Digital Humanities and Future Archives

SAMPO VIIRI

REPORTS

The Finnish Institute is a London-based private trust. Our mission is to identify emerging issues relevant to contemporary society and to act as catalyst for positive social change through partnerships.

We work with artists, researchers, experts and policy makers in the United Kingdom, Finland and the Republic of Ireland to promote strong networks in the fields of culture and society. We encourage new and unexpected collaborations and support artistic interventions, research, the creative industries, foresight and social innovation in new, socially central areas.

The Reports of the Finnish Institute in London is a series of publications, which publishes research, studies and results of collaborative projects carried out by the institute. The reports provide evidence and ideas for policy-makers and civic society organisations dealing with contemporary social and cultural challenges.

Unit 1, 3 York Way
N1C 4AE London
United Kingdom

T +44(0)20 3764 5090
www.finnish-institute.org.uk

August 2014

Table of Contents

Preface

What are the future archives like and how can disciplines such as open data and digital humanities facilitate innovative archival practices and support the work of historians? This has been a question which underlies our work on exploring the emerging issues in the field of sustainable long-term preservation and accessibility of digital cultural heritage.

Our thinking starts from the premise that culturally valuable digital contents should be released under the requirements of open definition whenever possible, both in order to support historical research and to celebrate the rich variety of our cultural domains, but also to facilitate citizen empowerment and creative innovations.

Thus, our idea of future archives does not only reflect the issue of preservation, but parallels wider questions regarding participatory historical culture and ideals of open knowledge.

Participatory historical culture aims at both improving historical consciousness and offering citizens a possibility of addressing their individual concerns and thinking over how to make a better future for themselves. This is in parallel with the ever-increasing urge to further empower citizens to take part in public life and open up the public discourse for equal advancement of interests. One of the goals is indeed to provide ways for people to educate themselves in preserving historically and culturally valuable information, and to increase the historical knowledge.

With this setup we commissioned this survey, which looks on the emergence of digital humanities – a discipline which includes elements of open knowledge, digital preservation and collaborative research. The survey addresses simultaneously the promises and challenges of open GLAM (galleries, libraries, archives and museums) movement and the increasingly important concept of long-term preservation and accessibility of digital contents.

I wish to thank researcher Sampo Viiri, who has addressed the difficult concept without fear and managed to compile an enjoyable, coherent and informative report in a fairly limited time.

Antti Halonen
Head of Society Programme
The Finnish Institute in London

Executive Summary

This paper presents recent trends in digital humanities research and digital cultural heritage preservation in the UK, Ireland and Finland. With the main focus areas being history and archives, the survey draws an argument on how the higher education and the GLAM (galleries, libraries, archives, and museums) organisations could collaborate in digital humanities projects and also promote the development and use of open knowledge.

The key findings of the survey were:

- Access to digital content online has benefited researchers and can improve geographic and socioeconomic equality.
- Digital history may narrow the gap between academic and popular history, and increase cross-disciplinary and international collaboration.
- New digital tools have not yet changed most historians' research methods substantially but digital tools and methods should be included in history curriculum.
- Libraries, archives and universities see future digital preservation challenges similarly, which encourages further collaboration.
- Digitisation projects are work hour intensive and require new skills and attitudes.
- Private-public partnerships in the culture heritage sector are in many ways controversial but may offer opportunities for collaboration.
- Measuring the societal value of digital humanities is difficult but might be possible in the near future.

This paper cites good projects and recent literature from the respective countries in order to circulate ideas and facilitate discussion. Direct comparisons between the impacts of projects in different countries would be challenging but further research in that area should also be done.

1 Introduction

1.1 Objectives

The evolution of digital technology and online networks is shaping our lives and societies. At some point in the recent past, digital stopped being something separate from the "real" world for many of us. Now we are constantly in the network and using digital tools without even noticing it ourselves. Digital information is everywhere and largely available for everyone.[1]

In the United Kingdom and other countries there has also been a wide interest in open data and open knowledge, a theme that the Finnish Institute in London has closely followed recently.[2] Moreover, in Finland the Open Science and Research Initiative, established by the Ministry of Education and Culture, aims at Finland becoming the leading country in openness of science and research by the year 2017 and that the opportunities of the open science will be widely utilised in Finnish society.[3]

However, archivists and IT experts respectively have expressed their concerns on the so called digital dark age: a phenomenon where significant amount of important cultural heritage is lost due to rapid digitisation of information and lack of reliable long-term preservation methods. [4]

How does this change influence humanities research and teaching and the preservation of cultural heritage?

Digital humanities, the intersection of digital technologies and humanities, has been a buzzword of the humanities field in the last few years. In the open data field the United Kingdom has so far been seen as a pioneering country. [5] There are also many digital humanities centres in the UK[6] and English language literature scholars are especially present in the international digital humanities discussion. We see the fields

---

[1] About the thoroughly digital world, Vuorinen 2014.

[2] See Halonen 2012 & *The Open Book*. This survey concentrates on digital humanities and cultural heritage, not particularly on open government data although we address political history, parliamentary data etc.

[3] Open Science and Research Initiative, http://openscience.fi/

[4] See for example: Koehl 2013; Colvile 2013; Digital Preservation Handbook.

[5] Open Data Index. https://index.okfn.org/country/

[6] The International Directory of Digital Humanities Centers. http://digitalhumanities.org/centernet/centers

of open data, digital humanities and cultural heritage preservation deeply intertwined.

In Finland the concept of digital humanities is known but there is still fairly little of formal contribution. CenterNet, an international network of digital humanities centres, lists no centres in Finland, and most European centres are geographically in the Western Europe. The other Nordic countries each have one centre[7]. However, the field is blooming in Finland and recently there has been progress in the area.[8]

The main objectives of this survey were to find out:

- What kind of research has been done in the UK and Ireland on digital humanities and how is the discipline likely to evolve?
- What is the status quo in Finland and what kind of practices should be brought in from the UK?
- Focusing on the fields of history and the archives, we want to know how could study of social sciences be supported via methods of digital humanities?
- In more precise terms, what could the future archives be like and how can disciplines such as digital humanities facilitate innovative digital preservation practices.
- How can digital humanities and open data projects support societal change?

History and digital preservation are interesting areas to survey because archives and other cultural heritage organisations face enormous challenges and possibilities when information is being digitised and new information is increasingly born-digital.

