutions."

gained from connecting the metadata, images and texts from the IISH-holdings and that of its sister institutes.

A first module in this direction could be developed relatively easily, as the IISH has recently digitized its 100th collection. Tis could serve as a 'poster'-project to introduce and develop the idea of an integrated, European portal on the workers movements' histories in Europeana (Research). Some of the IISH's sister institutes, by country, are:

- *Germany:*
  Archiv der sozialen Demokratie und Bibliothek der Friedrich-Ebert-Stiftung

  www.fes.de

  Friedrich-Ebert-Stiftung (FES), Bonn

  Contact: Dr. Anja Kruke (Archiv)

- *Italy:*
  Fondazione Giangiacomo Feltrinelli

  www.fondazionefeltrinelli.it

    o Milano
      Contact: David Bidussa

      Fondazione Antonio Gramsci

      www.fondazionegramsci.org

    o Roma
      Contact: Dr Dario Massim

- *Spain*
  Fundación F. Largo Caballero. Archivos del Movimiento Obrero (FFLC)

  www.ugt.es/fflc/

  Alcalá de Henares (Madrid)

  Contact: Nuria Franco

- *Switzerland*
  Schweizerisches Sozialarchiv

  Stadelhoferstrasse 12

CH-8001 Zürich
Contact: Urs Kälin

Stiftung Studienbibliothek zur Geschichte der Arbeiterbewegung

Quellenstrasse 25

CH-8031 Zürich

Contact: Brigitte Walz-Richter

Topic questionnaire

In an attempt to gain assistance in refining the above list of collections and in selecting specific subsections from them, a questionnaire was sent out to 10 researchers in Europe (see the Annex, below). Additional aims of the consultation were to query the researchers on their (digital) research practices, their knowledge of Europeana, and to solicit input for identifying materials that the team had not located by themselves. Time restrictions did not permit the team to cast its nets wider; and only 2 researchers participated.

Both respondents reported that they limited their digital research materials to 'images' and 'text' and stated that they had not performed any data analysis or visualisation. But they diverged considerably in their professional stage (PhD candidate / assistant professor), their digital literacy, and their capacity to identify additional content (-modules) for the topic. Compared to the listing above, the single addition that could be extracted from the information provided was to select colonial propagandistic materials from the collections of the National Archives of the Republic of Indonesia (ANRI). Whether this could possibly within the scope of Europeana's ambitions and operations remains a question.

Discussion / recommendations

In order to establish the validity of "Political Propaganda" as a developing module for content aggregation for Europeana, a sustained and substantially more appreciable involvement of researchers is essential. Time and effort will have to be allocated to develop mechanisms by which specialized researchers are sufficiently triggered to submit suggestions (on content, tools and functionality) from their specific subjects and methodological niches (that are, by their nature, very fragmented) to the broader, overarching theme and subject.

A similar challenge presents itself with regard to the management of the thematic module on the subject: editorial involvement of specialists is required to monitor progress and establish priorities in the content acquisition phase. A sizeable number of content suggestions listed in the previous pages describe subsections from larger collections of material, but the process of actually culling relevant materials from the larger collections, their listing, classification, acquisition management and deployment as Europeana materials within the thematic collection is one that requires deployment of expertise from subject specialists.

Finally, during the course of the development of this theme by WP1-collaborators it has proven unattainable to actively assess licensing issues of the materials that could potentially be incorporated into the thematic collection. It should also be noted that some of the materials are actually located in US repositories, and the WP1-team has not performed any research into appraising copyright/fair use stipulations and their possible divergences in the US and European domains.

6.4 Children's Literature

Point of departure for defining children's literature as one of the thematic topics for Europeana Research was the successful DARIAH web survey conducted in 2014 on digital practices in the Arts and Humanities. In the question "Identify your more specific field(s) or area(s) of research", out of 1833 replies, "literature" was a commonly met term. This popularity of the term "literature" led to the first search in the portal of Europeana resulting in 104 163 only text-based material found under this term. Setting literature as the wider thematic topic, next step was to narrow it down to a more specific area of literature, thematically or even chronologically. Children's literature thus emerged as a narrower sample of literature leading to a second search in the portal of Europeana resulting this time in 1596 only text-based material; nevertheless, these results were multilingual and spanning in a period of 500 years. While this number is quite low, it thus showcases that there is room for further research and exploration for enforcing the presence of this field in the new portal of Europeana Research.
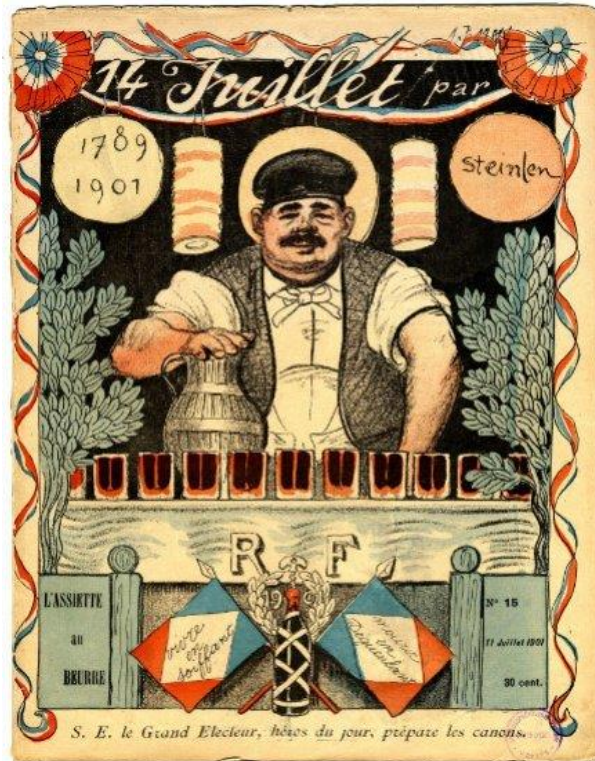
Title: Obálka časopisu L´Assiette au Beurre, č. 15
Creator: Steinlen, Théophile Alexandre
Data provider: www.esbirky.cz
Provider: Partage Plus
View item at: www.esbirky.cz
CCO

The methodology followed in researching this topic was a mixed methods approach. Exploring Children's Literature as a research field, inside and outside Europeana, required at first an extensive desk research to record its current state of the art. In this way, the aim was to understand the particularities and characteristics of the field, the extent of its digital, mainly, presence in Europeana and in other libraries and archives around the world, the tools and services build so far on top of such content and its recent orientation as an academic field. This would also allow delimiting the topic in chronological and geographical range for the purposes of the project. Apart from that, desk research was further complemented by virtual meetings held with researchers of the field. Semi-structured interviews were designed and conducted via Skype while in some cases the designed questionnaire was answered offline by the researchers, if their schedule did not allow a virtual real-time conversation. Having questions on

research career and background, research practices, resources and digital tools and finally on future developments needed or expected, researchers were asked to share their own expertise and experience of the field and also their vision for its ideal future development. The results of this methodology will be further analyzed below and will ultimately lead to findings and suggestions for the platform of Europeana Research.

Definition of the topic

"As a term, "children's literature" does not easily fit into any cultural or academic category; rather, it is a diverse and paradoxical area of study."[206] This suggestion mirrors the large number of definitions and theories found in the literature around this term as well as the challenge of setting specific chronological and geographical boundaries in this current research. According to the International Companion Encyclopedia of Children's Literature it "*is (among many other things) a body of texts (in the widest senses of that word), an academic discipline, an educational and social tool, an international business and a cultural phenomenon*". Even its definition is quite controversial and meets various explanations. Karin Lesnik-Olberstein in 1999 suggested that children's literature is a "category of books the existence of which absolutely depends on supported relationships with a particular reading audience: children".[207] If we adopt this suggestion then "children's literature is defined by audience in a way other literature tends not to be".[208] Apart from the various definitions that this topic accepts, it also seems to embody a number of different content types, genres and interested disciplines. For the purposes of this project, children's literature will be approached as a research theme rather than as type of content concerning thus academic audience.

In an effort to capture its historical dimension, we render here Peter Hunt's suggestion:

"In the early stages of a printed literature, there are few or no books published specifically for children. There are perhaps a few books intended for broadly educational purposes, such as the courtesy or behaviour books printed in the fifteenth or sixteenth centuries in European countries, or the twentieth-century text books published to support the formal school curriculum in

---

[206] "To Instruct and Delight: A History of Children's Literature" in Random History.com, published in January 2008 http://www.randomhistory.com/1-50/024children.html (Accessed" 13 October 2015).
[207] Hunt (Ed.), *Understanding children's literature*, Routledge, 2006.
[208] Ibid.

developing countries. In this situation, children, as they learn to ready, also take over adult books which appeal to them." [209]

It is evident that attempting to specify the topic chronologically and geographically is quite problematic. Projects in this area, or even individual research initiatives, do not usually approach this kind of literature by countries of origin or by chronological period. Even if they do, this is quite restrictive for the aims of the present task. For this reason, research conducted here was quite inclusive focusing mainly on European literature of the 18th to 20th centuries – thus the 'Golden Age' of literature.[210] However, earlier works, when found and documented, were also listed as potential content for the platform.

Matching academic disciplines corresponding to this topic with possible research questions helps contextualizing children's literature in an academic sphere. Hence, below is a list of research questions as these emerged from the children's collection and research conducted at the Henry Madden Library in California, US.[211]

1.	**Comparison of books to the movies that were made from them.**
2.	**Comparison of different editions of the same book.**
3.	**It used to be thought that novels written for children or young adults had to follow a chronological plot or be told from one point of view. How have those theories changed over the years and why? What books were influential as examples of the contemporary view?**
4.	**Selection and research of children's and young adult books with ambiguous endings.**
5.	**What picture books present subplots in the illustrations that are not referred to in the texts? Are there common themes among these subplots?**

---

[209] Hunt, "Children's literature", p.4.

[210] "To Instruct and Delight: A History of Children's Literature" in Random History.com, published in January 2008 http://www.randomhistory.com/1-50/024children.html (Accessed" 13 October 2015).

[211] Children's Literature in Arne Nixon Center in http://libguides.csufresno.edu/c.php?g=288884&p=1927022 (Accessed: 18 September 2015).

6.	New writing trend: Books by celebrities. Quality and effect on other authors.
7.	Ethnicity/gender of authors: should Caucasians be able to write about Native Americans? Women about men? Can the reader tell if a man is writing under a woman's name or vice versa?
8.	What is the history of multiculturalism in children's literature?
9.	Comparative Literature: Comparison of award-winning books over time.
10.	The study of the history of the portrayal of a particular group such as women, Muslims, children, the elderly, the disabled etc., in 20th century children's literature.
11.	Comparative Literature: Comparison of books written about a specific event.
12.	Comparative Literature: Comparison of book reviews for the same titles in publications in the fields of library science, education, and English. What other professions review children's books?
13.	Illustrations: our collection emphasizes cat books, the Alice books by Lewis Carroll, and Oz books. What roles do cats play in children's books? How is Alice depicted by different artists from different countries? How have Oz illustrations changed over the years?
14.	The study of types of literature (e.g.) graphic novels.
15.	Comparison of the books of controversial authors who write on similar topics.
16.	History of censorship in children's literature: Which subjects get censored? Do these change over time?
17.	Comparison of the same fairy tale as told in different cultures.
18.	The history of illustration techniques used in children's books.

These types of research questions could correspond to a list of disciplines such as: History, Cultural History, Literature and Languages, Education, Library Studies, Philology, Textual Studies, Linguistics and Media Studies. According to Wolf et al., scholarship on children's literature was primarily conducted in three different disciplinary fields: literary studies / cultural

studies, library and information science and in education.[212] Apart from disciplinary variations, the audience of Children's Literature could be distinguished according to its professional status, from literary scholars to critics, teachers, authors or simply readers and amateurs. As Peter Hunt suggests, "it is attractive and interesting to students (official or unofficial) of literature, education, library studies, history, psychology, art, popular culture, media, the caring professions etc… Its nature…has been to break down barriers between disciplines, and between types of readers".[213]

Literature aimed specifically at children has been written and published since the 15th century around the world absorbing elements of folktales, different cultural traditions, fairy tales and oral tradition. Intended to educate, instruct and entertain, this distinct type of literature gradually developed rich literary variations according to the time period, geographic or cultural context, author etc. Thus, six genres have been identified by literature critics determined by technique, tone, content of length:

- Picture books
- Traditional literature: further divided in subgenres of myths, fables, legends and fairy tales
- Fiction: fantasy, realistic fiction and historical fiction
- Non-fiction
- Biography and autobiography
- Poetry and verse

Content-wise research

As introduced before, Europeana holds a relatively important number of records on Children's Literature. According to the search term or the search filtering option, the number of retrieved results varies as well as their actual relevance to the topic in question. This suggests that there may be more relevant records in Europeana that could not be retrieved due to different metadata used to describe them.

---

[212] Wolf et al, 2011.
[213] Hunt, "Children's literature", p.1.

In order to document relevant datasets that Europeana currently holds, a number of various searches have been conducted to explore the portal. To begin with, the first search conducted had 'children's literature' as search term.[214] This resulted in records:

- Text (1596)
- Image (194)
- Video (25)

These records ranged chronologically from 1450 to 2014 and geographically from all over Europe, mainly by the United Kingdom.[215] The information provided for these results suggest that the main provider of children's literature in Europeana is the European Library with 359 records. It should be noted here that the providers listed there vary from the data providers which refer to the actual institutions holding the resources. Therefore, as most of the content has been ingested into Europeana through projects or digital libraries, such as TEL, the main content provider in this topic has been the Bodleian Libraries of the University of Oxford. Interestingly, the British Library, which holds a significant number of resources in the field of children's literature and is leading research in this topic with specially dedicated space to its webpage on such issues, it has incredibly weak presence in the portal of Europeana.[216]

Examining the retrieved results of the first search, it was observed that a number of them are quite irrelevant and could not support research in this area. Therefore, a second search was conducted using this time quotation marks to accompany the same search term used before.[217] This resulted in half the records presented before. More specifically '"children's literature"' resulted in records:

- Text (809)
- Image (126)
- Video (3)

---

[214] http://www.europeana.eu/portal/search.html?query=children%27s+literature&rows=24

[215] Note that the chronological range refers to the date of the object while the geographical refers to the country providing the object and not to its geographical origin.

[216] http://www.bl.uk/romantics-and-victorians/themes/childhood-and-childrens-literature (Accessed: October 20 2015).

[217] http://www.europeana.eu/portal/search.html?query=%22children%27s+literature%22&rows=24

The chronological range was different as well, dating records from 1709 to 2014 provided mainly from the United Kingdom. It could be said that the aim of retrieving more relevant results was in a large measure achieved in this second search despite the fact that this observation is again not absolute. Since using quotation marks in search terms enforced the relevance of the results retrieved various searches were then conducted using Boolean queries.

Evaluating the results retrieved as potential sources for research in the field of children's literature, some drawbacks but also potentials have been observed. To discuss drawbacks first, as introduced before, resources retrieved in Europeana after searching on this topic are not always relevant, at least to their full extent. For example, the textual results have rarely to do with the actual literary works for children. Even when searching for *Alice in Wonderland* for example, it is not clear if among the results there is the actual book for further research. Apart from that, even when tracing relevant or useful results, it is not ensured that the actual digital object, and thus link, is still online and accessible. What is useful on the other hand is that the retrieved results include various articles commenting on this work in different national contexts. Europeana thus offers a brainstorm of results relevant to the search term that may lead the researcher to new ideas on how to explore the topic or to new interpretations. Even the idea of displaying a variety of text, image and video results on children's literature sorted randomly could prove to be useful for research as it sheds light in content that the researcher might not have thought of searching and accessing or might not have known of its digital existence.

What is going on outside Europeana?

Identifying material outside Europeana has been a rather interesting procedure with incredibly rich results. It seems that the topic of Children's Literature has attracted academic and public interest in various countries that sought to create associations to host such content, discussions and events. Focusing on digital content already available to the community, there is a number of digital libraries that host collections of this field. Below is a list of such stakeholders that are, as can be seen in the column "Institution", mainly deriving from the United States. This does not suggest that European libraries and archives do not conduct significant work on this topic. It is observed however that the latter offer mostly online services and catalogues that support research based on analogue content found in the shelves of the library. The institutions listed below on the other hand provide open access to digital content on children's literature.

Title/URL	Features	Institution	Timespan	Access/License
International Children's Digital Library	Repository of 2,827 children's books in 48 languages viewable over the Internet	ICDL – Network of partner organizations, advisors etc. Goal: to build a collection of books that represents outstanding historical and contemporary books from throughout the world. Ultimately, the Foundation aspires to have every culture and language represented so that every child can know and appreciate the riches of children's literature from the world community.	15th – 20th centuries	Freely accessible on the web / Copyright policy: all rights reserved
Baldwin Library of Historical Children's Literature	Over 5800 fully digitized children's texts from the Baldwin Library of Historical Children's Literature. A continually growing archive of children's books published in the United States and Great Britain.	University of Florida	19th century	Copyright
Literature for Children	This collection is part of the Publication of Archival Library and Museum Materials project comprised of digitized volumes from the Departments of Special Collections at several of the State University System of Florida libraries. It is a collection of digitized titles published predominantly in the United States and Great Britain.	University of Florida	17th-20th centuries	Copyright
Internet Archive Children's Library	Books for children from around the world. [Item description rather than collection description – each item has information on its copyright policy, date of creation, content contributor etc]	From University of California Libraries (list), the University of Florida's "Literature for Children" Collection, the National Yiddish Book Center, the New York Public Library, International Children's Digital Library, and other libraries.		
The John MacKay Shaw Collection	Collection of books, poems, songs and other works written for children and for childhood with illustrations, photos, prints, and music. They represent some of the most famous children's stories, poems, songs and rhymes of the 18th through the 20th Century in the Shaw Collection.	Florida State University.	18th-20th centuries	

The libraries and archives listed here are just indicative examples of rich content found in leading - to this topic - associations freely accessible in the Web. While not exhaustive, this content concerns world literature works dated since the 15th century. It is thus quite inclusive as

potential content for the platform of Europeana Research consisting of texts and images (illustrations).

Engagement with the community

Apart from the digital libraries and archives presented before, research led in tracing and documenting an even more significant number of associations on children's literature worldwide. Most of them function as networks for either public discussion among teachers, librarians and parents or academic discussion and research. In general, they connect stakeholders with an interest in Children's literature thus creating a community networks. Creating a web space with digital content has not much to offer unless it is linked to a community. Thus, this section has been largely explored as engagement with the community has been considered as an extremely important factor in the initiative of Europeana Research.

By tracing and documenting below a number of research societies existing worldwide on the topic of children's literature, it was attempted to monitor the wider community and their current needs. Most of these societies function as networks for discussions, research and events rather than content repositories.

- Children's Literature (http://childrenslit.com/)

To begin with, Children's Literature is an association founded in 1993 functioning as an independent review source. It is not affiliated with any publisher and accepts no advertising. Its mission is to help teachers, librarians, parents, and childcare providers make appropriate literary choices for children. It has thus developed two services:

1) Children's Literature Author/Illustrator Booking Service: to assist schools, conferences, and other organizations in identifying authors and illustrators

2) Children's Literature Comprehensive Database (CLCD): spin-off from Children's Literature. It uses reviews from a variety of sources, including Children's Literature, to provide an Internet based acquisition, research and reference service.

Based in Washington DC, the people currently involved as reviewers in this network are more than 125.