Compared to some other academic disciplines history has traditionally had a public facing role. History is also one of the important links between the humanities and social sciences, although digital history can be included as a part of the larger digital humanities area.[9]

---

[7] centerNet. http://digitalhumanities.org/centernet/. The definition of a DH centre is of course problematic and one can ask whether digital humanities is more about centres or networks anyway.

[8] Examples: Media Lab Helsinki in the Aalto University, Digital humanities professorship jointly in the University of Helsinki and the National Library's digitisation centre, Digihumanities Finland seminars and community, informal meetings, and closely connected to the field but not necessarily using DH vocabulary, Open GLAM and Open governmental data initiatives.

[9] More about digital history see for example Cohen & Rosenzweig 2005. Also: Parland-von Essen & Nyberg 2014, a Swedish language collaboration between the Finnish and Swedish scholars.

After defining the digital humanities and how the field is understood in this survey, the second chapter examines how the digital is helping history research and the public's perception of history. The third chapter surveys the possibilities and challenges for cultural heritage preservation in the digital age. Finally the state of the digital humanities field is summarised and some future collaboration possibilities are suggested.

1.2 Methods

The survey was conducted by going through existing literature, web discussions in both English and Finnish, and interviewing professionals associated with digital humanities. Seven semi-structured interviews were conducted face to face in London, four via email. All interviewees were asked the same questions, still allowing a free flow of conversation.

The interviewees' answers have been anonymised but a list of the interviewed people can be found at the end of the report. Although they come from various institutions the statements do not necessarily represent the views of their employers. We would like to thank all the interviewees and other people who have contributed in some way for their deep insights, ideas and suggestions.

The interviewees were asked the following questions:

- How would you define digital humanities?
- Explain your own work on digital humanities and what have been your goals?
- How expensive have different digital humanities initiatives been? What has the workload been like? Have digital humanities projects been worth it?
- It is said that the interest in history has grown rapidly. Is there evidence of that?
- Is there evidence of the digital improving the operating conditions of historians? Can it be quantified and if so, how?
- What is the future of digital preservation: Utopia/dystopia in the year 2030.
- What is the value of open data for humanities and the societal value of digital humanities? Is there evidence on improved decision-making?

The interviews contribute the most contemporary material, which is particularly useful in the digital humanities field where a book published a few years ago may feel terribly outdated. Digital humanities is constantly on the move and there is a steady

flow of blog posts and tweets about the subject. The internet itself is also constantly evolving and the speed of computing increases and the prices decrease, opening new research possibilities.

The interviewees represent a rich variety of perspectives for the digital humanities theme: librarians, archivists, academic researchers and grassroots activists. Researchers, the government, the GLAM (galleries, libraries, archives, museums) sector and the broad public view the field from different standpoints. We argue that the opportunities and challenges in digital research and digital preservation are similar throughout the GLAM sector and for the academics using that material.

The hype around digital humanities is so strong[10] that it's important to adopt a critical view on the research projects. We have tried to collect information on what the societal impact of digital humanities projects has been and what have been the costs and benefits. By reviewing some of the challenges we can improve the quality of future projects, avoiding falling into some pitfalls.

The survey is "humanist" in the way that it both tries to answer some existing questions and pose new ones. This survey contributes to the existing discussion by circulating good ideas to the Finnish context. We want to promote international discussion and strengthen the connections between Finland, UK and the rest of the world. The discussion is already international so the footnotes reference important contemporary literature and projects, mainly from different parts of the Anglosphere.

1.3 Defining Digital Humanities

Some scholars have argued that by including digital methods, traditional humanistic research may ask genuinely new questions and transform the whole field of study. Digital humanities has been labelled as a saviour to the field where tightening budgets and limited research funding leave scholars in desperate need to demonstrate their value to societies.[11] It may increase the humanities' value for societies and also involve mass participation by the public in projects traditionally done by lone scholars in the academia.

[10] More about the recent discussion Kirschenbaum 2014; Terras 2014.

[11] A very optimistic view can be found for example in the introduction of the handbook *Digital_Humanities,* Burdick, Drucker et al. 2012.

The hype around digital humanities has led to a certain new interest towards humanities but also created some confusion when everybody may want to label their research as digital humanities simply to get research grants.[12] One quote describes perfectly the confusion and frustration in defining digital humanities. "Digital humanities is what digital humanists do."[13]

So what does digital humanities actually mean? Articles and even whole books have been written on the definition and history of digital humanities.[14] The simplest broad definition would be:

> *Digital humanities involves the use of digital tools in research, teaching, scholarship and publication in humanities disciplines*.

This can mean answering new kind of questions and presenting data in ways that would have previously been impossible. The projects can add a certain extra layer to old contents or transform the research process thoroughly. Using maps and GIS technology is one of the most prominent examples of this.[15]

Digital humanities can also mean using digital tools in answering fairly traditional questions. For example digitised material and text search tools make the research process faster and more efficient. Distant reading and text mining are one of the most prevalent textual methods.[16] Another strand of research that is sometimes labelled as digital humanities is examining humanity in the digital age. This encompasses especially research of digital media.

Overall the field is in constant change. Nowadays every humanities scholar uses computers in word processing, reading texts, finding references or communication. Usually it is argued that communication in Twitter or sharing files in Dropbox still do not qualify as digital humanities activities. However, as one of our interviewees also emphasised, the impact of those tools should not be minimised because they really do change the working practices.

---

[12] Kirschenbaum 2010, 5.

[13] Quote by Rafael Alvarado. Reprinted in *Debates in the Digital Humanities*, chapter: "Day of DH: Defining the Digital Humanities".

[14] *Defining Digital Humanities*. See especially chapter 20, "'Definitions from the Day of Digital Humanities 2009-2012", 279-287.

[15] More examples in chapter 2.1.

[16] List of some UK projects:
http://www.jisc.ac.uk/whatwedo/topics/datatextmining.aspx?page=1&filter=Projects

One common feature in many digital humanities projects is building things. Whereas traditional humanities research usually outputs a text, in digital humanities the output can also be a database or some other piece of digital infrastructure. If you need to build things to be a digital humanist, do you also need to code? That is another question which has raised different opinions.[17]

If the 'digital' part of digital humanities is somewhat controversial, also the 'humanities' part of the concept is rather ambiguous. The infograph by UCL Digital Humanities Centre and 4Humanities defines humanities:

> *"The Humanities are academic disciplines that seek to understand and interpret the human experience, from individuals to entire cultures, engaging in the discovery, preservation, and communication of the past and present record to enable a deeper understanding of contemporary society. The Humanities are academic disciplines that seek to understand and interpret the human experience, from individuals to entire cultures, engaging in the discovery, preservation, and communication of the past and present record to enable a deeper understanding of contemporary society."*[18]

Digital humanities however seeks to cross and redefine the borderlines among the humanities, the social sciences, the arts, and the natural sciences. Digital humanities projects are usually collaborational and in many cases also cross-disciplinary, including significant representation from information studies, computer science, and library and archives professionals. By inventing new forms of inquiry, digital humanities can expand the scope and quality of research and reach new audiences for humanities studies.

Some academic disciplines such as political history or media studies lie in the borderline between humanities and social sciences. Whereas digital humanities has definitely been a buzzword for years, there has been virtually no talk about digital social sciences. Have digital methods been so thoroughly engaged in social sciences that they don't need to talk about digital research as a specific area of study? Is this the same case in humanities as well? And if both 'digital' and 'humanities' parts of the term are controversial, is there really a future for digital humanities scholarship

---

[17] Gold 2012.

[18] The Humanities Matter. http://4humanities.org/wp-content/uploads/2013/07/humanitiesmatter300.pdf

or should digital humanities simply be called humanities or digital scholarship/research?

Digital humanities has a history of many decades but the methods and objectives have changed a lot with the rapid technological progress within the last few years. The beginning of digital humanities scholarship is usually pinpointed to the Italian Jesuit priest Roberto Busa, one of the pioneers in the usage of computers for linguistic and literary analysis. In the 1940's, working with IBM, he developed a tool for performing text searches within the corpus of Thomas Aquinas's works.[19]

Since Busa's early research the technological possibilities in the usage of computational and quantitative tools in humanities research have of course expanded enormously. Prior to the concept digital humanities, the term humanities computing was usually used which traditionally concentrated on textual analysis. [20] This had as much to do with the limitations of computational power. Nowadays sound, images and video have been more incorporated in the field.

The digital humanities community relies heavily on open data, and the challenges are connected to the future of digital cultural heritage preservation. Open data can be defined as: "A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike." [21]

In this paper a broad interpretation of digital humanities has been adopted. As one interviewee mentioned, the digital humanities community is interesting because people may come from really different standpoints and have different notions of what digital humanities is all about, but everybody shares one "hook", the concept of digital humanities. That leads to a discussion in which professionals may share ideas more freely, not just inside their own specific circles. However, there is also a danger that digital humanities becomes an empty concept and a tactical word that people employ casually, without thinking what it means.

---

[19] Vanhoutte 2013, 125–127.

[20] Digital Humanities itself only began to be used after the publishing of *A Companion to Digital Humanities* in 2004.

[21] Open Definition, http://opendefinition.org/

2 The Researcher's Perspective

Digital history, which can be included as a branch of digital humanities, refers to the use of digital media and tools for historical practice, presentation, analysis, and research.[22] Digital possibilities help both the popular interest and the academic research in history. Although history books are read by a wide public there has always been a certain gap between popular and academic history, a gap that some scholars have argued has widened dangerously. Digital tools could also help to bridge this gap.[23]

The main change so far has been the improved access to resources, both primary and secondary sources. The digital tools have changed many day-to-day practices of historians but still the underlying research methods remain usually the same as before. This chapter concentrates on the digital humanities possibilities in the field of history. However, we see history as a very broad area of interest to the human past and not just confined to the academic discipline named history.

2.1 Access to Historical Resources and Public Interest

When asked about the growth of public interest in history, our interviewees agreed that there seems to be growth in the interest for the past but that it would be quite hard to measure and quantify. The most obvious area of improvement is the access to resources. The sheer volume of material available online has made a difference for many researchers and for the wider public.[24]

In the United Kingdom the most thriving area of popular history is family history, also called genealogy. This interest has geared archival digitisation towards family history content because of the financial value to be exploited. The digitised material helps academics too. However, those interviewees who mentioned family history also saw the private companies' role as controversial because they simply want to make money with content that should in principle be publicly available for free. The

[22] One of the most authoritative and comprehensive works on digital history is *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web* (Cohen & Rosenzweig 2005). However, since the book was published there has been a great deal of evolution in the nature of the web.

[23] Armitage & Guldi 2014.

[24] List of digital resources http://www.history.ac.uk/digital/resources.

business opportunities are manifested in the success of commercial websites such as Ancestry or Find My Past.[25]

Collaboration between publicly funded organisations has lead to popular family history television projects.[26] Also for history web portals such as British History Online, most of the users are not academics but family history researchers. [27]

The demand for public history probably existed before the growth of online content but people didn't know how to access the material. There is still a lot of naivety about the research; people think that everything is online. The percentage of people that want to learn is growing however, and for example the UK National Archives has an extensive education program. [28]

The digitisation of finding aids has been an important advancement for both the public and the academics. Archival catalogues were the first to be accessed online, which can already mean a big step forward when the researchers have knowledge of the collections before they enter the physical archives. For the US context, Ithaca S+R report from 2011 states that "Most interviewees said they are not traveling less for research because of digitized finding aids and collections, but they have been able to travel more strategically".[29] Knowledge about the collections may also invite members of the public to the archives that they previously saw as uninviting or even a bit scary.

During the last 100 years history has been captured in an unprecedented way. An enormous body of history in the form of texts, pictures, video etc. is stored in our societies, some of it in family attics but also the culture heritage organisations have vast archives that are increasingly put available online.

Different history projects have used maps as presenting history in a new way. The popular Historypin website offers a new way to put historical photographs into

---

[25] www.ancestry.com; http://www.findmypast.co.uk/. See also Prescott 2014a about paywalls in digital content.

[26] For example the BBC series *Who Do You Think You Are?* that the public broadcaster created together with the National Archives of UK. The series has been adapted to many countries including Finland (produced by the public broadcaster Yle). Who Do You Think You Are? http://www.bbc.co.uk/programmes/b007t575; Kuka oikein olet?http://yle.fi/ohjelmat/1325513.

[27] Interview. British History Online is maintained by the Institute of Historical Research. https://www.british-history.ac.uk/.

[28] The National Archives, http://www.nationalarchives.gov.uk/records/start-here.htm and http://www.nationalarchives.gov.uk/education/.