- Sussex Centre for Folklore, Fairy Tales and Fantasy, University of Chichester (http://www.sussexfolktalecentre.org/, http://www.chi.ac.uk/research/research-centres/folklore-fairy-tales-fantasy)

Moving on to the United Kingdom, the Sussex Centre for Folklore, Fairy Tales and Fantasy is quite keen in giving a space of dialogue and research on the three aspects of folklore, fairy tales and fantasy as creative forces in literature and culture. With national but also international scope, its aim is to provide a forum where writers and scholars from various disciplines can discuss folk narratives, fairy tales and fantasy works, both as independent 'genres' and also in terms of the resonances and dissonances between them, and other cultural forms. The services it currently hosts are:

1) Sussex Folklore Index: An online, multilingual, multi-authored, annotated bibliographic index consisting of links to primary sources of folktales, fairytales and fantasy works available in the public domain, as well as to secondary sources for scholarly discussion on these subjects.

2) FOLKTALES Mailing List (https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=FOLKTALES): Used to facilitate interdisciplinary discussion of folktales, fairy tales and fantasy literature.

Constituted mainly by academics from various leading universities, American and European, it provides a more scientific perspective on research in children's literature.

- International Research Society Children's Literature (IRSCL) (http://www.irscl.com/index.html)

With the aim of bringing together scholars interested in children's literature globally, the International Research Society for Children's Literature is an international scholarly organization established to support and promote research in the field of children's literature with members from over forty countries worldwide. Acting as the worldwide point of reference in this area, it serves in facilitating cooperation between researchers in different countries and in different branches of learning and in enabling researchers in different countries to exchange information, share discussion of professional and theoretical issues, and initiate and coordinate research.

Similar initiatives are documented in various European countries such as France, the Netherlands or Germany which mainly focus on the aspect of networking rather on providing

access to digital content. For the purposes of this project, the networks are listed here either as potential communities to be engaged in Europeana Research or as examples of academic collaboration in this field.

- The Web-Site of the French 'Graduate Centre International d'Etudes en Litterature de Jeunesse' (http://www.ricochet-jeunes.org/)
- Children's Literature in Italy (http://users.unimi.it/childlit/)
- Spanish Research Society Children's Literature (http://anilij.uvigo.es/)
- The Irish Society for the Study of Children's Literature (https://issclblog.wordpress.com/)

In addition to the initiatives listed before, research on children's literature is also conducted in a number of research centers developed within university institutions. Similarly connected to the topic of children's literature, academics and students develop and promote scholarly work in this field.

- Center for Children's Literature and Culture, University of Florida (http://cclc.english.ufl.edu/)

Its aim is to encourage the exploration of this vital area of cultural life through scholarly and critical investigations; through meetings, symposia, and seminars; and through the development of innovative ways to make the research and concerns of its members available to the general public.

Interestingly, this center provides, apart from its networking activities, a list of online resources divided into categories of community and local resources, blogs, illustrations and visual art collections, exhibitions, film and theatre, institutions, organizations and museums and photography.

- Centre for International Research in Childhood, University of Reading (http://www.circl.co.uk/index.htm)

Established in the English Department at the University of Reading in Reading, England, in October 1996, CIRCL aims to promote and coordinate international and interdisciplinary academic research in childhood, focusing particularly on research in children and culture, children's literature, and children and the media.

- The Folklore Society (http://folklore-society.com/)

The Folklore Society (FLS) is a learned society, based in London, founded in 1878, devoted to the study of all aspects of folklore and tradition, including: ballads, folktales, fairy tales, myths, legends, traditional song and dance, folk plays, games, seasonal events, calendar customs, childlore and children's folklore, folk arts and crafts, popular belief, folk religion, material culture, vernacular language, sayings, proverbs and nursery rhymes, folk medicine, plantlore and weather lore.

- The Cambridge/Homerton Research and Teaching Centre for Children's Literature (http://www.educ.cam.ac.uk/centres/childrensliterature/)

Both the Faculty of Education of the University of Cambridge and the Homerton College have a long tradition in children's literature and a strong national and international profile for high quality research and teaching. The Centre brings together researchers, teachers and students in a vibrant, thriving scholarly community.

Aim for:

- o The creation a favourable environment for high-quality and innovative research in children's literature
- o The combination of literary, aesthetic and educational approaches to the subject
- o The accommodation of theoretical, empirical and interdisciplinary research
- o The support of practitioners through dissemination of research findings

- National Centre for Research in Children's Literature, University of Roehampton (http://www.roehampton.ac.uk/Research-Centres/National-Centre-for-Research-in-Children-s-Literature/)

The Centre promotes academic excellence in research into children's literature, primarily through thriving postgraduate MA and PhD programmes, conferences and staff publications.

This long list of research centres, institutions, groups and people surrounding children's literature in all its aspects clearly reveals the wide interest in this topic by the academic community but also the public. It could be said that one of the particularities of this field lies on

its diverse audience. In any case, research conducted here aimed in exploring and documenting the measure of interest and engagement of – at least - the scholarly community with Children's Literature as a significant factor in enriching this kind of content in the new platform of Europeana Research.

Tools and Services

As introduced before, the types of content found in Europeana but also outside Europeana in various institutions on this topic are mainly text and image. Apart from that, interestingly enough, none of the websites/research societies dealing with this field have reference of employing a digital tool or service when dealing with such content. Does this suggest that there is no digital potential in the area of children's literature?

Literature suggests that "the twentieth and twenty-first centuries have seen a great increase in the diversity in children's books, from picture books to flap books to online multimedia texts".[218] This "explosion of multimedia and intertextuality" as has been described, has led to an interplay with digital technologies in terms of storytelling, authoring or simply presenting such content.

In this context, a project called FABULA has been traced as an example of innovative use of technology in the area of Children's Literature.[219] Developed between 1998 and 2000 and funded by the European Commission, Fabula was a multidisciplinary, multinational project aiming to produce a simple-to-use tool for making bilingual multimedia story books in the lesser used languages of Europe. In this way, the project would help ensure that minority languages were not excluded from the Information Age and it would increase the perceived status of lesser-used languages by associating them with new technologies. While having a different focus from actual research on Children's Literature, it is presented here as indicative of technology development in education and story writing with a linguistic perspective. Its particularity also lies in the fact that its target audience is children who are not perceived as simply 'consumers' but they are rather actively involved in the process.

---

[218] "To Instruct and Delight: A History of Children's Literature" in Random History.com, published in January 2008 http://www.randomhistory.com/1-50/024children.html (Accessed" 13 October 2015).

[219] Edwards, Pemberton, Knight and Monaghan, "FABULA: A BILINGUAL MULTIMEDIA AUTHORING ENVIRONMENT FOR CHILDREN EXPLORING MINORITY LANGUAGES". Retrieved from http://www.it.bton.ac.uk/staff/lp22/fabula.doc (Accessed: 18 September 2015).

Based on the idea that "creating an artefact is a more powerful way of learning than consuming another's product", the project planned to involve its target audience, children, in the process of creating story texts based on fables, myths or fairy tales, audio recordings and images then combined into a full Fabula storybook. In this way, children would develop a range of social and organizational skills as well as deepening their understanding of Information and Communication Technology. Thus, one of the requirements of the software, apart from being simple to use, was to create an aesthetically pleasing environment. Interestingly, it was considered that creating digital versions of bilingual books "would give minority languages the high status attached to computer based material and give children enjoyable, engaging language learning/exploration tools". Digital content was thus highly valued as an asset in engaging the community to interact with this particular type of content.

What was also considered important was the fact that, for children, "the prospect of having their work published for the rest of Europe to see has proven highly motivating".[220] The software developed, apart from its potentials it offered to teachers and students, it also formed the focus for a community of users with a commitment to high quality and innovative language teaching, keen to look outwards to share their experiences". For the purposes of Europeana Research, knowledge rendered here would be:

- To actively engage users in editing and elaborating on resources, stating clearly the provenance of the new editions
- To build a network of users/researchers
- To enable researchers to share knowledge and publish views
- To create an environment simple to use

Another interesting technological innovation that should be discussed here is digital storytelling. "Digital storytelling at its most basic core is the practice of using computer-based tools to tell stories".[221] Applied mainly for educational purposes but also as a mean to enrich user experience in GLAMs, this service is in general the idea of combining the art of telling stories with a variety of multimedia, including graphics, audio, video, and Web publishing. As an increasingly popular technological asset, digital storytelling has various potentials and

---

[220] Edwards, Pemberton, Knight and Monaghan, "FABULA".
[221] "What is Digital Storytelling" in http://digitalstorytelling.coe.uh.edu/page.cfm?id=27&cid=27&sublinkid=29 (Accessed: 20 October 2015).

applications that have only been partially recorded or explored. In respect to education, it provides both teachers and students an effective instructional and learning tool respectively by encouraging the integration of new multimedia in the classroom or by enhancing the student experience through personal ownership and accomplishment. Despite the fact that this does not correspond directly to researchers of Children's Literature, it does provide a more contemporary view of how literature is taught and perceived, how digital means transform or influence the relation between user and content and sheds light in potentials given to children's literature to be enriched or presented to the public in the GLAM sector.
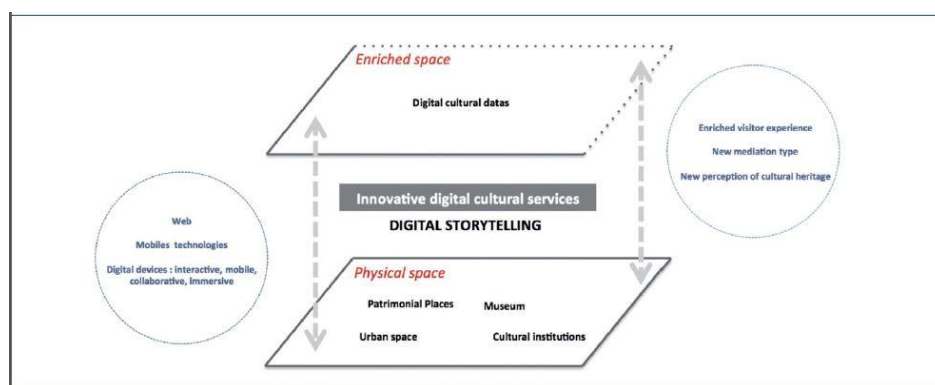
Figure 35. *Depiction of the enriched space visit in respect to digital storytelling, AthenaPlus Project Report, p.12.*

It could be thus suggested that such digital services have a dual function. Apart from enabling new uses or enhanced elaboration of digital resources, they also impose new research questions as they create a new reality of interaction between the user and the respective field/content. In this context, digital storytelling, as a new form of digital mediation of cultural heritage, has served, among other, in enriching the visitors' experience cycle by offering them new action and participation capacities. In order to guarantee cultural data attractiveness over time, in research, education or simply in visiting experience in libraries, museums and archives, such services are incredibly valuable and become objects of study themselves in respect to their function and outcomes.[222]

Despite the fact that there is no other reference to digital tools or services employed in research in Children's Literature, the type of tools and services that would be potentially useful in this field could be inspired by the type of content of the area, in particular text and image. Therefore, annotation tools would be such an example or digital illustration tools.

---

[222] "Digital Storytelling and Cultural Heritage: Stakes and Opportunities", AthenaPlus Project Report, 2015, p.14.

Researchers' views

Desk research was further complemented by informal discussions with researchers of the field aiming to explore and document their current digital practices and in particular their needs that Europeana Research could attempt to meet in the future. To this purpose a questionnaire was prepared on investigating how the community conducts research in this area, which methods it employs, available tools and resources it currently uses and future aspirations. This structure was not strictly followed but it rather encouraged an open discussion with researchers according to their interests or points that they would like to stress more which lasted 40 minutes approximately.

Research societies and networks of Children's Literature, as these were identified and documented before, served as the initial contact point for reaching researchers in the field. Eventually, the sample of researchers that contributed in this task was mainly mobilized through the Spanish National Association and Investigation on Children's Literature (ANILIJ).[223] In particular, three discussions were held in September-October 2015 with:

- Dr. Juan Senís Fernández, Department of Teaching of Languages and Humanities and Social Sciences, Faculty of Humanities and Social Sciences, University of Zaragoza, Spain
- Dr. Xavier Mínguez López, Department of Teaching of Languages and Literature, University of Valencia, Spain
- Dr. Patricia Hansen, Independent Researcher

While this cannot be considered a representative sample of researchers in Children's Literature, it will help to illustrate here the different parts of the interview questionnaire as well as points made earlier in presenting desk research. All three researchers have conducted extensive research in this area and had valuable insights to share. The issues that emerged from the discussions held were:

1) Geographic coverage of available digital material

---

[223] Spanish National Association and Investigation on Children's Literature (ANILIJ) http://anilij.uvigo.es/

Researchers observed a gap between national contexts in producing digital data and thus be represented online in digital libraries and archives. As digitization projects are not equally developed in all countries, their digital presence is also unequal. Apart from that, even in cases where digitized material is produced, its good quality is not ensured. Producing thus digital content was considered the first and most significant step before moving on to building tools and services.

2) Collections as means of organizing content

Building collections on children's literature was considered to be the best way to organize and orient research. Collections either thematically or content-wise would potentially serve as ready-made research material.

3) Easy access to references, databases, journals and books

Researchers heavily rely on textual resources, such as journals and books, to support their research argument and explore their research question. Thus, one of the issues stated in the discussions was the need for online access to such multilingual material.

4) Map of publications

In relation to the previous suggestion, it was considered useful to have a constantly updated record of publications on a specific research topic that would serve researchers to trace what are the new publication trends on Children's Literature. This could be translated into a map of publications which would potentially enforce networking aspects in the field.

Discussion

Following the desk research conducted as well as the discussions held with researchers of the field, it is suggested here that the area of Children's Literature could indeed serve as a research topic in the new platform of Europeana Research. The arguments supporting this suggestion are:

1) Enrichment of content: Currently the online presence of content related to Children's Literature in the portal of Europeana is important but it is still quite underrepresented. Therefore, there is room for further enrichment, considering also the number of digital libraries that have been traced and documented earlier with such available digital resources.

2) Networks: Evaluating the number of research associations that currently exist in this area, there is a surprisingly wide interest in academic circles that are already joined in various networks around the world. Mobilizing them in respect to new digital resources, tools and services will immediately embrace this research path in Europeana Research with potential users and community.

3) Interest for further development of tools and services: Despite the fact that there is only a small number of researchers in this field employing digital tools and services on such content, the prospect of developing such in the future gathers great attention among researchers. Moreover, the low engagement with digital technologies so far could be seen as an opportunity for new developments from the scratch in a field which is currently not well imprinted in respect to digital tools and services.

6.5 European School History of the 20th Century

Researchers that explore the school history of the twentieth century come from a wide range of disciplines including (but not limited to) history, sociology, archival science, the branches of education science, and the various disciplines connected to subjects taught in schools. Relevant types of material range from public statistics to personal diaries, from newspapers to policy documents. Research can cover anything from the history of a small village school to the development of a school system exposed to international influences. It is a subject that would benefit greatly from having online access to a variety of material of different types and from different countries.

Delimitations

This topic is delimited in three ways: focusing on the 20th century in Europe, and on the history of schools rather than education in general. That said, these limits are fuzzy. In particular the temporal endpoint has proved vague, shading into the early 21st century. Europe has been taken to include the European Union as well as some non-EU members such as Norway, Serbia, and Switzerland, as they either had contributed much material to Europeana relevant to this topic, or because they have prominent researchers in the field. It also proved difficult to determine exactly where education in schools ended and other forms of education begun; many researchers that we have been in contact with are researchers of the history of education, or of

specialised areas within the history of education in general, such as teaching children with special needs. This should be borne in mind when considering our discussion below.

Process Outline

The three basic steps of addressing the topic were 1) an explorative evaluation of material of possible interest that could be found in Europeana and of the ecology of European museums with collections pertinent to school history; 2) a series of mainly video conference and e-mail interviews with European researchers in the history of education, and examination of related material; and 3) a compilation and analysis of the findings. These steps are detailed below.

Explorative Evaluations

Europeana

The evaluation of the Europeana database was carried out by searching for a variety of terms related to schools and schooling in a number of languages (English, Norwegian, Swedish, German) and categorising the results, giving illustrative examples. Each media type (image, text, video, sound) was looked at separately to facilitate categorisation. The 3D media type was deemed to contain too little relevant material to examine. (The search terms we used can be found in Appendix 4.) In cases when a search yielded too many hits to examine all of them, random samples were checked. The results as a whole are thus not statistically representative, but give at least an indication of what material is available.

No obvious datasets resulted from these searches; instead, we found possibly relevant material spread out over several datasets, often with only a few objects in each dataset. The results were compiled and divided into rough categories for presentation to the researchers on one of the instruments we used as basis for the interviews.

The results of the exploratory study prior to interviews are in no way exhaustive. The following tries to give an overview of the types of material we have found by roughly grouping the results; it is only meant as a guide to what can be found, not a full description of everything there is to find in Europeana. The examples are only that: examples in case you want to check in Europeana (europeana.eu) what a group might include.

<u>TEXT</u>

- Historical texts, monographs and articles on individual schools or school disciplines

- History of education in individual European countries, e.g. National education of Ireland, History and present position of primary schools in England etc.

- Pupils' journals

- Ethnic (minority) schools (articles, books, short letters on the history, status, position and problems of schools where teaching is given in a minority language)

- Religious education (papers, books and research on the topic of religious schools, their persistence and problems over time etc.)

- Jewish schools - reports, bulletins and registers

- Political manifestos and campaigns on education and schools

<u>IMAGES</u>
- Photos of written documents, e.g. examination papers, maps.

- Pictures of school children/staff

    o Class/group photos

    o Children in class/during lessons

    o Various school staff

    o Portraits

    o Pupils in various activities, incl. sports

    o School performances (plays, concerts etc.)

- School interiors (Classes)

- School exteriors

    o Aerial photography

    o School buildings

- o "Mobile classrooms"

- o School buildings as backgrounds to group photos

- School objects

- School work/crafts

  - o Equipment & pedagogical tools (blackboard, ink well, notebook, test tubes)

  - o Didactic material (posters, tables, illustrations)

  - o Prizes, medals

  - o Clothing

VIDEO/SOUND

- News clips and newsreels including material from/on education, schools or school children, e.g. Finlandia-katsaus

  - o Coverage of school inaugurations and building ceremonies

- School documentaries of varying length, e.g. Ar Scoil/At School

- Projects by school children, e.g. within project "minneinsamling-i-skolen"

- School children interviews on various topics (like the Christmas Story)

- TV Scolaire / School TV

- Memories of what school life was like, e.g. "Interview with Daphne Glick"

Museums

Using web searches, we identified a number of European museums that contained collections of material which could possibly be of use for researchers interested in school history. They were listed along with information about whether they had already contributed material to Europeana or not. A possible categorisation emerged:

1. *National Museums of Schools and Education* – large national collections, often digitized, meant for research as well as public access.

2. *Research Museums of Pedagogy and Education* – connected to research institutions. Large collections meant for research. Good chance collections have been digitized.

3. *Major School Museums* – collections of thousands of objects, possibly digitized and with research facilities.

4. *Minor/Local School Museums* – collections of hundreds of objects, probably not digitized, either under the aegis of a larger institution, or run as a hands-on "experience" for families/school classes.

Most researchers that we interviewed agreed with this basic structure, although suggestions for adding some specific types of museums (such as historical museums and children's museums) were voiced, and occasional alternative structures were offered.

A great many school museums give the impression of being small institutions with small, often non-digital collections. Occasionally, they exist under the aegis of a larger museum which may have provided data to Europeana. Only a few of the museums in our pilot list could be found in Europeana. This goes even for national museums for school, education and pedagogics.

Researcher Interviews

Researchers interested and active in history of education were identified in various ways:

- International publications were checked for researchers who have published in areas relevant to our topic. Similarly, conference programs within the field were perused to find researchers who had presented papers on relevant subjects.