[29] Rutner & Schonfeld 2012, 10.

context by attaching them on a digital map that uses Google Maps services.[30] AddressingHistory is an online tool that enables users to combine data from digitised historical Scottish Post Office Directories with contemporaneous historical maps, which are also combined with a present day online map.[31] Old Weather is a US-UK collaborative online weather data project to digitise weather observations recorded in US log books dating from the mid-19th century onwards.[32] Bomb Sight maps the London blitz during the Second World War.[33] A Finnish example is Sodan jäljet ("the Traces of War") that maps people who died during wartime in the 20th century.[34]

The American Ithaka S+R report found out that historians mostly felt that the GIS work was one aspect of their research, and would fit into their broader narrative and a monograph. These projects were generally not intended to replace monographs, articles, or other traditional historical works. GIS does, however, add a valuable layer of interpretation to the work.[35]

The First World War centenary in 2014 is also a good example on the expansion of digital history projects. The First World War was the first major war to be filmed and there is a wealth of all kinds of media material about the war, which offers completely new possibilities for the wide range of centenary projects that have been put online.[36]

Some of the most popular history related pages online are controversial because they lack any context or attribution for the shared content. One of the most notable examples are the historical pictures on Twitter accounts like @HistoryInPics.[37] These sites may contain a risk of leading to a somewhat narrow and shallow conception of history. However, the popularity of these pages can also be seen as a wake up call for the universities and the GLAM sector to respond to the interest and engage the public in new ways.

---

[30] Historypin. www.historypin.com

[31] AddressingHistory. http://addressinghistory.edina.ac.uk/

[32] Old Weather. http://www.oldweather.org

[33] Bomb Sight. http://bombsight.org/

[34] Sodan jäljet. http://www.sodanjaljet.fi/

[35] Rutner & Schonfeld 2012, 30–32.

[36] The Imperial War Museum is leading the commemorative projects in the UK. http://www.1914.org/

[37] Onion 2014 is an article about historical pictures Twitter accounts in Slate magazine: http://www.slate.com/articles/life/history/2014/02/_historyinpics_historicalpics_history_pics_why_the_wildly_popular_twitter.html

The juxtaposition of the "bad" popular projects and serious academic work doesn't mean that good history projects are only done in the academia either. Many projects are grassroots projects that interested individuals have set up.[38] Some interviewees felt that there is a certain ivory tower attitude within the academia, which can hamper public engagement and also sharing research material.

The access to resources is not only about digitising and describing the old material. In contemporary history born digital online resources are getting more and more important. The mid-1990s are a legitimate era for history research and students born in the mid-1990s whose whole lives have been lived in the internet era are entering universities. Today's websites and online databases will be rich primary sources for the future historians.[39]

We asked our interviewees about the connection between digital humanities, open data and their societal value. The history resources are often linked to and can include government data. In the past the important data used to be a domain of politicians but nowadays the public can access it better. The politicians have less space to hide which can have powerful positive effects on democracy. In the UK the government has the Data.gov.uk service which is releasing public data in one searchable website in order to help people understand how government works and how policies are made.[40]

The interviewees thought it to be optimistic to think that the politicians themselves would considerably use the digital open data. Whether the analytical research reaches the decision makers is also doubtful but the digital ways to publish information, one of the next chapter's themes, can help in that area too. [41]

---

[38] For example Sodan jäljet or historical image projects such as What's That Picture? http://www.whatsthatpicture.com/ and Public Domain Review, http://publicdomainreview.org/. More about Public Domain Review on following page.

[39] TheyWorkForYou which primarily is designed for British citizens and voters to keep tabs on their representatives and see how they have voted. http://www.theyworkforyou.com/ Similar Finnish initiative is Kansan Muisti. http://www.kansanmuisti.fi/. See also for example DataShine http://datashine.org.uk/ which uses different kinds of UK census data (health, ethnicity, housing etc.) as an online map. Even though the statistics would have been available as spreadsheets anyway, the visual map can give researchers new ideas and raise new kinds of questions.

[40] http://data.gov.uk/

[41] Open government data is an important theme but not the focus of this survey. The Finnish Institute in London has published a report about the subject, see Halonen 2012.

The benefits of improved online access to resources are similar as the effect of open data generally. Open data is primarily a means to an end, not an end in itself. Quoting Rufus Pollock: "The real ends are the improved creation, processing and use information for the purpose of bettering our lives and the world around us."[42]

Some digital humanities scholars are keen to emphasize the barrier-breaking and revolutionary nature of digital humanities and may not be so interested in digital contents just “making things easier”. However, better access to contents may diminish barriers and problems that arise from peripheral location or underprivileged education status. Visiting the British Library is naturally easier for a researcher based in University College London than for a student from Finland or New Zealand. Also, not all of the institutions have same kind of financial resources as the world's top universities.[43]

Sometimes digital content is also freely available only behind an academic barrier, not for all interested individuals. You have to be a student or member of the staff of a university to access for example digital books, although the access is provided by the university libraries which (in Finland) are public libraries where everybody can read and borrow the physical books. The collections may be digitally accessible in public libraries and other public institutions but still the geographical barriers remain.

One way to counter the pay walls is for non-profit organisations to step in. The Wikimedia Foundation has recently addressed this concern by allowing some active English language Wikipedia editors access to a wide range of pay walled databases and the experience has so far been positive.[44]

Another new development in the United Kingdom is the alteration of the copyright law (since June 2014) that allows non-commercial researchers to legally copy material for a text and data mining analysis.[45] ContentMine is a new initiative that uses machine-reading to liberate 100 million facts from the scientific literature and make them free for everyone in Wikidata. The project is not strictly digital humanities research but same kind of methods could also be applied specifically in

[42] Pollock 2011.

[43] Cohen 2010 about the usefulness of Google Books in the higher education.

[44] Wikimedia blog, https://blog.wikimedia.org/2014/07/22/expanding-local-history-wikipedia-library/

[45] *Exceptions to copyright: Research.*

history research, especially social and economic history which use more quantitative data than some other areas.[46]

Empowering the public and bridging the knowledge gap[47] between different socioeconomic groups can improve social cohesion. In Finland high tuition fees are not (yet) a problem but geographically the country is far away from some important resources. Additionally, even within the country it can be hard for a student based in the north to access some important resources that are traditionally concentrated in Helsinki.

[46] ContentMine, http://contentmine.org/. See also Peter Murray-Rost's (2014) presentation at the Wikimania conference http://www.slideshare.net/petermurrayrust/contentmine-and-wikidata

[47] The knowledge gap was the focus of the Finnish Institute's previous survey. Jäntti 2014.