- Journal editors and editorial boards, conference organisers, and executive boards of international organisations were selected if their research profile matched the topic.

- Persons with relevant contacts in our own networks were asked to recommend possible researchers for us to contact.

- Once we got in contact with researchers, we also asked them for suggestions about whom else to get in touch with.

(Our interviewees are listed in Appendix 3.)

Following these four principles in compiling a list of researchers soon led to many of them recurring, suggesting some degree of "saturation" in coverage of the field. In selecting whom to contact, we aimed at a broad geographical coverage but also tried to include researchers of varying seniority and with different research focus. Often, we had little control over exactly who agreed to assist us, however, and in some cases, we were referred to people or publications that did not fit our initial template. Our approach was consistently pragmatic: if we were pointed to a particular source of information, we included it in our overall evaluation. This research field is not equally developed in every country, so it makes little sense trying to find corresponding types of researchers, publications, organisations, or conferences in each country. Instead, the material we have collected offers a wide range of perspectives on the requirements of the field, including various disciplinary approaches, which we have tried to aggregate in our discussion.

Slightly fewer than half of the interviews (9) were carried out via Skype or in person. Each interview was preceded by an information e-mail which contained our central questions and some information about them, as well as a pdf with the categories of what Europeana material we had found and the museum categories. The interviews took about an hour. They were unstructured and were allowed to focus on the expertise and research interests of the interviewee. The issues we explored centered around the following (with focus on 1.):

1. For researchers in History of Education,

   - what types of material from Europeana would be useful to include in Europeana Research?
   - what cultural-heritage institutions (may) hold material that would be useful to include in Europeana Research?

2. What sort of information (metadata) about the material would be necessary?

3. Comments or suggestions? (E.g. whom else we can contact regarding this.)

Ten interviews were carried out via e-mail. For these, the questions were accompanied by short explanatory texts and were presented as four separate questions rather than sub-questions.

We also conducted one group interview at the Sixth Nordic Conference on the History of Education in Uppsala (Sweden) in which four researchers took part.

Other Material

The miscellaneous material that came our way and which we included in our compilation and analysis of the topic included:

- Call for Paper for an international symposium on School Memories ("New Trends in Historical Research into Education: Heuristic Perspectives and Methodological Issues") in Seville 22–23 September, 2015.
- de Coninck-Smith, N. & C. Appel (eds). *Dansk skolehistorie, hverdag, vilkår og visioner gennem 500 år*. Aarhus: Aarhus Universitetsforlag: 2013–2015.
- Zounek, J. & M. Šimáně. "Resources mediating the history of primary schooling during the normalization period in the Czech Republic (1969–1989)." Unpublished conference presentation. ECER, Budapest: 2015.

Findings and Discussion

Below follows a brief exposition of the interviews with the researchers. It is worth pointing out that although we attempted to contact researchers in as many European countries as possible, we eventually had to settle for a smaller number, partly due to time constraints and partly because it proved difficult to get in contact with some researchers. The selection is not representative, and as we relied partly on our own networks, there is a bias towards certain countries. It is our opinion that the number of researchers and countries is still large enough to give a varied picture of research in the field as it stands today.

Sources and Material

Researchers who explore the school history of the twentieth century employ a wide range of material and sources in their scholarship. Exactly which material a particular researcher uses varies depending on nationality and general research area, but taken together, our sources give a fairly coherent picture.

It is worth pointing out that the examples of possible research material that we identified in Europeana were on the whole considered interesting material. Many researchers identified a number of the material types, while a few stated that all the examples could be used for interesting research. A couple of researchers also observed that the Europeana examples included material suitable for research on both macro (aggregate) and micro (individual school or district) level.

Another point that became clear from our various sources was the overwhelming importance of textual sources to current school-history research. *Dansk skolehistorie* is one source that stresses this, including the fact that for the more recent periods, printed matter dominates the sources. "Only to a very small extent have non-print archival sources been used" for the chapters in the final volume (from 1970 to present). Because of this focus on text, we will begin by detailing the various textual materials used by researchers.

Text books and reading primers are by far the most common type of material that researchers consider useful, and which was most often indicated as missing in Europeana. There is a major concern with textbooks as an important kind of research material, and a few researchers also added that textbooks for religious and civic education would be valuable. The interest in textbooks varied in different countries, however, and it is worth observing that there are already several centres (national and international) which hold major collections. Foremost of these is the Georg Eckert Institute for International Textbook Research (GEI) in Germany but there are others as well. Because of the pre-existing collections and the availability of textbooks in libraries, some researchers were of the opinion that metadata would be enough in Europeana Research.

A range of other sources came up during our investigations. While not easily subdivided, one group of sources comprise material in some way directly connected to the running of the schools. Rough categories include *pedagogical material*, *administrative material* and *propaganda material*, although there are no distinct boundaries between these. Pedagogical material would include such material that directly involved the meeting between teacher and student. Examples:

Pedagogical material
teaching instructions
pedagogical guidelines
curricula
data on student achievements
student notebooks

Administrative material describes the practical management of school life, including records "such as description of everyday expenditures, organization of everyday life, solving of simple

problems (and resources available to schools)" as one researcher describes it. Some specific examples that were brought up are:

Administrative material
reports and records
time schedules
teachers' diaries (in Spain)
"visitation reports" (in Denmark)

In some countries, during some time periods, classrooms contained material that in some form functioned as propaganda for the current regime. Some examples from Latvia include:

Propaganda material
portraits
slogans
textbook pictures
pictures of classrooms
booklets and propaganda posters

Several researchers also emphasised the need for a social context for the source material. Such context, it was stated, can be found in *political material* regulating the education system. A number of researchers explicitly wished for such contextual material to be available in digital form through Europeana Research, for instance including full-text legislation to facilitate comparative studies.

Political material
national, European and international legislation (laws, bills, ordinances, statutes etc.)
public announcements and proclamations
parliamentary papers
political manifestos
grey literature (unpublished reports)
public surveys and inquiries
material from teachers' unions and associations

Necessary context can also be provided through *statistical material*, which one researcher described as "the core information to start the analysis of historical transformation [without which] it might be hard to understand the discussions and processes surrounding [the] education system."

Statistical material
salary statistics
demographic data
number of pupils, schools, teachers
school locations

What appears to be an essential type of written sources are *publications*. These fall roughly into two kinds: primary or historical (contemporary commentary, debate, criticism) largely in the form of newspapers and trade periodicals (journals and reviews) but also pamphlets and other publications; and texts with a historical perspective. The latter kind include biographies, amateur histories and anniversary publications as well as research publications in the form of scholarly books, Ph.D. dissertations and journal articles. One researcher suggested that academic publications which were out-of-print or published Open Access be considered for inclusion in Europeana Research.

Former Yugoslavia provides a case in point when it comes to trade periodicals. One researcher suggested that digitized versions of *Učitelj* ("Teacher") from 1882 to 1941 (although not published during World War I) and *Pedagogija* ("Pedagogy") from 1946 would be valuable for researchers. The latter contains studies on schools from all six Yugoslav republics (Slovenia, Croatia, Bosnia & Herzegovina, Serbia, Montenegro and Macedonia). Through the former journal, it is possible to track the influence of German schools on the Serbian/Yugoslav school system, and through the latter, the Soviet and then Western influences can be similarly traced.

Another important group of text material comprises recollections, memories and so-called *ego-documents*. These include personal diaries and letters as well as published recollections, autobiographies/memoirs, and interviews. One researcher noted how autobiographies often have sections devoted to school memories which can be used as material for content analysis. Related to this is research into oral histories, interviews and recollections which are stored as text, sound or video data.

Photographs are less common in current research material, but are considered to be very interesting and valuable by a number of researchers (although one researcher observed that research based on portraits and class photos was always focused on local interest). Apart from the types of images that we found in Europeana, and presented in connection to the interviews, several other types of photos were mentioned.

Foremost among these types were pictures of *didactic material* other than textbooks, primarily teaching charts or posters, tables for teaching languages or maths and various illustrations, but also slides. In some countries, there are collections of these (Switzerland and Denmark); but researchers in other countries seemed equally enthusiastic about this kind of material.

Photographs of people and places also appeared to be popular with many researchers. Children in class or involved in sports, performances, or other extracurricular activities were brought up as possible research material, as well as images of teachers. School building exteriors, including when featured as background to group shots, aerial photography of schools, and pictures of "mobile classrooms" were other examples mentioned in the interviews.

One category is photographs of *physical objects.* Although occasionally unclear, mostly these were seen as possible ways to examine the objects if these were no longer extant, or available in a collection far away. Interesting objects would include equipment and pedagogical tools (such as blackboards, inkwells, desks, notebooks, test tubes, school bags), clothes and school uniforms, art in/of schools, and objects related to school activities such as arts and music, crafts and (in particular) design, and sports and gymnastics. A particular type of object of interest to one researcher was the "school plaque" – a plaque at the entrance of schools that commemorated one or more people.

A final type of material, which was pointed out by researchers as interesting especially for research of the latter half of the twentieth century, is *audio and video* material. Apart from what we had already found in Europeana, special mention was made of educational radio and TV programs (of great interest to some), school documentaries (about schools and for school use) and "propaganda" films produced by the national education system. Examples of such propaganda films were given from the Baltic countries, Spain and Belgium. One research project examined video documentation of school effort, and another researcher suggested that amateur movies could have value for researchers. The many film projects made by school children that appear in Europeana caused one researcher to observe that school projects are

interesting if they are old – this may be taken as an argument for not including that sort of material in Europeana Research.

It may be worth noting that the current interest in oral history also means that recordings of interviews – professional and private – might be a worthwhile type of material to include in a school-history collection in Europeana Research.

Some other observations regarding future material in Europeana Research were offered in the interviews. The inclusion of material concerning minority pupils and special schools was urged by a few researchers, as was material concerning extracurricular activities (our exploration of Europeana may have missed such activities due to the choice of search terms). Statistics were seen as important, and one researcher in particular mentioned statistics and documents describing cross-school relations (such as official competition between schools or cross-school agreements). Several researchers also brought up the need for translations, to make as much use of the multilingual material as possible.

Research into school history makes use of a multitude of document types, as is evident from the source and literature essays in *Dansk skolehistorie* as well as the interviews carried out within the framework of this deliverable. One thing that has become evident from our work is that most material used by researchers can be found in either archives or libraries, although some types of material, in some countries, are found in special (private or public) collections. A relatively small part of material is taken from museums, and predominantly from museums with a national or research mandate. An exception to finding sources in memory institutions is the collection of material from school web sites.

Another point made by a few researchers is that digital material can only take historians so far; for most material, they need to see the original document. Digital versions are helpful in that they make it possible to decide whether certain sources are relevant or not, and whether a particular archive, library or collection is worth visiting. It was also suggested that archives containing photo, video and interviews would be useful in showing different aspects of school history. The reliance only on (printed) text material might be coming to an end.

Areas of Research

Research interests in the field of school history vary between researchers and over different European nations. Although by no means ubiquitous in the education-history research community, a few areas can be identified as currently popular or up-and-coming.

Exploring school systems during particular, in most cases authoritarian, regimes of the twentieth century appears to be a research topic of great interest in several countries. The school form during the Nazi era in Central and Eastern Europe, Baltic schools during the Soviet era, Czech and Slovak education during the Normalization period (1969-1989) as well as the history of schools in Belgian and Danish colonies are some of the research areas indicated by interviewees. Propaganda in the classroom is a closely related, current topic.

National educational regulations and policies, including the influence of other European systems on local schooling, are noteworthy topics that require access to legislative documents, records, pedagogical journals, diaries, textbooks etc. in order to track school development. Other areas that seem to garner a great deal of interest include studies of school materiality and body discipline (which requires access to pictures of classroom objects or the objects themselves); memory studies (in which individuals' recollections of their schools days are examined); explorations of (changes in) classroom culture and school traditions; and the exploration of emotions and oral history.

Europeana, and in extension, Europeana Research, is a transnational project facilitating cross-border research collaborations. One researcher is keen to have Europeana Research to stress how European nations build education collectively and another is interested in conducting comparative research on national education policies.

Despite a clearly expressed willingness and certain fruitful regional cooperation among the interviewed researchers (a Swiss researcher refers to a broader cantonal cooperation for instance), significant international collaboration between researchers in this field is still lacking. The main reason for this, according to some interviewees, is lack of funding for such projects. The few examples of cross-border research projects in school history that we came across were either cases of researchers from different countries who published chapters on their respective countries in the same book; or they were the result of individual researchers or research groups moving from one country to another and bringing their material with them. Lack of funding and linguistic barriers often prevent collaboration even in cases where there is ample and interesting

material, such as in the case of the Austrian Monarchy: today, there is an archive full of material from the entire Monarchy, but no interest from the countries that used to be part of it.

The existing transnational cooperation is based on certain historical and political grounds. Despite weak financial support and limited accessibility to EU funds due to some countries being non-EU countries, school historians from former Yugoslavia regularly gather and discuss research on educational development in former Yugoslavia. Similarly, cooperation in the form of conferences and joint publications also exists among Nordic researchers. Researchers from the Baltic countries point to some collaboration related to the Soviet period and express an interest in further cooperation, especially in order to recover lost or destroyed data/material from that period. As a result of the cultural similarities due to multilingualism, there has also been some cooperation between researchers from Switzerland and Luxembourg.

Metadata

In order to enable Europeana Research to provide as useful resources as possible to school historians, the researchers were asked what metadata would be necessary and valuable for the resources to have. This section summarises their views on this topic. There is a certain overlap in the following headings, depending on, for instance, the approaches of different interviewees and the focus of their particular area of research.

The interviewees made some general observations regarding metadata and their structure. One researcher acknowledged that the many different methods for the collection and presentation of information (for example, about objects in museums) presents a challenge in terms of creating standardized metadata. Another researcher proposed the use of a metadata "template" in order that the relevant national, political and legislative context is captured. Finally, there was a concern with linguistic problems: the many European languages were felt to prevent comparative projects, and metadata that come in "all" European languages were seen as a partial solution.

Basics: Year and Place (for all source types)

Time and place are the two fundamental metadata according to interviewees, and their requirements are in line with the current Europeana metadata. For some, an exact *year* is essential, but its *decade* can also be useful, indeed preferable if the sources available do not point to a particular year. The *century* of a material is regarded as too imprecise, however. One researcher mentioned that the material's *time period* could also be useful.

Metadata about place was referred to in many different ways and at different levels including *source*, *origin*, *country*, *location*, *geographical place*, and *the school's number* (where applicable). This kind of information provides, for example, images with invaluable context. One specific recommendation by one researcher was to classify material origin in terms of *city* or *countryside* (where applicable).

Content and Keywords

A number of interviewees suggested that material (in particular images) should be accompanied by descriptions, or that its content or topic should be included. This could take the form of a description of what is depicted on the image or a description that provides some sort of context. One researcher suggested that purely "factual" descriptions of images or objects (such as often feature in school museums) lack an element that describes the importance and relevance of the material and that makes links to relevant aspects of the history of education.

Three researchers referred to the importance of keywords and it was indicated that the keywords currently used in Europeana would benefit from being more specific. Another researcher acknowledged the difficulty of implementing a set of keywords that meets the needs of all researchers and also suggested that keywords are often very research-specific. An awareness of the categories or subcategories that are widely recognised by historians/researchers in history of education would therefore be beneficial. An example provided was that of *ego-documents*, which includes texts such as letters, diaries and pupils' journals.

People

A number of researchers described how metadata that identify specific people would be relevant to the material: *authors*, *photographers*, *subjects* (those depicted) or *participants* (in one case expressed as "protagonists"). In the context of school material, this could mean any and all of the following: *teachers*, *inspectors*, *principals*, *school directors*, *school ministers*. In the case of an image, the name of the photographer and whether they are/were professional or amateur are useful information, given that their respective approaches to photographing a classroom would differ.

Relation between Material and Collection

Information about by what criteria a material was selected for digitization and inclusion in Europeana, and what relationship a material has to a larger collection or corpus is essential to

researchers. This has a practical value (assisting a researcher in planning their research when accessing materials in different locations) but is also helpful in placing material in a larger context or contexts. One interviewee emphasised that sampling should be done by the researcher; research cannot be conducted on whatever material happens to be available/digitized.

The ideal is a fully digitized collection. This is the hallmark of good (often larger) research collections, but it is not necessarily the norm. If a collection is not fully digitized then information about this must be included. The reasons for partially digitized collections might include lack of funding or time/resources. The status of an ongoing or planned digitization process may also be useful information.

In other cases, there may be other circumstances that account for the limited nature of a collection or information about it. These might include materials which were deliberately destroyed or removed for political/ideological reasons. One researcher referred to the process of "recreating" or "re-finding" a collection that had largely been destroyed.

Material with little or no existing metadata (for whatever reason) may be problematic, but not necessarily impossible, to utilise in research. Anonymous oral histories (recordings/transcriptions) are one example. Information about the interviewer or about how the interviews were conducted can be of use. This issue was raised in terms of the practical implications for the compilation of metadata, but no solution was offered.

The "selection" process might also be seen as the historical process of the creation of a collection. For example, a complete collection of school books might be seen as either all school books that were published or all school books that were actually used in schools. Therefore, information about usage (anything from legal sources that govern selection of textbooks to teachers' journals and diaries) and temporal and geographic variations of usage can also be useful for researchers. Collaboration with other institutions such as libraries can prove useful at this stage.

Historical and socio-political context

Several researchers agreed that more information about context is a desirable addition to existing metadata. Such wishes suggest that they would, in fact, like to have information about national legislative contexts available in connection to other material, with possible links

between relevant research material and policy documents or legal framework. One researcher made this particularly clear, as she would like to see "a systematic overview of important national policy documents on education (including legislation, full texts)" because it "would be nice to do comparative research in this field, using documents on Europeana Research."

Exactly where contextual metadata ends and contextual documents (research material) begins is difficult to say, but it is clear that researchers would like some indications of such context in connection to the material. A number of the interviewees referred to the national, political and legislative contexts of different historical periods that have a particular importance in countries with periods of occupation, totalitarian rule or repressive governments. These are referred to variously as socio-political or socio-cultural contexts. This could also apply to ideological influence (see for example "Défense [nationale] spirituelle" in Switzerland). Periods in which societies are in transition to democracy (from authoritarian rule) also have their own particular characteristics. Other time periods might also be relevant, for example the period of identity crisis in 1980s Switzerland.

A similar problematic metadata/data complex is metadata that provides (information about) access to, or indicates the existence of, related academic material in the field of the History of Education. Although various researchers mentioned, for instance, problematization, contextualisation and theory it seems likely that they really would like metadata elements that contain links to or information about related publications such as journals and books.

Possible Usage, Rights and Licenses

Several researchers stressed that information about how material can be used is important. Such information is already largely present in Europeana, although some researchers also would like guidelines for how to access material. Specific terms of use – whether material can be used for research or publication, copyright issues which may vary from country to country, Creative Commons, public domain – also came up repeatedly in the interviews. A few researcher referred to difficulties in the research process when encountering restrictive archival laws (in particular in terms of material containing personal data), an issue which will be a problem in any future Content Strategy involving school-history material.

International comparative studies

As has already been noted above, school-history research seems mainly to focus on the nation state boundaries. The pan-European nature of Europeana material could encourage more transnational research, but such encouragement places demands on the metadata.

Two researchers raised the issue of the role of metadata in supporting (or hindering) comparative research. Education is referred to as a collective European project, created and evolving through collaboration, exchange and influence. This should be borne in mind both when analysing and categorising material, they pointed out. The main criticism of the current Europeana metadata is that providing only a country for a material is insufficient and risks obscuring interesting associations and approaches. While this element serves an obvious purpose, it could be supplemented by additional information that takes into account the fluid nature of national boundaries and changes in (political) status over time, the role of European colonisation and exchanges of ideas. However, one of the researchers acknowledged that the process of adding suitable metadata may be difficult.

There are four specific recommendations from researchers. One is that all "foreign" references in the source material should be treated as potentially important and highlighted (in the form of relevant metadata). Another is that a territorial designation be added, as a way of augmenting the category of political state. It is also suggested that there should be a category which allows for a lower levels of governance. Finally, as was indicated above, inclusion of important national legislation and policy documents on education, in full text, would also aid comparative research in this field.

Language is identified as another potential barrier to comparative research by a few researchers. This stresses the importance of available metadata in all European languages, and information about what language a material is produced in, in order to facilitate this kind of research.

Other Institutions

Apart from adding material from Europeana, Europeana Research would gain value to researchers by bringing in material from other institutions. The following discusses some types of institutions that we discovered during our explorations of the field, or which were recommended by the interviewees as institutions and collections of relevance to their research or to research in the history of education in general. These institutions might therefore be of

interest to Europeana Research, whether in terms of inclusion of content or of metadata. As every country has a number of such institutions, however, we look at particular types and give a few of the more typical (or oft-mentioned) institutions.

Digitized collections

Major digitized collections

There are a number of large, established institutions with digitized collections which are unlikely to make their collections available through Europeana Research. Inclusion through metadata only might be the best (and indeed only available) option in these cases, when the actual content is unavailable. Two of the prime examples of these types of collections, brought up repeatedly by interviewees, were The Georg Eckert Institute for International Textbook Research (GEI[224]) and the German Institute for International Educational Research (DIPF[225]).

The Georg Eckert Institute provides access to digitalized historical textbooks in a range of collections from a number of European countries. They also allow users to carry out full-text searches in the GEI holdings that have been digitalized thus far.

The Research Library for the History of Education (BBF[226]) is an international research library for research on the history of education with its own archive. The BBF is part of the German Institute for International Educational Research, an institute belonging to the Leibniz Association. The BBF has digitized selected parts of its holdings and provides free access to these collections via the internet:

- *Scripta Paedagogica Online* is a digital text archive on the history of education in the German-speaking countries) which contains more than one million pages digitized from educational and religious educational journals, monographs, reference works, teacher directories, school programs and annual school reports[227].

---

[224] http://www.gei.de/home.html
[225] http://www.dipf.de/
[226] http://bbf.dipf.de/en
[227] http://goobiweb.bbf.dipf.de/viewer/browse/

- *Pictura Paedagogica Online* is the digital archive of images in educational history containing illustrations of books, historical postcards, photographs and pupils' drawings[228].

- *Vox Paedagogica Online* is a digital archive of audio recordings focusing on the history of education in German-speaking countries in the 20th century with self-portraits of retired education scientists[229].

There are examples of institutions that have created portals in order to provide access to various digitized sources, but which also contribute to Europeana. The Institute for Social History has a collection of digitized sources (from various countries) containing several texts, images and videos of interest for historians of education[230].

Smaller institutions with digitized collections

There are also smaller institutions with existing digitalized collections that could, potentially, be made accessible through Europeana Research. An example of such an institution is *The Estonian Pedagogical Archives and Museum* (EPAM) which focuses on the history of Estonian school and pedagogy. Most of the images/slides in the collection are digitized (and some of the film, audio and text material) and the collection is accessible through an online database[231].

There are also research groups that are creating digital resources for research in school history and history of education, sometimes with support from Research Councils (Switzerland) or Ministries of Education (Portugal). "Bildungsgeschichte Schweiz[232]" in the German-speaking part of Switzerland as well as its Italian counterpart "Storiascuola[233]".

"Bildungsgeschichte Schweiz", associated with Zurich University, includes: statistics, school programs, curricula and biographies of education ministers. "Storiascuola" includes school programs and other related documents organized by type of document, periodization as well as the type of school (preschool, primary and secondary). These projects were initiated (material was collected, evaluated, documented, digitized) by researchers/research groups, carefully evaluating the material according to scientific criteria for inclusion (Switzerland) as well as

---

[228] http://bbf.dipf.de/VirtuellesBildarchiv/
[229] http://bbf.dipf.de/en/digitale-bbf/vox-paedagogica-online
[230] http://www.socialhistoryportal.org
[231] http://arhmus.tlu.ee/cgi-bin/epam
[232] http://www.bildungsgeschichte.uzh.ch/index.html
[233] http://www2.supsi.ch/cms/storiascuola/2014/02/14/programma-per-le-scuole-maggiori-femminili-del-1867/

creating standardised (DCMI) and discipline-specific metadata descriptions of, and critical comments (evaluation) about, the included materials ("Bildungsgeschichte Schweiz", Switzerland).

In Portugal, Repositório Digital da História da Educação[234], supported by Portuguese Ministry of Education, holds digitalized sources on legislation, textbooks, reflections on education, curricular programs as well as periodicals related to education and schooling as well as a virtual museum containing other collections.

Non- or partially digitized collections

Many institutions are unlikely to digitize their collections for some reason or other, often because of lack of funding, while others have only partially digitized collections (where the digitization activity may have ended because it was only project funded, or funding ran out for some other reason). These can be roughly divided into larger and small institutions.

Large research archives/libraries

Many of these are unlikely to be digitized, at least within a foreseeable future. They may also be unlikely to want to hand over content to Europeana Research. A short-term fix, according to some researchers who work a great deal with archival material, which would also be very helpful, would be to include metadata from such institutions in Europeana Research. For archives, these often exist in the form of archival descriptions, which describe in some degree of detail what can be found in an archive. Incorporating archival descriptions in Europeana Research would greatly facilitate finding related material in various archives.

Smaller institutions (such as small school museums)

Europe is full of minor school museums (our category 4) whose collections are generally not (professionally) digitized. There is no consensus among researchers about their usefulness, but at least some researchers are of the opinion that their collections would prove valuable for research if only they were more easily accessible. By digitizing such collections and adding to a European research infrastructure, new possibilities for school-history scholarship would open up. The problems here are, of course, funding and selection: how to decide what to digitize and how to pay for it.

---

[234] http://193.137.22.223/pt/patrimonio-educativo/repositorio-digital-da-historia-da-educacao/

But even larger museums have problems with digitization projects, and may thus effectively escape the attention of researchers. An example is the Museum of Education in Ypres, Belgium, a national museum (our category 1) which does not have digitized collections, due to lack of available funding. Their primary focus is thus on attracting and informing visitors (primarily school children) rather than provide a resource for scholarship.

Institutions with partially digitized collections

These are possibly worth encouraging to contribute material, on the principle that some accessible digital material is better than none. However, it is important to recognise that this could be problematic, often for the same reasons that digitization is incomplete, that is to say, insufficient funding, resources or prioritization. School archives, for example, can have a great deal of visual material, especially schools with related extracurricular activities involving photography. Some of the material is digitized and even published on the schools' websites (for example, Roja secondary school of Latvia[235]).

Discussion

The ultimate Europeana Research resource for research into the school history of the twentieth century is if not impossible then at least unattainable for many years to come. Researchers in the field come from many different disciplinary backgrounds, and use a plethora of sources. Many of these sources cannot be added to a research database today because of copyright, personal information protection, ownership issues, or simply because the material is not digitized and there is no funding for digitization available.

That is not to say that Europeana Research could not become a valuable resource for such research. A large number of researchers have indicated that the kind of material that we identified in our preliminary exploration of Europeana would be interesting and relevant for research, and there is every reason to bring such content together. And there is no reason *not* to develop and add to that content.

Any future development of content for school-history research requires consideration of what the metadata for material need to look like, and what material should be added. Ideally, such considerations are carried out in collaboration with active researchers in the field, as was pointed out in one interview.

---

[235] http://rvskvesturesfoto.wordpress.com/

During the Expert Forums, researchers pointed out the need for improved metadata if the Europeana material were to be useful in scholarship. Regardless of what content is included for school history, some metadata requirements must be met. One piece of information which is currently unavailable in Europeana is selection criteria. Other metadata, such as date (or at least year) and place, are also essential. Place does not necessarily mean geographic location or country: town or even school name (or number) could make a difference between a photograph being valuable or useless.

Context is brought up in one way or another by every researcher that we have interviewed. To some extent, such context can be provided by metadata, but additions of national education policy material and links to previous research would significantly improve the value of the content. Taking into account the content added to Europeana Research under other topics would also increase the total value for researchers in school history: parliamentary papers, political propaganda, and children's literature are all areas that are of interest here.

Current content available through Europeana comes with some drawbacks. Foremost is the fact that there are not distinct datasets that have clear benefits for school-history research. To identify useful material, searches with keywords and phrases need to be carried out, and not all hits will be valuable – some searches have too low recall, others are too imprecise. Another drawback with that the Europeana material is that different countries are either more or less well-represented, something which one researcher noted would hamper comparative studies.

There are a number of sources that can provide more material. Identifying and incorporating those collections which are already digitized (and encourage further digitization in cases where only parts of collections have been digitized) would be the easiest way to proceed. But in a number of cases – for instance major collections that are already research portals in their own right; public archives that will not or cannot digitize their collections, or that for some reason (legal, financial or otherwise) will not release their digitized material; libraries where there are copyright issues – the best way forward would possibly be to incorporate metadata rather than content. Finding a record of a textbook from Braunschweig and a link or an archival description for an archive in Helsinki which contains a useful document would not be as convenient as finding the digital copy, but it is better than not finding the objects at all.

Completion is a point that was raised again and again. Researchers must know that what they find is a complete collection, or, if not, how to get hold of the entire collection. The research

process requires any sampling done to be the researcher's own. But the selection of what content to incorporate in the future would also benefit from active researchers' assistance. In future development of content for school-history research, further engagement with the scholarly community would be highly recommended.

6.6 The Ancient Mediterranean

The Ancient Mediterranean is a quite broad topic, pertaining to the study and interpretation of the remains, findings and written record of the Mediterranean region in the Antiquity. Geographically, this topic covers the areas surrounding the Mediterranean basin, which of course includes some of the major ancient civilizations, such as Greece, Rome and Egypt.

As far as chronology is concerned, we have opted to extend the periods under examination from Prehistory to Late Antiquity in order to cover as many relevant Europeana collections as possible.

This is a thematic area which could correspond to a number of academic disciplines, such as Archaeology, Ancient History, the Classics, Epigraphy, Anthropology, and History of Art. offers a significant added value by a large-scale ripple effect of implications across the disciplines, professions and processes related to archaeological knowledge work from archaeological scholarship to scholarship on archaeology, historical studies, climate science, societal planning, tourism and heritage management to mention a few examples.

In July 2015, Europeana Research invited Archaeologists and scholars working on the Antiquity (Ancient Historians, Classicists), in order to:

- discuss and evaluate available tools and content in the area;

- assess Europeana content in the area;

- provide insights to user requirements with particular emphasis on digital humanities methodologies;

- advise on feasibility of Europeana working in that area with a focus on issues concerning the creation and use of related datasets;

- indicate further content to be potentially aggregated by Europeana in that area.

To this end, the workshop focused amongst other areas, on tools and services available to archaeologists and classicists as well as on applications of visualization technologies in archaeology and related datasets.

Europeana collections include text resources, images, audio, video and 3D relevant to the research topic "Ancient Mediterranean". Especially the datasets contributed by providers such as 3D Icons and CARARE as well as some LoCloud datasets contain a variety of relevant resources. 3D Icons provides digital representations of architectural and archaeological monuments and buildings in 3D through the CARARE aggregation service. The CARARE Best Practice Network, which ran from 2010 until 2013 was also related to architectural and archaeological content and established a network of institutions in order to support researchers, archaeological field units, site museums and local institutions to make their content available. The project had over 2 million items ingested by Europeana by 2013. Additionally, the pilot map-based search interface provided by CARARE (see Figure 36) can serve as a useful tool for users interested in this research topic, since the content to which it is related is relevant and the geographical region within which one might want to perform a search can be selected in the map.
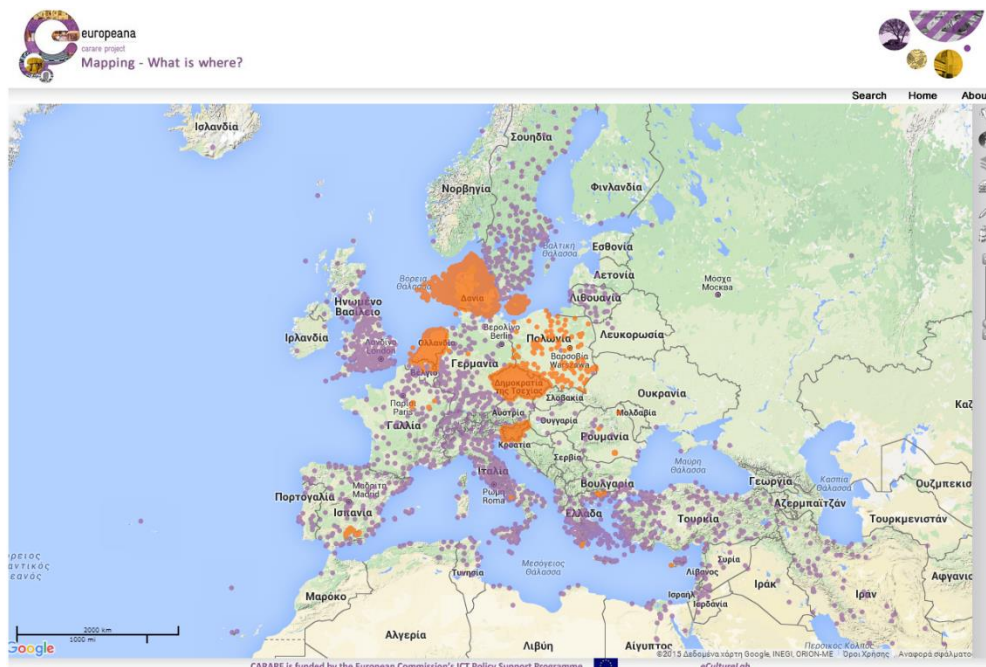
Figure 36. Snapshot of the pilot map-based search interface provided by CARARE

These resources can be accessed in the Europeana portal through various searches related to the thematic area and the specific topic of interest explored. As the Ancient Mediterranean is a broad topic covering diverse geographical sub-areas, disciplines and periods of time, the related Europeana datasets can be identified through varied searches and queries. Searching the Europeana resources through Boolean queries containing the topic under consideration as well as the academic discipline or the chronological span relevant to the study has been proven useful in order to locate specific collections. On the other hand, Boolean queries can sometimes limit considerably the search results, leaving related resources out of the search results.

The examples provided below illustrate the path that a user interested on a specific topic related to the Ancient Mediterranean can follow in order to identify relevant collections within Europeana. For instance, a historian interested in resources about temples in ancient Greece can initiate his/her search by entering a Boolean query including the terms "Greece" and "Temple". In this case the results are very rich (755 items) and a variety of resources is available to him/her: text, image, 3D, sound and video. The results come from numerous providers, namely the Hellenic Aggregator at Veria Public Library, Linked Heritage, the Judaica Europeana, Europeana Photography, the European Library, Athena, 3D Icons, CultureGrid, Universitätsbibliothek Heidelberg, Athena Plus, LoCloud, CARARE, Médiathèque Valais - Martigny, DISMARC, Europeana Local Austria, Institut National de l'Audiovisuel, DM2E, EUscreen Project, Europhoto, Europeana Travel and Federacja Bibliotek Cyfrowych. Moreover, the results include various description languages such as English, Greek, French, German, Catalan, Italian, Serbian, Magyar, Polish, Romanian and Swedish and come from many different countries, namely Greece, Ireland, the United Kingdom, France, Germany, Spain, Austria, Switzerland, Israel, Italy, Cyprus, Belgium, Serbia, Hungary, the Netherlands, Poland, Romania, and Sweden. The image below presents some of the resources that can be found in the Europeana collections following this query. The resources include images of temples, images of artefacts related to temples, historical texts, 3D representations, videos related to temples and audio recordings. Similar broad searches have been performed with a combination of query criteria such as "Greek" and "temple", with homologous results.
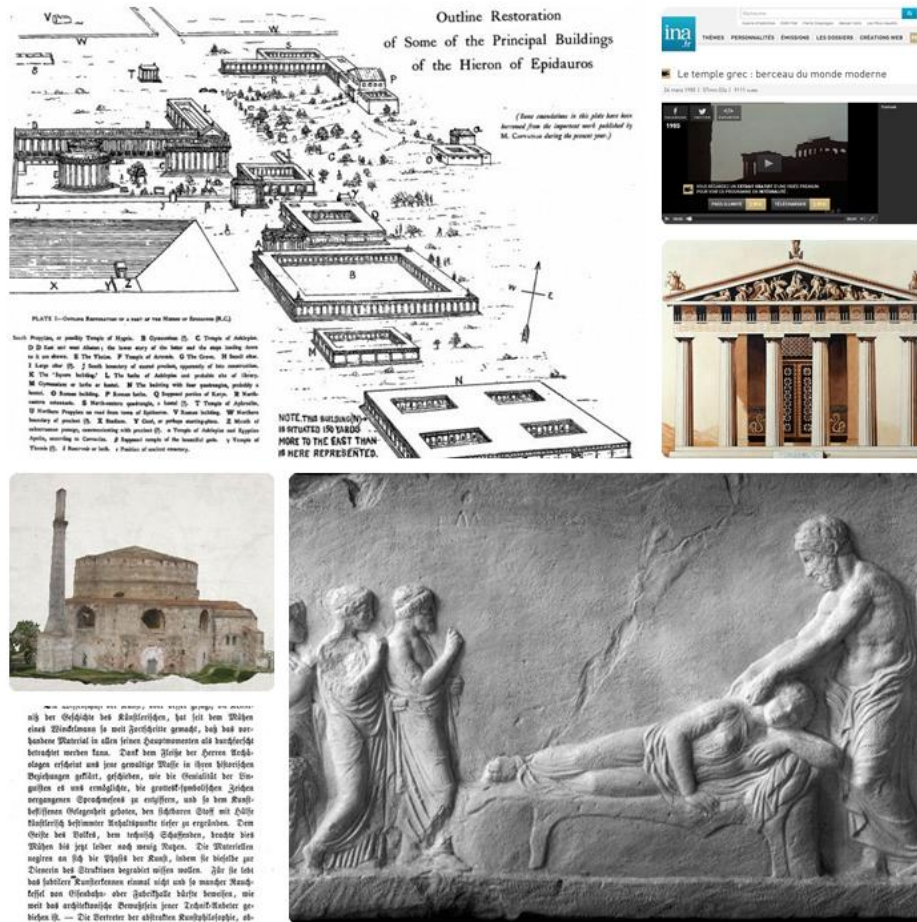
Figure 37. Representative items for the search results of the Boolean query "Greece" and "Temple"

Since these results include Christian and other temples from periods of time exceeding the late Antiquity, a refining of the results can be achieved with the addition of the criterion "ancient" in the query. In this case, however, the results are very limited (72 resources). Additionally, some resources which were brought back by the previous queries and belong to the designated time period do not appear under this search and a more specific research must be made in order to access them. For example the fragment of a bas-relief in the form of a temple showcased in the image above belongs to the timespan covered under the term "ancient" but does not appear in the relevant search results.