Case example: The Public Domain Review

All works eventually fall out of copyright and in doing so they enter the public domain, becoming material that everyone can enjoy, share and build upon without restriction. The Public Domain Review, founded in 2011, is an online journal and not-for-profit project dedicated to promoting and celebrating the public domain in all its variety and richness – like a pathway to a network of archives and storage rooms that lie beyond.

"With a focus on the surprising, the strange, and the beautiful, we hope to provide an ever-growing cabinet of curiosities for the digital age, a kind of hyperlinked Wunderkammer – an archive of materials which truly celebrates the breadth and variety of our shared cultural commons and the minds that have made it."

The Public Domain Review is a project of The Open Knowledge Foundation (recently rebranded as Open Knowledge) – a project dedicated to promoting open data and open content in all their forms. Public Domain Review and works closely alongside the Open GLAM initiative, that promotes free and open access to digital cultural heritage held by the GLAM sector.

The Public Domain Review has been a response to wide digitisation of cultural content. Suddenly lots of content and archives that used to be buried somewhere became available online. Nowadays there is simply too much to explore, so Public Domain Review tries to curate and highlight some of the content, offering a pathway to the more obscure material that maybe would not find its way to academic research.

The website is an example of a grassroots project that mainly has required the working hours of its editor Adam Green. The material was first displayed in a blog and then it gradually grew to a bigger project. However, when the site became bigger, there are also increasing costs. The Public Domain Review shop is one way to counter the costs, offering products made with the featured public domain content. This highlights the public domain's possibilities: anyone can reuse the contents even commercially, something that the traditional museums and libraries can be afraid of, losing "their" content to the public.

http://publicdomainreview.org/

2.2 New Tools and Methods

Beyond mere access to content, digital material may change research work substantially. For example, when texts are enriched as data sets where you can do different kind of searches, with methods such as text mining and distant reading, researchers can use larger amounts of data than in the traditional case where you

closely read paper material. Also presenting the findings can include not only texts but sophisticated data visualisations, maps etc. [48]

There certainly have been improvements in the access to resources and also indications of a growing public interest towards the past, but has this changed the fundamentals of historical research? Based on the interviews and literature[49], we would argue that the underlying core research work has so far stayed the same as before even if the digital era has already changed the day to day practices fundamentally.

UK organisations and history scholars have been active to publish digital tools and tutorials for them.[50] However, their usage has so far been usually quite low. It seems that there is a smaller number of researchers who engage all kinds of digital methods but the majority is not so eager yet. It seems that in digital history there are not yet any "killer applications" that a lot of scholars would use and which would really change history research.

Workload in the digital projects usually means pure working hours. The interviewees had not usually needed money to buy special equipment, although when presenting and preserving the research outputs, server storage can become costly at least for private small scale projects that operate outside of big institutions.

Although the technology used in digital humanities projects is nowadays usually far from cutting edge or really high-end, the learning curve for new skills such as data mining can be quite steep. For example in the Digging into Linked Parliamentary Data project the project team found that it would be quite difficult to use students as research assistants. The researchers had to spend more valuable working hours in doing things themselves, which naturally slowed metadata creation.[51]

---

[48] More about new digital humanities methods: Burdick, Drucker et al. 2012, 39–49.

[49] *Reinventing research? Information practices in the humanities* for the British humanities research practices. See also Rutner & Schonfeld 2012 which examines the changing research methods and practices of academic historians in the United States.

[50] Online courses by the Institute of Historical Research http://www.history.ac.uk/research-training/online. List and links to tools http://www.history.ac.uk/history-online/tools. Making Digital History by University of Lincoln http://makingdigitalhistory.co.uk/. Also the Programming Historian is an online, open access, peer reviewed suite of about 30 tutorials that help humanists learn a range of digital tools, techniques, and workflows to facilitate their research. The project has members from the US, Canada and the UK: http://programminghistorian.org/. See also School of Data http://schoolofdata.org/ and William J. Turkel's website, http://williamjturkel.net/how-to/.

[51] Interview.

Teaching digital tools and methods should therefore be included in the history curriculum to some extent. Ithaka S+R published a report on the changing research methods and practices of academic historians in the United States. It also recommends training in developing a dissertation proposal recognising resource constraints, in the adoption and use of research practices and methods, in the use of non-textual sources, and in the use of new forms of scholarly expression.[52]

Viktor Mayer-Schönberger, a big data and networked economy researcher at the University of Oxford, argues that in the long-term skills are not a problem in big data digital research because the area offers so many lucrative opportunities. He thinks that the access to resources remains the main obstacle. [53] So when more and more researchers in different academic fields engage in big data methods, it would seem natural that access to tutorials and the quality of training would also improve. This could also benefit the history community even though history research may not offer direct opportunities to make money.

There are also frameworks for measuring the overall digital literacy skills of the students and staff, such as Open University's Digital and Information Literacy Framework. [54] Perhaps such a framework could be modified and imported to digital history skills too. However, it is also important to remember that many historians would find it off-putting if they all had to learn to write code. This could be detrimental in recruiting the most interested people to study history.

Putting sophisticated data mining tools aside, it is clear that the day to day practices of researchers have changed with tools such as Google Search and Google Books, Dropbox, Twitter and so on. We all have become very gifted in research technology. Previously librarians did things that all people do with search engines today.

The vast amount of data calls for skills to find out the relevant material and organise the files in new ways. The widespread use of digital cameras and scanners to capture source materials is also a significant shift in research practices among historians. To be able to share this material good knowledge about copyrights and licences is needed.

---

[52] Rutner & Schonfeld 2012.

[53] Mayer-Schönberger at the Wikimania 2014 conference, London, 8 August 2014. http://new.livestream.com/wikimania/friday2014/videos/59350594

[54] Digital and Information Literacy Framework. http://www.open.ac.uk/libraryservices/subsites/dilframework/

This takes us back to the definition of digital humanities: do these tools count as digital humanities? Most would say no, but still they have changed a lot. Nowadays there are free tools that feel trivial but 10–15 years ago they were worth hundreds of pounds and considered cutting edge technology. Going forward 10–15 years from this day digital humanities may not even exist as a specific area. The nature of digital tools and online networks is constantly evolving and computational power is increasing. In its core the research objectives are the same as before and along with new skills, traditional expertise like basic source criticism is still needed.