Furthermore, a parallel search in the CARARE pilot map-based search interface was performed, which however showcases results coming only from the collections ingested through CARARE.
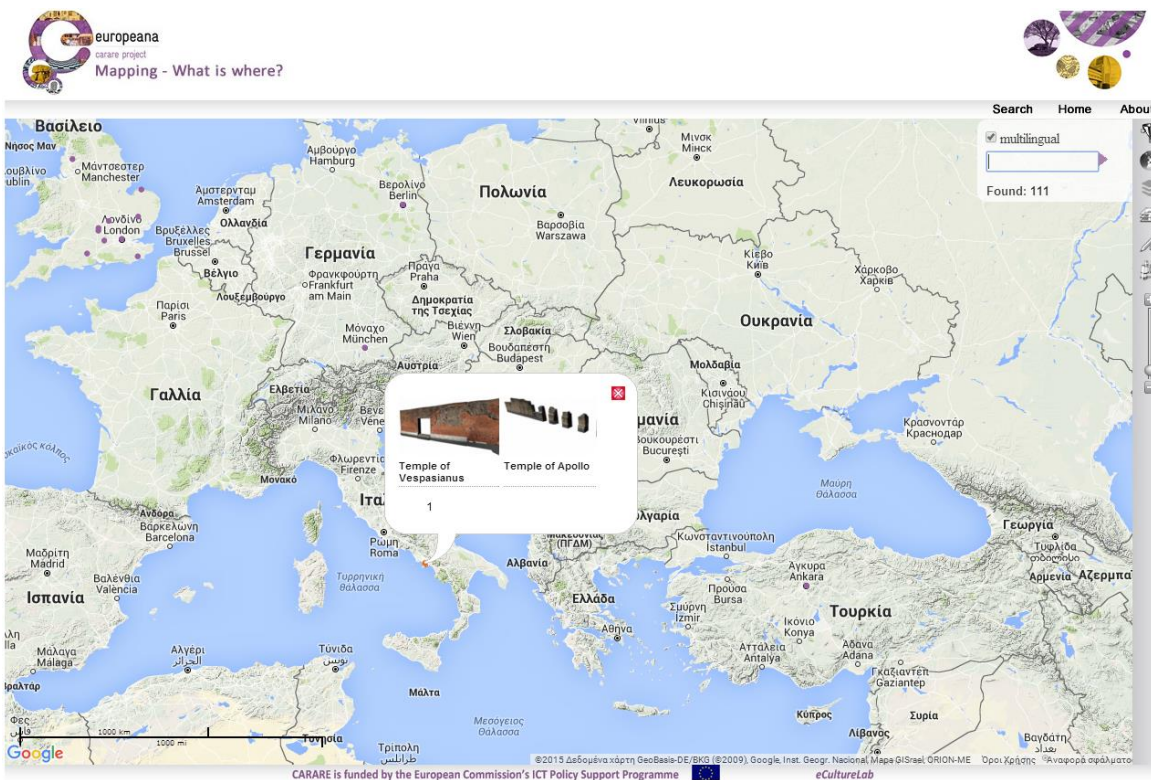
Figure 38. Results of the CARARE map for the query "Temple"

The searches performed on the Europeana portal and on the CARARE map on the relatively broad subject of ancient Greek temples turned a number of results in various formats but failed to provide the set of objects that one would expect to get through such a search. For instance, the results of the search "Greek" and "Temple" and "Ancient" do not include any items related to the Athens Parthenon, probably the most famous ancient Greek temple. Yet the Europeana collections include 2751 items related to the search term "Parthenon", which however are not accessible through the searches described up to here.

The second example, thus, focuses on a specific term around which the search will revolve, namely the Parthenon. As mentioned above the search results for the term "Parthenon" alone return 2751 results including images, texts, sounds and video which represent a good starting point for discovering assets as well as for refining the search results. Some of the results returned by this search are showcased in the Figure 39.
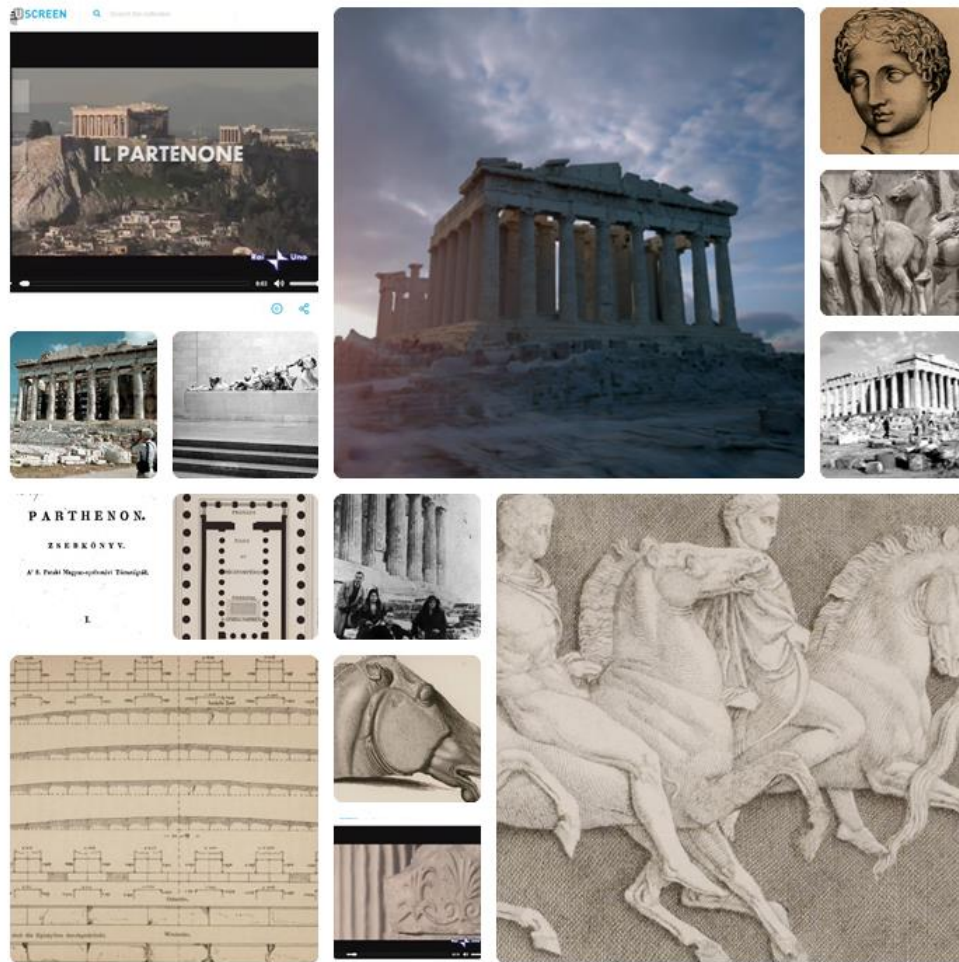
Figure 39. Representative items for the search results of the simple query "Parthenon"

Because this search is much narrower and contains a very specific term, it was found that it is easier to make further filtering questions based on it. For instance a Boolean query including the terms "Parthenon" and "Athens" and "Column" or another one with the terms "Parthenon" and "Athens" and "Sculpture" return results in close connection with the queries, as showcased below (Figure 40).
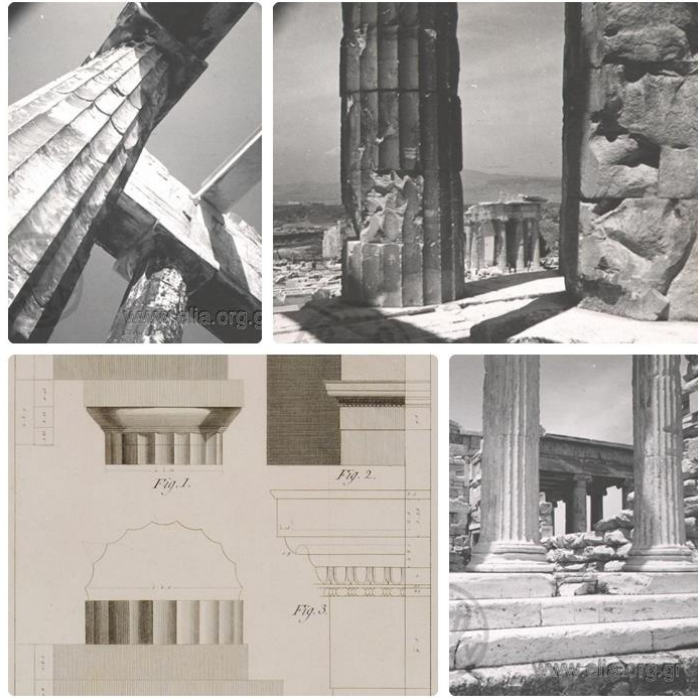
Figure 40. Results for a Boolean query with the terms "Parthenon" and "Athens" and "Columns"

Figure 41. Results for a Boolean query with the terms "Parthenon" and "Athens" and "Sculpture"

On the whole, searching through the Europeana collections might be a challenge for researchers interested in broad topics related to the Ancient Mediterranean which cannot be uniquely identified, like in the case of ancient Greek temples. On the other hand when looking for specific monuments or artefacts or geographical locations the search results can be very useful and can be further filtered. The main problem that has to do with further filtering which has been identified through this process is that when narrowing the terms some items that exist within the previous searches are detached, although the search terms should be sufficient in order to be included in the results.

As far as other sources of data are concerned, since the topic is a very popular one and covers many geographical regions and many time periods, a significant variety of digital resources is available from other providers as well. Firstly, most European national libraries have noteworthy datasets associated to their collections. Many of those are digitised and accessible online. For instance, Gallica, which is a digital repository linked to the general catalogue of the French national library and is as well a partner of Europeana, includes numerous digitised resources on the ancient Mediterranean divided in categories such as books, gazetteers, objects, images, manuscripts, cards, audio and scores. Figure 42 showcases some objects related to ancient Athens.
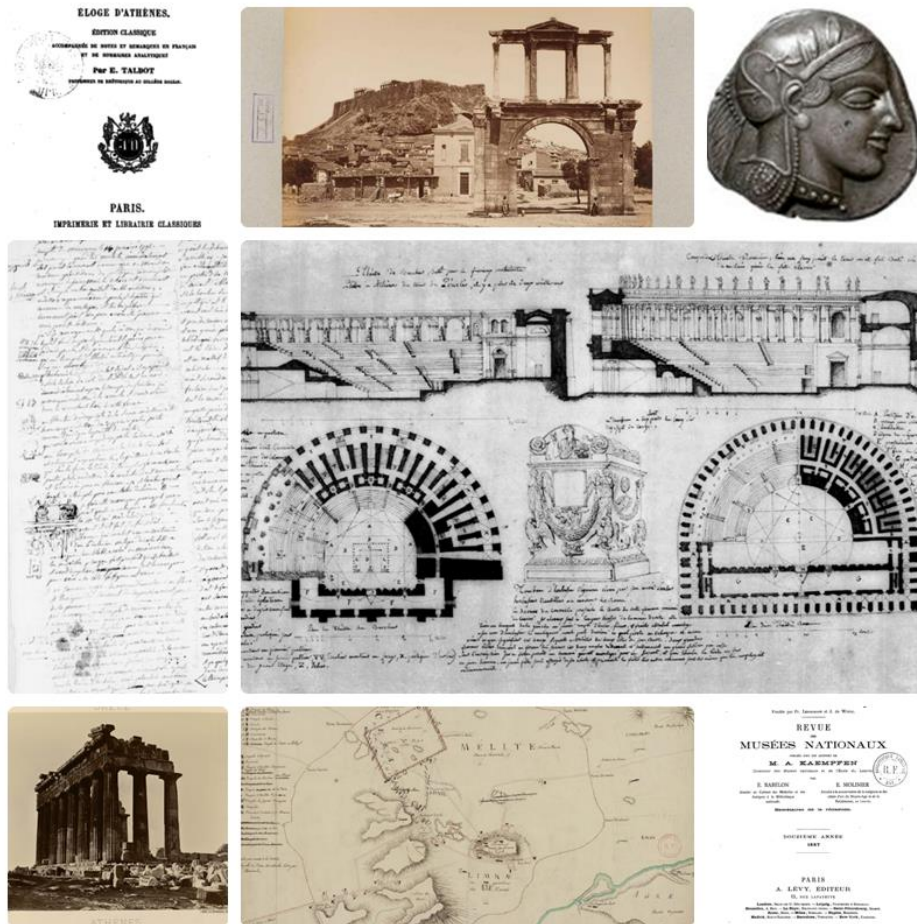
Figure 42. Resources related to "Athens" in Gallica

Other collections include thematic digitised content provided by projects such ARIADNE, a project funded by the European Commission under the Community's Seventh Framework Programme, which brings together various distributed datasets from existing archaeological infrastructures targeting at harmonizing archaeological data and allowing homogenous access to it. ARIADNE provides its online services giving integrated access to several archaeological datasets through the portals of ADS, FastiOnline, ARACHNE and ZENON DAI.

Pelagios, which stands for "Enable Linked Ancient Geodata in Open Systems" provides linked open data related to ancient Greece and Rome and is being extended with data annotations covering Christian, Islamic and Asian Geography. Pelagios' search is based on a map of the Mediterranean where one can search for a geographical region and find related datasets and linked data. For instance the map of Pelagios leads to 45368 references for "Athenae". These references are contained in 17 different datasets from various providers such as Arachne,

SquinchPix, CLAROS, Pleiades Annotations in the Perseus Digital Library etc. This geographical aggregation of datasets permits the user to identify data related to a specific geographical region. Figure 43 is a snapshot from the map of Pelagios.
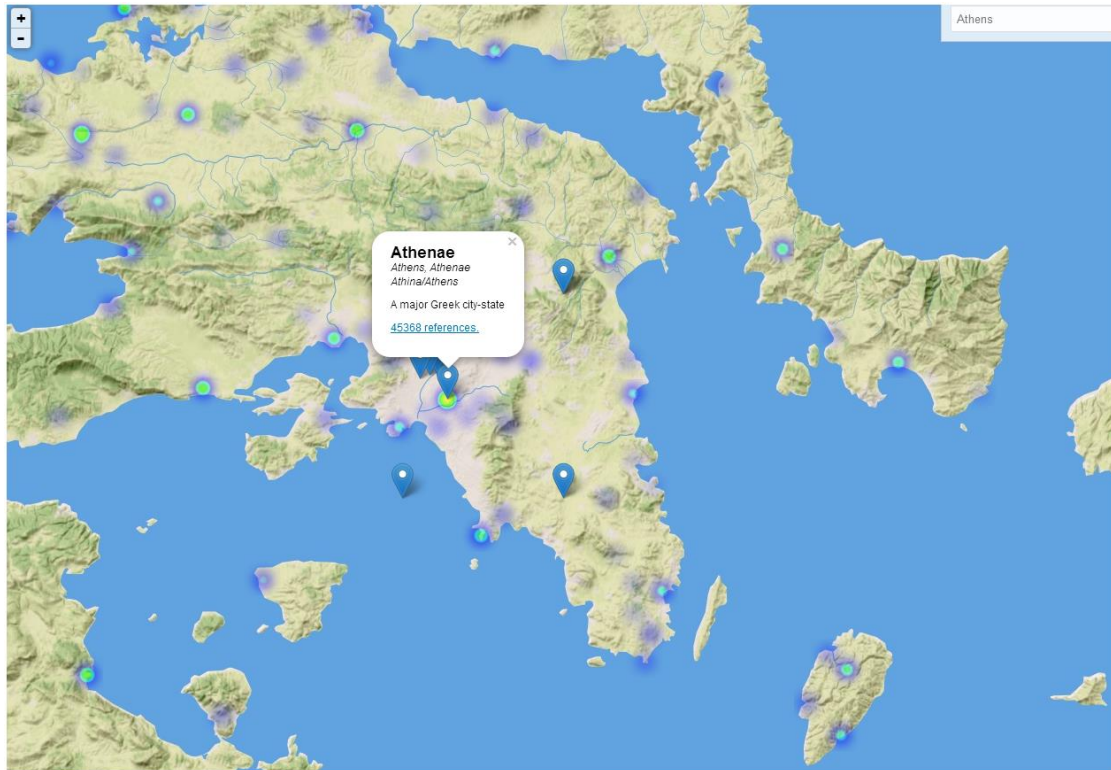
Figure 43. The map of Pelagios showing the number of linked data referring to Athens

Pleiades, also bases its search functions on the geographical information and covers the ancient Greek and Roman world while being extended also to the Ancient Near Eastern, Byzantine, Celtic, and Early Medieval geography. Places and locations are searchable through a map which supports ancient terrain, modern terrain, modern streets and the Roman empire. The catalog of Pleiades objects can be searched via simple or advanced forms and all published content is accessible under open license.

Perseus Digital Library is a project running since 1985 and focuses on ancient Greece and Rome, while it is planned to include also other topics such as early modern English, the American civil war etc. Perseus support multilingual search amongst its collections and a suite of search tools including word and phrase search. The primary and secondary sources for the study of ancient Greece and Rome included in the Perseus Digital Library are accessible as a

list or can be searched through the Perseus Catalogue where they can also be browsed by author, work title, work original language, edition or translation year published, edition or translation language, series and subjects. A general search in the Perseus Catalogue for the term "Athens" returned 540 results each of which is related to one resource and leads to the location where a digitized copy of the resource can be found. Figure 44 showcases the path in order to locate the comedy "Clouds" of Aristophanes.
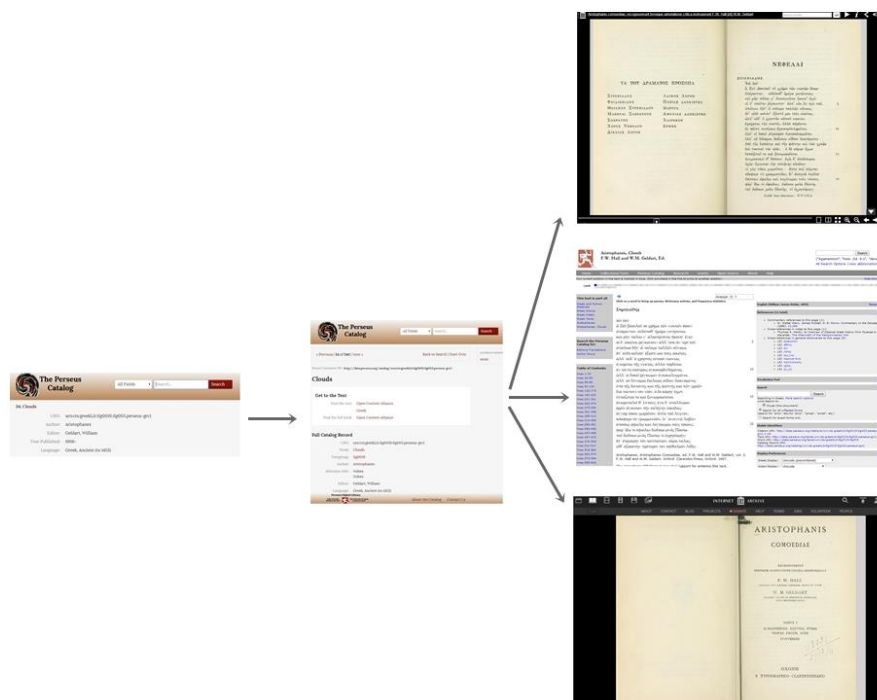
Figure 44. Perseus Digital Library - Path to the Clouds of Aristophanes through the Perseus Catalogue

Empirical research in the context of the EU Project *Preparing DARIAH* showcased the importance of annotation for scholarly work. It thus came as no surprise that, a few years later annotation tools are still considered to be a major need amongst researchers studying the Ancient Mediterranean. Textual and image annotation is one of the basic means of organizing and interpreting digital substitutes of findings and secondary sources.

Evidence-based data about the services and tools used by researchers working on the topic of Ancient Mediterranean and more specifically Archaeologists, Historians and Classicists can also be drawn from two relevant web surveys conducted amongst Humanists: the Europeana Cloud Research Community Web Survey and the DARIAH-EU Web Survey. Both surveys seek to describe and elucidate the scholarly practices and needs of researchers in the Arts, Humanities

and Social Sciences and they both include questions related to the use of digital services and tools. In the context of the Europeana Cloud Research Community Web Survey data about the frequency of use of specific services is available, as well as a list of services proposed by the respondents themselves. The DARIAH-EU Web Survey on Digital Practices in the Arts and Humanities, on the other hand, provides data about the use of digital services and tools for research purposes. For the purposes of this account, a cluster of the total sample of each of the aforementioned surveys has been used, in order to provide a discipline-specific account. Therefore the following data derived from the two web surveys refer solely to the respondents who stated that their discipline is either Archaeology, History or Classics.

The data provided by the Europeana Cloud web survey about the use of specific services suggest that Archaeologists, Historians and Classicists use very often services such as online library catalogues and specific online digital archives. The use of commercial portals of scholarly journals and of Google Scholar is less frequent, while they seldom use services such as museum websites, the Europeana Portal and the European Library (see Figure 45).
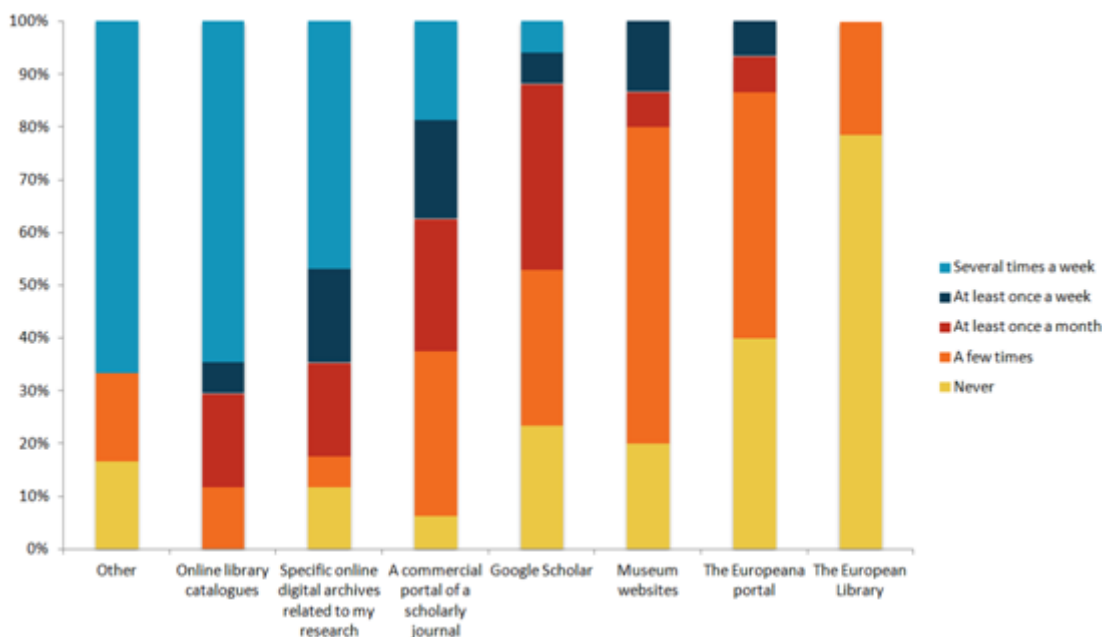
Figure 45. Europeana Cloud Research Community Web Survey, Sample of Archaeologists, Historians and Classicists, Frequency of use of services. N=30.

The findings of the DARIAH-EU Web Survey suggest that Archaeologists, Historians and Classicists mainly use web search engines and online library catalogues, while they also use

often digital archives, digital collections and data repositories. The data suggest that the use of services such as online scholarly journals and search engines of research publications is less frequent, while the use of social media for research purposes is not at all performed by almost half of the respondents (see Figure 46).
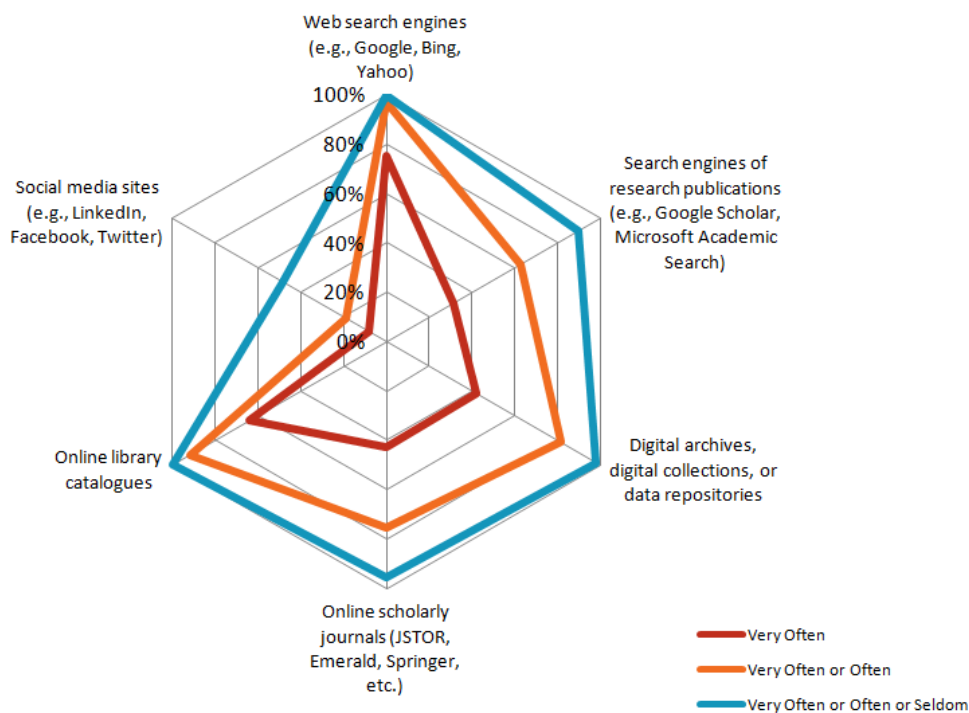
Figure 46. DARIAH-EU Web Survey, Sample of Archaeologists, Historians and Classicists, Frequency of use of services. N=610.

In a subsequent question of the same study, the respondents were invited to choose from a list of tools and services which ones they use to store and manage their research assets. According to their answers, almost all Archaeologists, Historians and Classicists use a word processor, 65,5% use a spreadsheet application, 44,4% use a database management system, 41,6% use some non-digital method to store and manage their research assets, 30,6% use a note-taking application and only 15,5% use a web-based content management system.
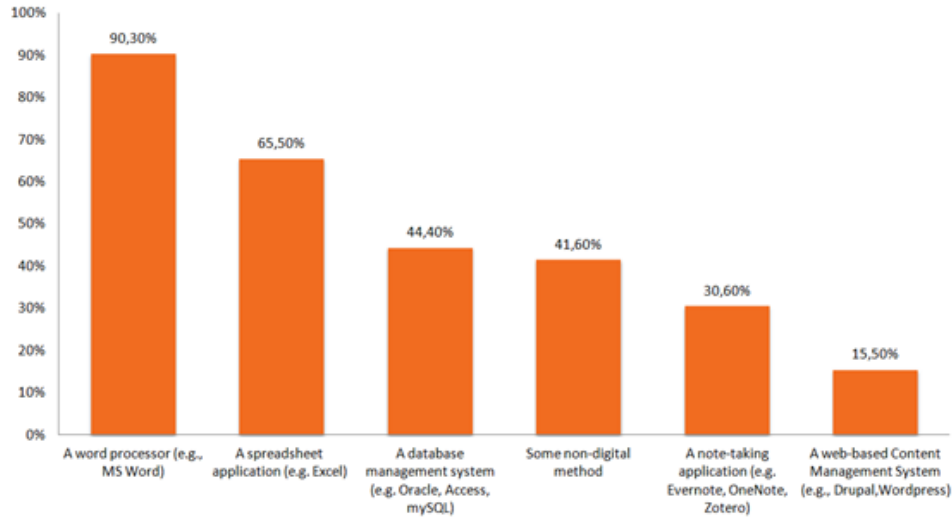
Figure 47. DARIAH-EU Web Survey, Sample of Archaeologists, Historians and Classicists,Use of applications to store and manage research assets. N=610.

The answers provided in the option "other" include databases, CAD programmes, archiving applications such as Dropbox or Google Drive, bibliographic management applications such as Endnote or Mendeley, GIS programmes, statistics software, word processing software such as Latex, applications such as Github. Less frequent solutions for storing and managing research resources include the use of one's own website or the use of a smartphone for storing notes.

7. User Requirements and Content Strategy Recommendations for Europeana Research

In the following section, user requirements and content recommendations for Europeana Research are broken down into those two main categories, often overlapping. All categories and specifics are considered required, although the report authors do not attempt in this document to define the tests for achieving a requirement or, for that matter, a recommendation. This document was not designed to be a development of functional specifications. The Project should analyze and evaluate both the individual features and the environment overall potential as a coherent service of the environment, including growth and sustainability potential as well as systemic weaknesses or limits.

- **Research Communities Web Survey**

Even though majority of the questions asked were related to the importance of different aspects in already existing practices, they clearly indicate the preferred routines and thus also requirements and expectations in the research community towards any new services.

SEARCHING / DISCOVERING RELEVANT RESEARCH MATERIAL
- ➤ Queries, comprising either specific terms or a combination of terms as well as finding resources through footnotes in articles and/or books and using the results of a search in order to conduct a new search, are judged to be of high importance in scholarly work of the research community;
- ➤ Browsing related resources and consulting collection inventories, indexes and summaries are also rated as important while activities such as asking peers or colleagues, finding relevant information on the basis of resources one just happens to stumble upon and asking reference librarians, curators or archivists seem to be less central in the process of seeking information.
- ➤ Most typical queries used by the respondents in relevant research communities were thematic, looking for a specific subject, period of time or geographical region and are phrased affirmatively; queries might also look for specific type or group of objects, such as "GIS" or persons; can be phrased in the form of questions, such as "what is the etymology of word X?"; or as Boolean queries.

USER REQUIREMENTS RELATED TO METADATA

> **Resource characteristics**. Respondents **prioritize aspects of the resources that are very specific, such as the dates and time spans, places and events related to a resource**, while characteristics which are indirectly or more loosely related to a resource, such as the kinds of objects, artefacts and art works, the classifications of places mentioned or represented or the kinds of people mentioned or represented are considered less important.

> **Specific properties of a resource**. The data suggests that the name of the author, creator or issuing authority of a resource is by far considered to be the most important property of a resource, followed by the fonds, collection or holding institution to which the resource belongs and the format of a resource. The rest of the properties under consideration are considered to be relatively less important, although none of them is judged to be not important at all by more than 20% of the respondents.

USER REQUIREMENTS: ORGANISATION OF UNPUBLISHED MATERIALS:

> Filing together **digital copies of resources** is considered to be very important and extremely important by most respondents;

> **Collecting and keeping references** to all resources of interest and **collecting lists of references to primary resources** in either digital or analogue form are "very important" or "extremely important" to most of respondents;

> Software application to organize resources, manually copying or retyping the content of resources and filing together photocopies of resources are also considered as relatively important, while the other activities - keeping a list of keywords for references to primary material, maintaining a manual index of keywords on resources and using catalogue cards to manage resources are considered to be of somewhat less importance.

USER REQUIREMENTS: ANNOTATING SCHOLARLY INFORMATION

> Underlining or highlighting relevant passages of texts and scanning texts quickly to identify relevant passages to be the most important activities,

> Exporting and keeping relevant passages and notes within the word processing file of a research draft, keeping margin notes on passages of texts and together with passages of texts, are also considered important relatively often.

USER REQUIREMENTS: COOPERATION & COLLABORATION

> Sharing copies or resources with colleagues, asking colleagues for their expert opinion and collaborating on joint publications or conference papers are considered most important.

> Researchers would be interested to know which scholars in their field work on a particular source or research question, and would be prepared to share information with colleagues on the sources or research questions they work.

USER REQUIREMENTS: RETRIEVAL OF RESOURCES

> Fetching and automatically importing bibliographic references into bibliography applications was most often considered as "very" or "extremely" important,

> Functionality of fetching and automatically importing elements like search results, online digital resources or metadata into a word processing file or research database / repository was also quite often considered as important (see also important activities in annotating scholarly information);

> Receiving automatic notifications on the existence of online digital metadata and resources of interest and fetching and automatically importing online digital resources (e.g. images) into a research repository or database have also been considered as important quite often;

> Fetching and automatically importing online digital resources (e.g. images) into a notetaking program, fetching and automatically importing search results on online digital resources into Excel and fetching and automatically importing metadata about online digital resources into Excel were considered to be less important.

- **Humanities and Social Sciences Case Studies**

> Provide import and export functionality for a wide variety of formats

> Functionality that enables addition, import, export, and collaboration metadata for Europeana records should be flexible and easy to use

> Functions for mark-up, organization, and collaboration would be highly useful

> Provide basic analytic functionality and the possibility of documenting research decisions

> Support for (the creation of) multilingual data and metadata would greatly increase the material's usefulness.

- **Parliamentary Papers**

➢ Key for Digital Humanities researchers: material available in full-text

➢ A portal/resource that brings together various types of Parliamentary Papers from a wide array of countries

➢ Enriched full-text would enable researchers to pose new, transnational questions

- **Population Displacement Research**

User requirements regarding tools and services in accessing, processing and analysing data:

➢ The **accessing data** could be most efficiently done **via an API** - however this would be considered quite advanced technical knowledge and might not be so common among the typical Humanist research community;

➢ For **data processing** (?), Pinterest would be useful if there were more powerful features; Story Map Journal has also been useful; and Culture Collage has potential with this also, although it is not such a powerful tool at this stage.

➢ **An image annotation** tool such as HyperImage (for example) would be very useful in order to **create links between images** and **keep notes alongside the image** analysed;

➢ A Zotero-like tool would also be useful **for cataloguing the content** found and **creating links between the content**; if this could then be **downloaded as an index** that would **enable to find content and references more easily**;

➢ Being able to **use reference software alongside Europeana content with all annotations made on the Europeana Research Platform** also included would be highly beneficial to researchers in all sectors (not just in social history).

- **Political Propaganda**

➢ Material from former European colonies could be useful (but may be beyond the scope of Europeana's ambitions)

➢ A sustained and substantially more appreciable involvement of researchers is essential in order to establish the validity of "Political Propaganda" as a developing module for content aggregation.

➢ Editorial involvement of specialists is required to monitor progress and establish priorities in the content acquisition phase. (A point relevant for School History as well)

- **Children's Literature**

Many of these are not necessarily useful for researchers but other kinds of users. Occasionally hard to tell apart.

- ➢ Possibility
    - o for users to edit and enrich resources; clear information about provenance of new editions
    - o to build network of users/researchers
    - o for users to share knowledge and publish views
- ➢ A platform/environment which is simple to use
- ➢ Annotation tools
- ➢ Digital illustration tools
- ➢ More digitized content/production of digitized content
- ➢ Collections organized thematically or by content could serve as ready-made research material
- ➢ Online access to multilingual material
- ➢ Constantly updated record of publications (on specific research topics) to serve researchers to find what is new in the field
- ➢ Map of publications which could strengthen networking in the field

- **European School History in the 20th Century**

- ➢ Sources
    - o Identify and incorporate already digitized collections (and encourage further digitization in cases where only parts of collections have been digitized)
    - o Where appropriate incorporate metadata (e.g. from major collections that are already research portals or public archives) rather than content
    - o Researchers must know that what they find is a complete collection, or, if not, how to get hold of the entire collection
    - o Digital material can only take historians so far; for most material, they need to see the original document.
- ➢ Metadata needs to be improved if Europeana material is to be useful in scholarship

- Certain metadata requirements must be met: date (or at least year) and place, selection criteria (not currently available in Europeana)
- Also useful: refinement of keywords, context and terms of use

➢ National education policy material and links to previous research would significantly improve the value of the content

➢ Better representation of different countries in Europeana and greater flexibility in terms of categorisation according to current nation state boundaries would assist comparative studies

➢ In future development of content for school-history research, further engagement with the scholarly community highly recommended.

- **Ancient Mediterranean**

➢ Annotation tools are considered to be a major need amongst researchers studying the Ancient Mediterranean. Textual and image annotation is one of the basic means of organizing and interpreting digital substitutes of findings and secondary sources.

- **Recommendations for Europeana Research Content Strategy**

This deliverable has laid out many of the weaknesses and strengths of the current Europeana offering, and requirements of the research community in terms of re-use of tools and content. Given these user requirements, the following recommendations are made for a Europeana Research Content Strategy

1. Given that Europeana datasets has multiple ingested collections that provide usefulness for humanities research, especially when openly licenced, more should be made of these collections.
- **Collections of metadata should be highlighted on Europeana Research**. They should be made available via the portal, but also via the API, so that digital humanities scholars can access the the underlying metadata in bulk (export metadata ranks highly in the user requirements) It would also be useful if tools could be developed to create collections of content (held at the source cultural heritage institution) based on metadata records assembled by a user
- Collections should be as meaningful as possible for researchers. Many datasets ingested by Europeana are currently collections of data with no coherent strategy for

their alignment. In presenting collections on Europeana Research, each set of items should form a coherent whole.

2. Europeana Research also requires focussed accumulation of content to create in depth corpora for researchers to reuse. The work on research themes in section 6 highlights possible topics where Europeana may wish to concentrate collections building.

- Europeana as a whole needs to develop more coherent collections of metadata and related content to provide focus, especially for channels, but also to give collection strength within the Europeana Collections (the new version of the portal) and Europeana Labs
- Therefore future aggregation work for Content Strategy as a whole should be closely aligned with Research. However, they should not overlap - researchers require much more focussed collections than the broader channels.

Bibliography and References

Appendix A -List of Surveys

Appendix /// - Search terms and informants

Europeana Research - Implementing Collection Descriptions

A first step of the Europeana Research Content Strategy is to publish collection descriptions of data available in Europeana. The Europeana Collections website (ie the portal) allows searching for individual items, but the multiple records returned often hide the rich seams of collections that are available. Taking the approach of describing groups of items that have been delivered to Europeana as coherent collections will help increase knowledge and understanding of what is available.

Additionally, providing information on collections also allows more advanced users to exploit the API, and return information on the content related to the data.

It is suggested that Europeana Labs and Europeana Research follow a standard workflow for identifying collections that can be made available on each of the two sites. The collections

should then be described according to the needs of their respective audiences - Labs for the creative industries and Research for the research community.

While Labs follows a strict policy of highlighting collections that are openly available (ie with Public Domain, CC-BY or CC-BY-SA), Europeana Research could take a broader viewpoint including CC-BY-NC-SA and even Rights Reserved collections. Rights Reserved collections would not be open for content harvesting (ie users could not use the API to harvest the content at the source institution) but the relevant collection description would still play a useful role in highlighting collections available in the Europeana portal.

To describe collections on Europeana Research, the following fields will be used

1. (Edited) Title
a. This should be a clear title that describes the collection. Amendments can be made to ensure understanding. It should be a maximum of 100 characters

Description

. The description can vary according to the time the cataloguer has. Ideally each description will be a detailed review of the contents, highlights and provenance of the collection, and identifying strengths and weaknesses from a scholarly point of view. However, this is not feasible for every collection, but a description of the content is obligatory. This should be at least 100 words.