One interviewee noted that although using digital content is not necessarily considered as digital humanities, digital information is always structured in some way. In the process of using this kind of information the reader or researcher is constantly doing some kind of analysis about the infrastructure which will change our perception of the media.

When historians have traditionally used paper newspapers or archival files, they may have found things that are not strictly related to the research subject but give meaning to the context. For example by reading historical newspapers the reader may glimpse through the other news articles, advertisements and such. When flicking through an archive file the researcher can find interesting new findings next to the document he/she was really searching for. This kind of serendipity may be lost in the digital search query process, but it remains to be seen whether more time-efficient search methods balance this lack of serendipity.

Case example: Digging into Linked Parliamentary Data

The Digging into Linked Parliamentary Data project (Dilipad) delivers a common format for encoding parliamentary proceedings in different countries. Focusing on the UK, Canada and The Netherlands, it is a collaboration between the University of Amsterdam, the History of Parliament Trust, the Institute of Historical Research, University of London, King's College London, and the University of Toronto.

Parliamentary proceedings reflect historically significant events from centuries ago to the present day. They exist in a common format that has survived the test of time. With parliamentary proceedings becoming available in digital form in many countries, new research opportunities arise to analyse this data, on an unprecedented longitudinal scale, and across different nations, cultures and systems of political representation. Replicating the paper proceedings in digital format is just the beginning and not enough. With digital interfaces and metadata the proceedings can be enriched to enable completely new kinds of research.

A key feature of the Dilipad project is its use of the XML schema PML (Parliamentary Metadata Language) as its core metadata format. The goal of the project is to enable large-scale, transnational, cross-language, comparative, diachronic research on the proceedings of parliament. Specifically:

1) to enhance the existing corpus of parliamentary data, using natural language processing and linked data;
2) to develop new and adapt existing tools which will allow the comparative, longitudinal study of the enhanced data; and
3) to explore substantive research questions which will both test and inform the development of those tools, in an iterative process of co-design.

Comparative analysis of this kind, and the tools to support it, will inform a new approach to the history of parliamentary communication and discourse, and address new research questions. Quantitative methods can be used along with traditional qualitative methods. The researchers can use big structured data sets and for instance classify cases where some words or phrases appear. The explicit documentation of these queries also helps to open the historical research process, and in principle the research could be replicated in the future. (See also Ihalainen 2013; van Eijnatten, Pieters & Verheul 2013.)

According to Ihalainen (2013), United Kingdom, Ireland, the Netherlands, Belgium, and Austria are currently in the forefront of parliamentary digitisation. Finland has fallen behind the top countries. The plenary protocols of the Parliament of Finland are available online only from the year 1989 onwards.

The digitisation of the parliamentary proceedings would benefit significantly the Finnish political culture research. This would not come without costs but it should be seen as a long time investment. The Digging into Linked Parliamentary Data project's data scheme can be seen as one of the ways of how the digitisation could benefit from international cooperation, as the common format is meant to be relatively easy to export to different languages and parliamentary cultures.

More information: http://dilipad.history.ac.uk/

2.3 Sharing and Preserving

One important strand of digital scholarly resources is research data management. Traditionally humanities research notes could be scribbled notes on paper, content that is not really feasible to be opened and published. Nowadays many scholars use digital databases and content that could easily be included or linked to the output of research. However, the attitude towards opening research data varies a lot and humanities research data is often still conceived as the lone researcher's own domain.

The Expert Advisory Group on Data Access published recently a report about research data sharing, mainly in the UK context.[55] The UK Data Archive provides services for humanities and social sciences research data archiving.[56] For structured machine-readable data, Wikidata will also be another important possibility to store and share research data.[57]

In Finland the Finnish Social Science Data Archive[58] provides similar information and services but it is still not used to a great extent in the humanities field. There has recently been discussion about opening research data in Helsingin Sanomat, the main daily newspaper in Finland.[59] The Open Science and Research project will also hopefully steer the Finnish academic community into opening up the research practices.[60]

One of the main products of the digital era is that historians and scholars of other disciplines do not simply output a text and hope that somebody publishes it. They can publish it immediately online. This change may lead to a certain chaos and lack of authority (which I noted when searching for literature to help with this survey) but the creative chaos could also act as a leveller and increase democracy. When articles are published in blogs, students and the public can immediately react to them. This is not the case with prestigious peer reviewed journals. How this new nature of

[55] *Establishing Incentives and Changing Cultures to Support Data Access.*

[56] UK Data Archive, http://www.data-archive.ac.uk/

[57] Wikidata. www.wikidata.org

[58] Finnish Social Science Data Archive, http://www.fsd.uta.fi/fi/

[59] Wilhelmsson & Keinonen 2013 and the reply: Toivonen & Lehtomäki 2013.

[60] Open Science and Research, http://openscience.fi/

scholarly publishing affects the knowledge gap between socioeconomic groups would be an interesting area for further research.[61]

For publishers this change can mean a loss of profit and a need to change their ways of doing business. Of course there have been other media revolutions before but the internet is completely changing the ways of publishing. Publishers can also publish open access literature online but they have been charging the scholars high prices for getting the research published, sums that maybe only tenured professors in high-profile institutions could afford.[62]

Although this chapter has been about using the digital contents and tools for reading and writing, the new role of historians and other researchers providing access to scholarly works reaches beyond blogs, Twitter and free online journals. There are many cases where historians have used the network to build their own online exhibitions, presenting and preserving material that has previously been buried somewhere either in the attics or deep in the web. This is a new way of creating public history and also closely connected with the future of digital preservation, traditionally the playground of the GLAM sector and the topic of chapter 3.

2.4 Narrowing Gaps and Building Bridges

Digital humanities can be a way for researchers to:

- Increase cross-disciplinary and international collaboration and comparison in research projects.
- Engage the public better.
- Bridge gaps inside the academia, for example between the humanities and social sciences.
- Collaborate with the GLAM sector.