Item Type

. There is a controlled list of the main content types featured. This is restricted so as to restrict the number of available choices on the main browsing page of Europeana Research

a. Terms: Manuscripts, Books, Posters, Photographs, Prints, Letters & Diaries, Pamphlets. Drawings, Newspapers, Maps, Paintings, Sheet Music, Documents, Audio. Film / Video. Others are added as free tags

Spatial Coverage

There is a controlled list of the main content types featured. This is restricted so as to restrict the number of available choices on the main browsing page of Europeana Research

a. Terms: Northern Europe, Southern Europe, Western Europe, Eastern Europe, Middle East, Asia, Latin America, North America, Oceania. Which country fits into which specific category is defined by the [EuroVoc vocabulary](#)

Temporal Coverage

There is a controlled list of the main content types featured. This is restricted so as to restrict the number of available choices on the main browsing page of Europeana Research.

a. (Terms: Ancient World, Medieval, Early Modern, 18th century, 19th century, 1900-1945, 1945 onwards, Multiple Eras)

Link to Europeana portal

This links to the items on the Europeana portal

Link to Europeana API console (not obligatory)

This links to the items on the Europeana portal

Name of data provider / Source

. This is name of the original cultural institution that provided the collection (not the aggregator but the GLAM)

Image

. An image taken from the one of the items in the collection. Ideally, it should be accompanied by a caption but we do not have that facility yet

Link to Rights Statement

This must be one of the Europeana Right Statements

Link to original source

This should be a link to the same collection at the original cultural institution

References

Literature review and related work

Anderson, S., and Blanke, T. (2012). Taking the Long-View: From e-Science Humanities to Humanities Digital Ecosystems. *Historical Social Research, 37*(3), pp. 147-164.

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

Benardou, A., Constantopoulos, P., and Dallas, C. (2013). An approach to analysing working practices of research communities in the Humanities. *International Journal of Humanities and Arts Computing*, *7*(1-2), Edinburgh University Press, pp. 105-127.

Bia, A. (2012). The Digital Humanities Workbench. *Proceedings of the 13th International Conference on Interacción Persona-Ordenador.* New York, NY, USA: ACM, 50:1–50:4, doi:10.1145/2379636.2379685.

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

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

Council of European Social Science Data Archives – CESSDA. (2012). Retrieved from http://www.cessda.org/

Distributed Digital Music Archives and Libraries Lab (n.d.). Retrieved from http://ddmal.music.mcgill.ca/research/salami

Education for Change Ltd, SIRU (2002). *Researchers' use of libraries and other information sources: current patterns and future trends*, The Research Partnership, University of Brighton, pp. 1-84. Retrieved from www.rslg.ac.uk/**research**/lib**use**/LUrep1.pdf

Electronic Archiving System of Data Archiving and Networked Services – EASY. (2013). Retrieved from https://easy.dans.knaw.nl//ui/home

Europeana 2013 Vocabulary, Unpublished Report.

Griffiths, R., Dawson, M., and Rascoff, M. (2006). Scholarly Communications in the History Discipline. *ITHAKA Report Commissioned by JSTOR*, pp. 1-14. Retrieved from www.sr.ithaka.org/download/file/fid/485

Houghton, J., Steele, C., and Henty, M. (2003). *Changing Research Practices in the Digital Information and Communication Environment* (Doctoral Dissertation). Department of Education, Science and Training, pp. 1-186. Retrieved from www.cfses.com/documents/**Changing_Research_Practices**.pdf

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

Hughes, L.M. (2014). Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines, chapter in *The New Companion to Digital Humanities*, Schreibmann, S. and Siemens, R. (Eds.) (Oxford, Blackwells).

Hughes, L. M. (2014). Live and Kicking: The Impact and Sustainability of Digital Collections in the Humanities, in *Proceedings of the Digital Humanities Congress 2012*, Mills, C., Pidd, M, and Ward, E. (Eds.). Special edition of Studies in the Digital Humanities. Sheffield: HRI Online Publications.

Hughes, L. M., Ell, P., Dobreva, M., and Knight, G. (2013)*. Assessing and measuring impact of a digital collection in the humanities: An analysis of the SPHERE (Stormont Parliamentary Hansards: Embedded in Research and Education) Project, Literary and Linguistic Computing (Oxford University Press).

Hughes, L. M., and Ell, P. (2013). Digital Collections as Research Infrastructure, chapter in *'From Evolution to Transformation: Research Infrastructures and Scholarly Research*'. Special issue of the *International Journal of Humanities and Arts Computing* (Edinburgh University Press).

Hughes, L., Constantopoulos, P., & Dallas, C. (2015). Digital Methods in the Humanities: Understanding and De-scribing their Use across the Disciplines. In S. Schreibman, R. Siemens, & J. M. Unsworth (Eds.). *A new companion to digital humanities.* Oxford: Wiley-Blackwell, 150-170.

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

Meho, L., and Tibbo, H. (2003). Modeling the Information-Seeking Behavior of Social Scientists: Ellis's study revisited. *Journal of the American Society for Information Science and Technology*, *54*(6), pp. 570-587.

Metadata Offer New Knowledge (n.d.). Retrieved from http://monkproject.org/

Mihajlovic, J., Tojic, K., and Simu, M. (n.d.). *kulturklammer*. Retrieved from http://www.kulturklammer.org/view/64

Palmer, C., Pirmann, C., & Teffeau , L. (2010). *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development.* OCLC Research . Retrieved from: www.oclc.org/research/publications/library/2009/2009-02.pdf

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

Service Integration, data interconnect research and education – Isidore. (2013). Retrieved from http://www.rechercheisidore.fr/

Share More, Learn More (n.d.). Retrieved from http://www.seasr.org/

Sherratt, T. (2011). It's all about Stuff: Collections Interfaces, Power and People. *Journal of Digital Humanities, 1*(1). Retrieved from http://journalofdigitalhumanities.org/1-1/its-all-about-the-stuff-by-tim-sherratt/

Staten-General Digitaal. (n.d.). Retrieved from https://www.kb.nl/bronnen-zoekwijzers/dataservices-en-apis/staten-generaal-digitaal

Terras, M. (2009). The potential and problems in using high performance computing in the arts and humanities: The Researching e-Science Analysis of Census Holdings (ReACH) project. *Digital Humanities Quarterly, 3*(4).

Thaller, M. (Ed.). (2012). *Historical Social Research*, *37*(3).

Virtual Research Environment for the Humanities – TextGrid. (2012). Retrieved from http://www.textgrid.de/

Warwick, C.; Terras, M.; Galina, I.; Huntington, P.; Pappa, N. (2008). Library and information resources and users of digital resources in the humanities. *Program: Electronic Library and Information Systems*, *42*(1) pp. 5-27.

Unsworth, J. (2000). Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this?. *Symposium on Humanities Computing.* K. College (ed.), London. Retrieved from http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html

Und ich dachte immer. (n.d.). Retrieved from https://kars1918.wordpress.com/

Unsworth, J. (2003). Tool-Time, or 'Haven't We Been Here Already?'. *Presented at the Transforming Disciplines: The Humanities and Computer Science*. Washington, DC. Retrieved from http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html

Research Communities Web Survey

Boone, H.N. Jr. & D. A. Boone. (2012). Analyzing Likert Data, *Journal of Extension*, *50*(2).

Clason D.L. & T. J. Dormody. (2010). Analyzing Data Measured by Individual Likert-Type Items. *Journal of Agricultural Education*, *35*(4).

Jamieson, S. (2004). Likert scales: how to (ab)use them. *Medical education*. 38(12), pp.1217-1218.

Humanities and Social Sciences Case Studies

Addo, P., and Besnier, N. (2008). When Gifts Become Commodities: Pawnshops, Valuables, and Shame in Tonga and the Tongan Diaspora. *Journal of the Royal Anthropological Institute, 14*(1).

Aldhous, P. (2012). NodeXL for Network analysis. *NICAR 2012*. St Louis.

Ceni, A., Hogan, B., and Smith, M. (n.d.). Social Network Importer for NodeXL. Retrieved from
http://socialnetimporter.codeplex.com/

Dempster, P.G., and Woods, D.K. (2011). The Economic Crisis Though the Eyes of Transana. *Forum Qualitative Sozialforschung*, *12*(1), Art. 16.

Europeana Cloud Deliverable 1.2: State of the art report on digital research practices, tools and scholarly content use
http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.2%20State%20of%20the%20art%20report%20on%20digital%20research%20practices,%20tools%20and%20schol.pdf

Europeana Cloud Deliverable 1.4: Content Priorities for Humanities and Social Sciences research communities
http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.4%20Content%20priorities%20for%20Humanities%20and%20Social%20Sciences%20research%20communities.pdf

Europeana Cloud Deliverable 1.5: Expert Forum Case Studies Report
http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.5%20Expert%20Forum%20Case%20Studies%20Report.pdf

Europeana Cloud Deliverable 1.5 (1): Expert Forum Tools & Content for Humanities Research Report
http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.5%20Expert%20Forum%20Tools%20&%20Content%20for%20Humanities%20Research%20Report.pdf

Europeana Cloud Deliverable 1.5 (2): Expert Forum Tools & Content for Social Science Research Report
http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.5%20Expert%20Forum%20Tools%20&%20Content%20for%20Social%20Science%20Research.pdf

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative* inquiry, *12*(2), pp.219-245.

Hansen, D., Smith, M., and Shneiderman, B. (Eds.). (2010). *Analysing Social Media Networks with NodeXL*. Morgan Kaufman: New York, NY.

Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, *12*(2), pp. 281-316. Retrieved from http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf

Smith, M., et al. (2009). Analysing (Social Media) Networks with NodeXL*. Proceedings of the fourth international conference on Communities and technologies*.

Smith, M. (n.d.). Third-Party NodeXL Graph Data Importers. Retrieved from

http://nodexl.codeplex.com/wikipage?title=Third-Party%20NodeXL%20Graph%20Data%20Importers

Teddlie, C. and F. Yu. (2007). Mixed Methods Sampling: A Typology With Examples. *Journal of Mixed Methods Research*, *1*(1), pp. 77–100.

Wakita, K., and Tsurumi, T. (2007). Finding Community Structure in Mega-scale Social Networks. *Proceedings of the 16th International Conference on World Wide Web*. Banff, Alberta, Canada.

Yin, R.K. (2013). *Case study research: Design and methods* (5th ed.). Sage: Los Angeles, CA.

Documentation for Transana

Hillman, T. (2012). A geography of connections: Networks of humans and materials in mathematics classrooms using handheld technology. *Forum: Qualitative Social Research*, *13*(1), Art. 12. Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs1201121

Hillman, T. (2011). The inscription, translation and re-inscription of technology for mathematical learning. *Technology, Knowledge and Learning*, *16*, pp. 103-124.

Hillman, T. (2011). Where's the math? Locations of mathematical activity in classes with handheld technologies. *Peer reviewed conference paper: European Association for Research on Learning and Instruction*.

Hillman, T. (2013). Finding space for student innovative practices with technology in the classroom. *Learning, Media and Technology*. DOI: 10.1080/17439884.2013.803256.

Lundberg, A. L. V. and Hillman, T. (n.d.). Student-driven proportional reasoning approaches to an early algebra task. In Lindmeier, A M & Heinze, A (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education*, *5*. Kiel, Germany: PME.

Hillman, T. (forthcoming). Tracing the construction of mathematical activity with an advanced graphing calculator to understand the roles of technology developers, teachers and students". *International Journal for Technology in Mathematics Education*.

Software:
Transana [computer software] v. 2.53-MU. University of Wisconsin-Madison Center for Education Research, Madison, Wisconsin.

Homepages:
http://www.transana.org/
http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/index.htm

Documentation for HyperImage

Warnke, M. (2008). God Is In The Details. The Filing Box Answers. *GAZING INTO THE 21st CENTURY*. MIT Press. Im Druck. (Vortrag auf der International Conference: GAZING INTO THE 21ST CENTURY. Second

International conference on image science. Kloster Göttweig, Österreich 16.10.2008 - 18.10.2008, 18.10.2008.)

Warnke, M. (2012). Explicit Art Historical Image Referencing on a Big Scale. *LIBREAS. Library Ideas, 8*(21).

Conference Presentations by the HyperImage Team:

Loebel, JM. (2013). Retrieved from
http://hyperimage.ws/wp-content/uploads/imeji-Day-2013_Slides_Loebel_HyperImage.pdf

Warnke, M. (n.d.). Retrieved from http://www.uni-
lueneburg.de/hyperimage/hyperimage/downloads/bmbf-praesentation.pdf

Documentation of NodeXL

Research publications by Dr Bernie Hogan for which he used NodeXL as an analytical tool:
Hogan, B. (2010). Analyzing Facebook Networks. In: Hansen D., Smith M. and Shneiderman B. (Eds.), *Analyzing Social Media Networks with NodeXL*. Morgan Kaufman: New York, NY.

Hogan, B. (n.d.). *Pinwheel Layout to Highlight Community Structure.* Retrieved from
http://www.cmu.edu/joss/content/issues/2010jossviz/2_Hogan.htm

Parag, Y., Hamilton, J., White, V., and Hogan, B. (2013). Network approach for local and community governance of energy. *Energy Policy*, *62*, pp. 1064-1077.

Software:

NodeXL (Network Overview, Discovery and Exploration add-in for Microsoft Excel) [computer software] v. 1.0.1.251. The Social Media Research Foundation.

Homepage:
http://nodexl.codeplex.com/

APIs in Humanities and Social Sciences Research

Bradley, J. (2012). Towards a Richer Sense of Digital Annotation: Moving Beyond a "Media" Orientation of the Annotation of Digital Objects. *Digital Humanities Quarterly, 6.*
http://digitalhumanities.org:8080/dhq/vol/6/2/000121/000121.html

Dinmore, M. D., and Curtis Boylls, C. (2010). Empirically-observed end-user programming behaviors in Yahoo! Pipes. *Psychology of Programming Interest Group*.

Downie, J.S and Bainbridge, D. (2013). Integrating Independent Discover and Analysis Tools for the HathiTrust Corpus: Enhancing Fair Use Digital Scholarship. *JADH2013 & DH-JAC2013 Conference Abstracts 29-30*.

Lehane, R. (2012). Access to online archival catalogues via web APIs. *Proceedings from the 2012 International Council of Archives Congress.* Retrieved from http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00233.pdf

Macdonald, S. (2011). Addressinghistory: A Web2.0 Community Engagement Tool And API. *IASSIST Quarterly*, *35*(3): 17-20. Library, Information Science & Technology Abstracts.

McGregor, N. (2013). Mash-ups, APIs and Linked Data, Oh My!. *The British Library Digital Scholarship Blog.* Retrieved from http://britishlibrary.typepad.co.uk/digital-scholarship/2013/02/information-integration-mashups-apis-and-linked-data.html

Sherratt, T. (2011). It's all about Stuff: Collections Interfaces, Power and People. *Journal of Digital Humanities*, *1*(1). Retrieved from http://journalofdigitalhumanities.org/1-1/its-all-about-the-stuff-by-tim-sherratt/

Terras, M. (2009). The potential and problems in using high performance computing in the arts and humanities: The Researching e-Science Analysis of Census Holdings (ReACH) project. *Digital Humanities Quarterly*, *3*(4).

Research Themes

Parliamentary Papers in Europe

Debates Parliament. (n.d.). Retrieved from http://debates.parlamento.pt/?pid=mc

Digging into Linked Parliamentary Data. (n.d.). Retrieved from http://dilipad.history.ac.uk/

European Information and Research Network on Parliamentry History. (n.d.). Retrieved from http://euparl.net/9353213/v/

Gartner, R. (2014). A metadata infrastructure for the analysis of parliamentary proceedings. In Big Data (Big Data). *IEEE International Conference on Big Dta*, pp. 47-50.

Linking Parliamentary Records through Metadata. (n.d.). Retrieved from http://www.liparm.ac.uk/?page_id=61

Piersma, H., Tames, I., Buitnick, L., Doornik, J., & Marx, M. (2014). War in Parliament: What a Digital Approach Can Add to the Study of Parliamentary History. *Digital Humanities Quarterly, 8*.

WIP: War In Parliament. (n.d.). Retrieved from http://www.clarin.nl/node/410

Population Displacement as a result of conflict in the 20th century

Akritas, J. (2013). The Greek-Turkish Population Exchange: A Necessary Evil? Retrieved from http://hellenicantidote.blogspot.co.uk/2013/02/the-greek-turkish-population-exchange.html

Alpan, A. (2013). But the Memory Remains: History, Memory and the 1923 Greco-Turkish Population Exchange. *The Historical Review*(9), pp. 199-232.

Anghel, R. (2015). Migration in Differentiated Localities: Changing Statuses and Ethnic Relations in a Multi-Ethnic Locality in Transylvania, Romania. *Population, Space and Place*.

Baldwin-Edwards, M. (2006). Migration between Greece and Turkey: from the 'Exchange of Populations' to the non-recognition of borders. *SEER Review for Labour and Social Affairs*, pp. 115-122.

Beeley, B. (1978). The Greek-Turkish Boundary: Conflict At The Interface. *Transactions of the Institute of British Geographers*.

Clark, B. (2007). *Twice a stranger: how mass expulsion forged modern Greece and Turkey.* Granta.

Coufoudakis, V. (2008). International Law and Minority Protection: The Fate of the Greeks of Imbros and Tenedos. *Meditteranean Quarterly, 19*(4), pp. 14-28.

Csaba, B., Byrne, M., & Rainer, J. (2002). *The 1956 Hungarian revolution: a history in documents.* Central European University Press.

Dobkin, M. (1998). *Smyrna 1922: the Destruction of a City.* New Mark Press.

Drew, C. (2006). *Population transfer: the untold story of self-determination.* Cambridge University Press.

Erden, M. (2004). The exchange of Greek and Turkish populations in 1920s and its socio-economic impacts on the life in Anatolia. *Crime, law and social change, 41*(3), pp. 261-282.

Fassman, H., Reeger, U., & Sievers, W. (2009). *Statistics and Reality: Concepts and measurements of migration in Europe.* Amsterdam University Press.

Granville, J. (2005). Caught with Jam on Our Fingers": Radio Free Europe and the Hungarian Revolution of 1956. *Diplomatic History, 29*(5), pp. 811-839.

Gursoy, Y. (2008). The effects of the population exchange on the Greek and Turkish political regimes in the 1930s. *East European Quarterly, 42*(2).

Hirschon, R. (2014). The long-term significance of the 1923 Greco-Turkish exchange of populations. In V. Lytra, *When Greeks and Turks meet: Interdisciplinary Perspectives on the Relationship Since 1923.* Publications of the Centre for Hellenic Studies, King's College London: 15.

Hirshon, R. (2003). *Crossing the Aegean: an appraisal of the 1923 compulsory population exchange between Greece and Turkey.* Berghahn Books.

Klapsis, A. (2011). Research Note: American Initiatives for the Relief of Greek Refugees, 1922-1923. *Genocide Studies and Prevention, 6*(1), pp. 98-106.

Kollugku, B. (2013). Excesses of nationalism: Greco-Turkish population exchange. *Nations and nationalism, 19*(3), pp. 532-550.

Kritikos, G. (2005). The agricultural settlement of refugees: A source of productive work and stability in Greece, 1923-1930. *Agricultural History*, pp. 321-346.

Lausanne Peace Treaty VI. Convention Concerning the Exchange of Greek and Turkish Populations Signed at Lausanne, January 30, 1923. (n.d.). Retrieved from http://www.mfa.gov.tr/lausanne-peace-treaty-vi_-convention-concerning-the-exchange-of-greek-and-turkish-populations-signed-at-lausanne_.en.mfa

Milton, G. (2008). *Paradise Lost: Smyrna 1922: The destruction of Islam's City of Tolerance.* Basic Books.

Pentzopoulos, D. (2002). *The Balkan exchange of minorities and its impact on Greece.* Husrt & Co. .

Santayana, G. (1998). The life of reason. *Fixed-versus Random-Effects Models, 1*, 1905.

Schopflin, G. (1996). *The Hungarian Revolution of 1956: reform, revolt and repression, 1953-1963.* London: Longman.