Digital humanities has often been said to change humanities research practices to include more collaboration and cross-disciplinary work.[63] There are many international projects including UK contribution such as the Dilipad project (see case

[61] More about open access publishing: Cohen 2011; Open Access Working Group of Open Knowledge, http://access.okfn.org/; Wikiproject Open Access https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Open_Access

[62] Melissa Terras (Professor of Digital Humanities, University College London) about publishers' open access policies: Terras 2013. http://melissaterras.blogspot.co.uk/2013/11/im-not-going-to-edit-your-10000-pay-to.html

[63] McCarty 2012; Spiro 2011. See also Scheinfeldt 2012.

example) and CENDARI (Collaborative European Digital Archive Infrastructure), a European Union funded research infrastructure project aimed at integrating digital archives for the medieval and World War One eras.[64]

Does this collaboration include the day to day research or is it just looser co-operation? The interviewees were a bit sceptical whether the collaboration in projects has really yet changed the core research work. Much of the work is team based and for example Twitter is used to a great measure in scholarly communication but the core research is still usually done alone. This notion is supported by a recent article which argues that single-authored papers are statistically predominant in the digital humanities field.[65]

The so called digital humanities projects can also include for example pure linguists, pure historians and pure computer scientists. Whether all these people would call themselves digital humanities professionals is questionable. That is of course more of a conceptual question than really about the nature of work.

Public engagement is an important part of the humanities projects. In the UK public engagement for the higher education is supported by the National Co-ordinating Centre for Public Engagement.[66] David Armitage and Jo Guldi argue that as academic history has become more and more specialised, the time periods and themes that historians focus on have become narrower and narrower. Because of this evolution the historians have been cut from wider reading public and deprived of the influence on public policy and global governance they once had. Due to the increased availability of large amount of historical data and the digital tools necessary to analyse it the historians could once again examine longer time periods and make more publicly engaging research.[67]

Wikipedia is another important, although controversial, meeting point of academic and popular history. For the public it can be the only information consulted, for more serious research it can serve as an entry point to sources. Wikipedia is one of the most important manifestations of how the academics' monopoly in information production has been eroded: no one needs a PhD or even a bachelor's degree to edit

[64] CENDARI. http://www.cendari.eu/

[65] Nyhan & Duke-Williams 2014.

[66] National Co-ordinating Centre for Public Engagement. http://www.publicengagement.ac.uk/

[67] Armitage & Guldi 2014.

Wikipedia articles. There are of course problems, such as the inherent bias or neutrality of the editors or simply the lack of good articles in smaller languages such as Finnish.[68]

This could however be an exciting possibility for history graduate training, teaching both the source criticism and the editing sides of wiki projects. Wikipedia and Wikidata information can also be used to present history in new ways: for example the UK based Histropedia company is making historical timeline visualisations using data from the online encyclopedia.[69]

Crowdsourcing, the process of leveraging public participation in or contributions to projects and activities, is an important tool for the academics to reach to the public and engage it during the research process. Success in crowdsourcing projects is as much about the communication as it is about the infrastructure. Just telling the public that "we need this to be done" leads nowhere, but if the public has a feeling that they are a community that is building something, it can lead to a wide success. The most successful crowdsourcing projects are often instigated by the GLAM sector. Collaboration between the universities and the GLAM sector would be beneficial because the GLAMs are naturally public facing and they can act as a channel for the academic research.[70]

Digital humanities can also help with bridging the gap within the academia, for example between the humanities and social sciences. Political and economic history are interesting themes because they are in the borderline between the humanities and social sciences. Lisa Spiro has recently discussed the collaboration possibilities of digital humanists and digital social scientists. Spiro acknowledges history (along with other disciplines such as anthropology) as one of the key areas where social sciences and humanities are linked. "By developing a deeper awareness of how social scientists use computational methods to address research questions, humanists might gain new insights into how they can apply similar techniques to their work — and vice versa."[71]

---

[68] About historians and Wikipedia from an American perspective: Campbell 2014; Cronon 2012.

[69] Histropedia, http://www.histropedia.com/

[70] Dunn & Hedges 2012.

[71] Spiro 2014. About computational projects in humanities and social sciences mainly concentrating to US projects see also Williford & Henry 2012.

Collaboration could help both digital humanists and digital social scientists to train new scholars to use digital tools and perhaps use the computational equipment and the services of computer scientists together. Both humanists and social scientists need to ensure that the digital strand of research is recognised and supported inside the academia.[72]

Besides social sciences, arts & design are also important areas where digital humanists can endeavour. Willard McCarty calls for closer examination what the artists would have to say on digital humanities projects, [73] whereas in the *Digital_Humanities* textbook the possibilities of design are emphasised. The addition of graphic supports such as charts, graphs, and animations are often essential in making an argument in digital humanities.[74]

Many of the digital humanities projects are also international. When reading about the projects it is sometimes not even clear whether they were instigated in Britain, United States or other English speaking countries. For the Finnish context the language is an obvious challenge to include foreign researchers to deal with the Finnish language content itself, but many frameworks could be easy to import and export.

2.5 Measuring the Impacts

Practically all our interviewees noted that measuring the societal impact of digital projects has proven to be difficult. There are no numbers widely available but in a few years the situation can change when more projects and evaluations are made, a feasible topic for further research.

One problem in measurement is citation and referencing. People still refer to the digitised books and other content as if they had read the physical copy. Hence beyond page views it's hard to measure the impact of the digital editions. Digital citations can also be very long, making it frustrating to write and read them. The publishers may also delete links from print editions.[75]

---

[72] Spiro 2014.

[73] McCarty 2014.

[74] Burdick, Drucker et al 2012, 12–16.

[75] About citation in digital humanities: Blaney 2012.

Also when citing an URL there is always a possibility that it quickly becomes obsolete. Persistent identifiers for archival resources have been developed and tested both in the UK and Finland, which is one way to encounter this problem.[76] There are also different online tools for measurement of the usage and impacts of digitised scholarly resources, such as the TIDSR, Toolkit for the Impact of Digitised Scholarly Resources, created by the University of Oxford.[77]

However, some of our interviewees argued strongly that quantitative measurement or international comparison of these projects is mainly pointless, and the focus should be in quality. Research in history and other humanities subjects is often organised nationally. The impact of digital resources and research written in the Finnish language that concentrate on the Finnish context cannot be directly compared internationally, as simply counting page views and number of references tells us nothing. The quality of metadata and the user interfaces should be measured.[78]

Although measurement can be tricky, as one interviewee noted, digital tools provide a concrete possibility to link the research to the wider society and also engage in international collaboration, which should be actively utilised.

[76] In the United Kingdom research in this area has been done for example in the UK Data Archive (Woollard 2012) and in the Data Curation Centre in Edinburgh (Davidson 2006). The National Archives Service of Finland has a project that is currently in pilot phase, and should be ready for wide public use in 2015 (Lahtinen 2013).