Sheridan, V. (2010). The inadvertent cosmopolitan: a refugee tale from Hungarian to English. In P. O'Donovan, & L. Rascaroli, *The Cause of Cosmopolitanism: Dispositions, Models, Transformations.* Peter Lang AG, Internationaler Verlag der Wissenschaften.

Shields, S. (2013). The Greek-Turkish Population Exchange Internationally Administered Ethnic Cleansing. *Middle East Report*(267), pp. 2-6.

Sirkeci, I. (2003). *Migration, ethnicity and conflict: the environment of insecurity and Turkish Kurdish international migratio.* Sheffield: University of Sheffield.

Stroebel, W. (2014). Distancing Disaster: Trauma, Medium, and Form in the Greco-Turkish War and Population Exchange. *Journal of Modern Greek Studies, 32*(2), pp. 253-285.

The Joint Czech and Slovak Parliamentary Library. (n.d.). Retrieved from http://www.nrsr.sk/dl/?lang=en

Today's Zaman. (n.d.). Retrieved from Nansen's solution: Revisiting the Turkish-Greek population exchangeby Şener Aktürk: http://www.todayszaman.com/op-ed_nansens-solution-revisiting-the-turkish-greek-population-exchangeby-sener-akturk-_314148.html

Tsilfidis, A. (n.d.). *Pontosroworld*. Retrieved from The Exchange of Populations Between Greece and Turkey: http://pontosworld.com/index.php/pontus/history/articles/295-the-exchange-of-populations-greece-and-turkey

Yildirim, O. (2006). The 1923 population exchange, refugees and national historiographies in Greece and Turkey. *East European Quarterly, 40*(1), pp. 45-70.

Yildirim, O. (2007). *Diplomacy and displacement: Reconsidering the Turco-Greek exchange of populations, 1922–1934.* Routledge.

Political Propaganda

Enenkel K. and Pfeijfer I. L. (Eds.). (2005). *The Manipulative Mode: Political Propaganda In Antiquity. A Collection Of Case Studies*. Brill Academic Pub.

Children's Literature

Random History.com (2008). To Instruct and Delight: A History of Children's Literature. Retrieved from
http://www.randomhistory.com/1-50/024children.html

Hunt, P. (Ed.). (2002). *Understanding children's literature* (2nd ed.). Routledge.

Wolf, S., et al (2011). *Handbook of Research on Children's and Young Adult Literature.* Hoboken : Taylor
    and Francis.

Edwards, V., Pemberton, L., Knight, J. and Monaghan, F.(n.d.). *FABULA: A BILINGUAL MULTIMEDIA
AUTHORING ENVIRONMENT FOR CHILDREN EXPLORING MINORITY LANGUAGES*. Retrieved from
http://www.it.bton.ac.uk/staff/lp22/fabula.doc

What is Digital Storytelling (n.d.). Retrieved from
        http://digitalstorytelling.coe.uh.edu/page.cfm?id=27&cid=27&sublinkid=29

Digital Storytelling and Cultural Heritage: Stakes and Opportunities (2015). AthenaPlus Project Report.

Appendices

Appendix 1: Europeana Cloud Research Community Web Survey

In the context of Europeana Cloud: Unlocking Europe's Research via The Cloud, we are conducting a web survey to gather information on digital research practices, tools and content and to collect evidence-based data from the Humanities and Social Sciences research communities, focusing in particular on the potential use of content from Europeana and the European Library within Europeana Cloud.

We are conducting a quick questionnaire research on behalf of Europeana Cloud (http://pro.europeana.eu/web/europeana-cloud).

Your answers will help us find out how Humanities and Social Sciences researchers discover and use content from Europeana and the European Library within Europeana Cloud.
The full questionnaire consists of fourteen questions. It should take you between 10-15 minutes to respond to all questions.

The initial results of this questionnaire will be released at the end of 2013.
Thank you for taking the time to respond!

There are 14 questions in this survey

- **Assessment of Research Practices and Needs**

Tell us about what you find important in the way conduct your research and your digital technology needs.

1. How often have you used the following services in the last 12 months?

(1 = never, 2 = a few times, 3 = at least once a month, 4 = at least once a week, 5 = several times a week)

Please choose the appropriate response for each item:

	1	2	3	4	5

	1	2	3	4	5
The Europeana portal	○	○	○	○	○
The European Library	○	○	○	○	○
Google Scholar	○	○	○	○	○
Specific online digital archives related to my research	○	○	○	○	○
A commercial portal of a scholarly journal	○	○	○	○	○
Museum websites	○	○	○	○	○
Online library catalogues	○	○	○	○	○
Other	○	○	○	○	○

2. **Identify the digital archives or other services you stated you have used in your response to the previous question.** (Use one line per archive or service you meant)

Please write your answer(s) here:

- Digital archive 1

- Digital archive 2

- Digital archve 3

- Other service 1

- Other service 2

- Other service 3

3. **Assess how important are the following activities as you seek relevant information for your research.**

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	2	3	4	5
I search using specific query terms in order to find relevant information.	○	○	○	○	○
I search using a combination of more than one query terms to find relevant information.	○	○	○	○	○
I conduct a new search within the results of a previous search.	○	○	○	○	○
I consult collection summaries to find information that interests me.	○	○	○	○	○
I consult collection inventories or indexes to find information that interests me.	○	○	○	○	○
I browse or leaf through related resources to find those that interest me.	○	○	○	○	○
I find relevant information on the basis of resources I just happen to stumble upon.	○	○	○	○	○
I find resources through footnotes in articles/books.	○	○	○	○	○
I ask peers/colleagues.	○	○	○	○	○
I ask reference librarians/curators/archivists.	○	○	○	○	○

4. **Please give an example of a common query you may ask as you seek relevant information for your research.**

Please write your answer here:

5. **Assess the importance of the following aspects of the content of a resource as you seek relevant information for your research.**
   (1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	2	3	4	5
The names of people mentioned in, or represented by, a resource. (e.g. Alexander the Great, Napoleon, Hiltler, Mozart, Mona Lisa)	○	○	○	○	○

	1	2	3	4	5
Kinds of objects, artefacts, art works. (e.g. portrait, statue)	○	○	○	○	○
Classifications of people mentioned or represented (social, occupation, gender, age, etc.) (e.g. colonel, admiral, king)	○	○	○	○	○
The specific places mentioned or represented (e.g. Athens, Falkland Islands, Auschwitz)	○	○	○	○	○
Classifications of places mentioned or represented (e.g., river, mountain, city, harbour etc.)	○	○	○	○	○
The dates, time spans or periods related to a resource	○	○	○	○	○
Names of specific events related to a resource (e.g., 1st World War, coronation of Queen Victoria, battle of Marathon, the French Revolution)	○	○	○	○	○
Classifications of events, activities, or processes mentioned or represented (e.g., war, buidling, hunting, family life, wedding, Olympic games)	○	○	○	○	○
Other things, ideas or entities related to a resource, not mentioned above	○	○	○	○	○

6. **Assess the importance of the following properties of a resource as you seek relevant information for your research.**

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	2	3	4	5
The style, period, group or movement related to a resource (e.g., impressionism, gothic, Hellenistic, pre-Rapahelites, Halstatt, magical realism)	○	○	○	○	○
The format of a resource (e.g., text, image, video, .jpg, .pdf etc.)	○	○	○	○	○

	1	2	3	4	5
The object type related to a resource (e.g., oil painting, etching, pottery, albumen print, book, church)	○	○	○	○	○
The genre of a resource (e.g., correspondence file, short story, landscape painting)	○	○	○	○	○
The name of the author, creator or issuing authority of a resource	○	○	○	○	○
The fonds, collection or holding institution to which a resource belongs	○	○	○	○	○

7. Assess how important is each of the following activities you undertake as you organize unpublished materials.

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	2	3	4	5
I collect lists of references to primary resources in a paper document or word processing file.	○	○	○	○	○
I manually copy or retype the content of resources that interest me.	○	○	○	○	○
I file together photocopies of resources that interest me.	○	○	○	○	○
I file together digital copies of resources that interest me.	○	○	○	○	○
I use catalogue cards to manage resources that interest me.	○	○	○	○	○
I collect and keep references to all resources that interest me.	○	○	○	○	○
I use keywords in order to identify the topic of resources that interest me.	○	○	○	○	○
I maintain a manual index of keywords on resources that interest me.	○	○	○	○	○
I use a software application to organize resources that interest me.	○	○	○	○	○

	1	2	3	4	5
I keep a list of keywords for references to primary material	○	○	○	○	○
Other	○	○	○	○	○

8. Assess how important is each of the following activities as you organize published materials.

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	2	3	4	5
I collect relevant bibliographic references in a paper document or word processing file.	○	○	○	○	○
I obtain, copy and/or print and store physical copies of published materials relevant to my research.	○	○	○	○	○
I keep a list of keywords for bibliographic references I have collected.	○	○	○	○	○
I download and store digital copies of published materials relevant to my research.	○	○	○	○	○
I use a bibliographic reference management application to manage references and/or published materials.	○	○	○	○	○

9. Assess how important is each of the following activities as you study and annotate information relevant to your research.

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	2	3	4	5
I study relevant texts systematically from beginning to end.	○	○	○	○	○
I scan texts quickly for relevant passages.	○	○	○	○	○

	1	**2**	**3**	**4**	**5**
I keep margin notes on passages of texts that interest me.	○	○	○	○	○
I keep relevant passages and notes within the word processing file of my research draft.	○	○	○	○	○
I underline or highlight relevant passages from texts I read.	○	○	○	○	○
I keep my own notes together with passages of text I find important.	○	○	○	○	○
Other (please specify and rate importance)	○	○	○	○	○

10. Assess how important is of each of the following activities as you work with others for your research.

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	**2**	**3**	**4**	**5**
I ask colleagues for their expert opinion on my initial research ideas.	○	○	○	○	○
I ask colleagues for their expert opinion on specific resources I work with at an advanced stage of my research.	○	○	○	○	○
I share my own notes on specific resources with colleagues.	○	○	○	○	○
I collaborate with colleagues on joint publications or conference papers.	○	○	○	○	○
I share copies or resources I find interesting with colleagues.	○	○	○	○	○
I collaborate with colleagues on developing shared information resources, databases etc.	○	○	○	○	○
Other (please specify and rate importance)	○	○	○	○	○

11. Specify how much you agree with each of the following statements.

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	Yes	Uncertain	No
I would like to know which scholars in my field work on a particular source, or research question.	○	○	○
I would be prepared to share information with colleagues on the sources or research questions I work on.	○	○	○
I would like to work in collaboration with others towards joint publication or common research results.	○	○	○
I find journal papers and books more trustworthy than online publications.	○	○	○
I find paper finding aids more trustworthy than online finding aids.	○	○	○
I would share interesting resources with colleagues if I was allowed to (e.g., overcoming copyright restrictions).	○	○	○
I regard copyright or privacy issues as important obstacles for my research.	○	○	○
I find resources in a physical archive or collection more trustworthy than those in a digital archive.	○	○	○

12. Assess how important is each of these functions to help you retrieve to your local environment different kinds of resources.

(1 = not at all, 2 = somewhat, 3 = moderately, 4 = very, 5 = extremely)

Please choose the appropriate response for each item:

	1	2	3	4	5
Fetching and automatically importing search results on online digital resources into Excel	○	○	○	○	○
Fetching and automatically importing search results on online digital resources into a word processing file	○	○	○	○	○

	1	2	3	4	5
Fetching and transferring metadata in XML or RDF from into my computer	○	○	○	○	○
Fetching and automatically importing metadata about online digital resources into Excel	○	○	○	○	○
Fetching and automatically importing metadata into a word processing file	○	○	○	○	○
Fetching and automatically importing metadata into a research repository or database	○	○	○	○	○
Fetching and automatically importing metadata into a notetaking program (e.g., Onenote, Evernote)	○	○	○	○	○
Fetching and automatically importing online digital resources (e.g., images) into a word processing file	○	○	○	○	○
Fetching and automatically importing online digital resources (e.g., images) into a research repository or database	○	○	○	○	○
Fetching and automatically importing online digital resources (e.g., images) into a notetaking program (e.g., Onenote, Evernote)	○	○	○	○	○
Fetching and automatically importing bibliographic references into bibliography applications (e.g., Zotero, Endnote, Refworks or Mendeley)	○	○	○	○	○
Receiving automatic notifications on the existence of online digital metadata and resources of interest	○	○	○	○	○

- **Personal Information**

Thank you very much for your response to our questionnaire! And now, give us some brief information about yourself.

   1. **How would you describe yourself?**

Please select at least one answer:

   ☐ Researcher working in an academic institution

☐ Researcher working in a research institution outside the academic sector

☐ Researcher, freelance PhD or postgraduate student

☐ PhD or postgraduate student

☐ Undergraduate student

☐ Archivist

☐ Librarian

☐ Curator

☐ Museum professional

☐ Amateur researcher

☐ Casual user

☐ Other:

2. Identify your primary and, if applicable, secondary fields of research.

(Please take a moment to examine available options before responding. In this question you have to chose at least one primary, one secondary and one not applicable option.)

*Please choose the appropriate response for each item:

	Primary field	Secondary field	Not applicable
African and Oriental Studies	○	○	○
Anthropology	○	○	○
Applied Ethics	○	○	○
Archaeology	○	○	○
Art and Art history	○	○	○
Classics	○	○	○
Communication & Media Studies	○	○	○
Criminology	○	○	○
Cultural Studies	○	○	○
Drama & Theatre	○	○	○
Economics and Economic History	○	○	○
Education	○	○	○

	Primary field	Secondary field	Not applicable
English Language and Literature	○	○	○
Ethnic, Gender and Cultural Studies	○	○	○
Family Studies	○	○	○
Folklore	○	○	○
Geography	○	○	○
History	○	○	○
History of Science, Technology & Medicine	○	○	○
History of Social Sciences	○	○	○
Human Ecology	○	○	○
Jurisprudence	○	○	○
Languages and Literature	○	○	○
Law	○	○	○
Linguistics	○	○	○
Medieval Studies	○	○	○
Music	○	○	○
Paleontology	○	○	○
Pedagogical and educational research	○	○	○
Philosophy	○	○	○
Philosophy, Ethics and Religion	○	○	○
Political Science & Public Administration	○	○	○
Psychology	○	○	○
Social and behavioural sciences	○	○	○
Social Sciences	○	○	○
Sociology, Demography and Social Statistics	○	○	○
Theology and Religious Studies	○	○	○

Appendix 2 Europeana Research Questionnaire – Political Propaganda

The WP1-team sent out this questionnaire to 10 researchers in Europe.

Two responded.

Europeana Research Questionnaire – Political Propaganda

Information to be included in the Invitation
Europeana Cloud is a project within Europeana that aims to ultimately develop a web-based platform, to provide researchers in the Humanities and Social Sciences with access to digital content from institutions across Europe, and also to offer a space in which researchers can analyse that content with tools and services. This platform will be called **Europeana Research**. For the purposes of the platform Europeana Research, a team from Work Package 1 of the project is exploring different topics and assessing the current research requirements and available tools and resources for that topic. This will help Europeana assess the feasibility of aggregating data in this area. Having selected "Political Propaganda" as one of the research topics to be explored, the aim of this task is to investigate how the community conducts research in this area, which methods it employs, available tools and resources used. Your time and assistance is much appreciated.
1. Research career / background
1.1. Can you describe your research (academic?) career / background? 1.2. What is your institutional affiliation now? What is your title/position? 1.3. Do you consider yourself digitally literate? What is your experience with digital applications? 1.4. What is you research interest regarding the topic of Political Propaganda?
2. Research practices
Now I would like to ask more about your research in the area of Political Propaganda.

2.1. Can you list a few keywords/research questions that summarize your research in this field?

2.2. Is the digital aspect central in this? Why not?

3. Resources & digital tools

3.1 Which are the main resources you employ in your research on Political Propaganda? Do you consult digital *and* non-digital material?

3.2. What types of data (*text, images, video, audio, 3D*) have you mostly worked with in your research?

❏ Image

❏ Text

❏ Audio files

❏ Video

❏ 3D

3.3 How would you characterize the importance of digital resources for your research area?

To what extent did the 'digital turn' influence the formulation of new research questions in your field/your research in general?

3.4 Have you heard of Europeana before? Have you used it in your research for accessing research content?

3.5 Do you employ digital tools/services in your research? If yes, what?

3.6 Can you describe an application of digital tools in your most recent research?

- Is it data collection, data analysis, data visualisation?

- At what stage of your research would this usually come in?

3.7 Which digital tools/services would you consider to be most useful when processing research data?

3.8 If we think about the digital tools/services in question, how extensively are they used in your field / research community? Why do you think it is so?

4. Future developments and content

4.1. Can you suggest specific collections of research materials in archives, libraries, museums or other institutes for Europeana to acquire and aggregate, to facilitate and enrich your future research?

Please be as *specific* as you think is called for.

4.1.a. are they digital

4.1.b. are they non-digital

4.1.c. or both?

4.2. What are the current problems you face in your research in the area of Political Propaganda and can you see Europeana Research helping you to overcome them in the future?

And if so, how?

Appendix 3 Parliamentary Papers in Europe

See attached .PDF

Appendix 4: European School History of the 20th Century

Participants

The following people have contributed information to this topic, for which we are immensely grateful:

Hilda T.A. Amsing, University of Groningen, the Netherlands

Ingrid Brühwiler, Department of History, University of Lausanne, Switzerland

Ning de Coninck-Smith, Danish School of Education, University of Aarhus

Henrik Åström Elmersjö, Department of Historical, Philosophical and Religious Studies, Umeå University, Sweden; ed. *Nordic Journal of Educational History*

Alexandre Fontaine, University of Geneva, Switzerland and ENS-Ulm Paris, France

Mikelis Grivins, Baltic Studies Centre, University of Latvia, Latvia

Wilfried Göttlicher, Department of Education, University of Vienna

Patrícia Hansen, Institute of Social Sciences, University of Lisbon, Portugal

Rita Hofstetter, Faculty of Psychology and Educational Sciences, University of Geneva

Iveta Ķestere, Department of Pedagogy and Art, Latvia University, Latvia

Anna Larsson, Department of Historical, Philosophical and Religious Studies, Umeå University, Sweden; ed. *Nordic Journal of Educational History*

Germund Larsson, Department of Education, Uppsala University, Sweden

Maria Núria Padrós Tuneu, Department of Pedagogy, University of Vic – Central University of Catalonia, Spain

Eulàlia Collelldemont Pujadas, Department of Pedagogy, University of Vic – Central University of Catalonia, Spain

Vadim Rõuk, Estonian Pedagogical Archives-Museum, Tallinn University, Estonia

Sarah Van Ruyskensvelde, Faculty of Psychology and Educational Sciences, KU Leuven, Belgium

Wolfgang Sahlfeld, Department of Teaching and Learning, The University of Applied Sciences and Arts of Southern Switzerland (SUPSI)

David Sjögren, Department of History, Uppsala University, Sweden; ed. *Nordic Journal of Educational History*

Irēna Stonkuviene, Faculty of Philosophy, Vilnius University, Lithuania

Branko Šuštar, Slovenian School Museum, Slovenia; President of the Historical Association of Slovenia

Joanna Wojdon, Faculty of Historical and Pedagogical Sciences, University of Wroclaw, Poland

Nataša Vujisić Živković, Faculty of Philosophy, University of Belgrade, Serbia

Search Phrases for Exploration of Europeana

school life (subject)

school children (subject)

school building (subject)

school group (titles)

education policy (subject)

skolehistorie (subject)

school history

schulgeschichte

schulmuseum

skolen (subject)

skoleprosjekt

classroom

classroom (as subject)

education (subject): too small precision

education and training: small precision (yielded results concerning the armed forces and practical skills)