[77] TIDSR: Toolkit for the Impact of Digitised Scholarly Resources. http://microsites.oii.ox.ac.uk/tidsr/welcome

[78] Hupaniittu 2012, 90–91. The same argument was found in our interviews too.

3 The Preserver's Perspective

3.1 Future Archives, Utopia and Dystopia 2030

Computer technology and the internet are forcing the cultural heritage sector to think about how they want to be represented on the web and how to deal with digitised analogue works. This evolution has been acknowledged in European Union wide strategies.[79]

The digital humanities research is closely connected with digital preservation of data. Last decade has seen a tremendous increase in digital preservation tools and services, delivering great benefit in ensuring the longevity of digital data. As the previous chapter already illuminates, digitising paper material has become an important aid for humanities research. Digitisation and the Internet improve the access and usability of the material, but it can also help preserving old and frail papers, tapes etc.

In addition, digital preservation is not only about digitising analogue content. Contents are increasingly born-digital, and include data sets that are expanding in scale, complexity and importance. Although this chapter mainly uses the word 'archives' when addressing digital preservation, the possibilities and challenges of digital preservation are similar throughout the GLAM sector.

We asked the interviewees what they thought would be the best and worst scenarios, utopia and dystopia, for the digital preservation and digital humanities field in year 2030. There seems to be a clear drive for open knowledge which is however not without challenges. The future is linked to the present state and these challenges need to be addressed now. The most common answers were:

Best-case scenario:

- Everything digitised and available online free, with good infrastructure:
    - linked with metadata.
    - theorised properly.
    - clear rights.

[79] Pekel, Fallon & Kamenov 2014. See also speech by Neelie Kroes (2014), the Vice-President of the European Commission responsible for the Digital Agenda in Open Knowledge Festival, Berlin 2014.

- Archiving the web will be much better and more important. Rich data about the last decades.

Worst-case scenario:

- Archives are overwhelmed and outpaced by the technological change.
- Digital black hole: Lost information in the change beginning from the 1990s.
- Physical collections are neglected and underresourced as the emphasis shifts to digital – major uncatalogued backlogs remain.
- Private companies own the data.
- Everything is behind paywalls.
- GLAMs are closed silos that don't talk to each other.
- Poor economic situation, not enough resources for humanities and social sciences research.

3.2 What is Available and how?

Mass digitisation and online availability of widely-consulted and publicly funded collections was mentioned as the best case future scenario by nearly everyone. The interviewees however acknowledged that in reality there will probably never be enough resources to digitise all cultural heritage. In the utopia there would also be a wide common understanding of why the archives should be open.

The rapid increase in online digital content has provided public access to an unprecedented amount of material. However, the sheer amount of data calls for someone to curate the material. Content that used to be buried somewhere is suddenly available but to find the relevant information, the professional skills of archivists and librarians are needed. The archivist is critical in the history research process as a pathway to the content. In the future researchers may visit the physical archives less and less. That is why archives should focus on good online and telephone services, and for example tutorials on their websites.

Just changing the papers to bytes is not enough. A vital part of the digitisation is preserving the original context of the documents, improving usability with different kinds of metadata and interfaces and building new links to other online material. Lots of data is not big data (to use the fashionable term in digital research) if the data is not properly linked and people cannot access multiple things at one time. In this area collaboration between different organisations is vital: one of our interviewees

suspected that common standards may be too difficult to achieve but even a common vision and a common way of referencing the collections would help in collaboration. [80]

The joint European search portal Europeana[81] is a good example of a common interface for the GLAM sector. Still, to make serious research possible via the interface, lots of resources are needed to build enough critical mass. In Finland the National Digital Library project is participating in the development of Europeana. The National Digital Library aims to ensure that electronic materials of Finnish culture and science are managed with a high standard, are easily accessed and securely preserved well into the future. It is one of the key electronic research and culture infrastructures currently under construction in Finland.[82]

The National Archives of UK has online guidelines and tools for digital preservation of records. [83] In the United Kingdom the Digital Preservation Coalition also acts as a collaboration platform for bringing people together to enable higher quality research, better informed standards and faster deployment of solutions. It was established to foster joint action to address the urgent challenges of securing the preservation of digital resources in the UK and to work with others internationally to secure our global digital memory and knowledge base. Its members are both from the UK and Ireland. They are also involved in joint Europe-wide projects.[84]

The Digital Preservation Coalition argues that

> *"...the fragmentation and impenetrability of digital preservation research is now an impediment and disincentive to those who need solutions urgently. There are growing risks that tool developers find it hard to reach a market meaning their solutions are under-deployed and their investment under-exploited; standards developers struggle to achieve consensus meaning their approaches are under-consulted or ignored; and problem owners find it hard*

[80] For example in the DILIPAD project that enhances digitised parliamentary proceedings, the MPs are given personal identity numbers. However, different organisations in the UK have given different identifiers so there are problems to put the systems to work together. See also *Digital Preservation Handbook*, and Jeurgens 2013, 33–34.

[81] Europeana. http://www.europeana.eu/

[82] The National Digital Library. http://www.kdk.fi/

[83] http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/

[84] Digital Preservation Coalition. http://www.dpconline.org/

to locate solutions, increasing the short-term risks of data loss and the long-term costs of deployment."[85]

There have been different standpoints to what the digitisation of information means for more marginal data. A positive and somewhat utopian standpoint is that when everything is digitised and online, previously neglected groups get their voice heard better.[86] However, some of our interviewees as well as literature have raised concerns that limited resources may direct digitisation towards the traditionally "important" material such as government politics, neglecting the vast collections that stand beside them.[87]

This is an important issue. Because resources are limited, there will always be decisions made on what is digitised and what not. Born-digital content may also seem virtually easy to store on hard disks but there are still substantial preservation costs.

The GLAM sector doesn't operate in a political vacuum and moving to the digital era doesn't mean getting rid of the politics and power influences. Politics are always present in the decisions on what is to be archived, preserved and digitised. According to Elizabeth Shepherd, professor of records management at UCL, the British archivists should be more proactive and need to speak up to make their voice heard in the society.[88] Perhaps new digital tools and contents could be a method for demonstrating the value and impact that the archives have for the society, facilitating transparency and preserving societal memory.

When choosing content to be digitised, widely consulted collections are an obvious choice. Copyrights and the sheer effort of digitisation (because of the scale and nature of different collections) have a big influence too. As mentioned in the previous chapter, online content may distort research to the areas that are digitised – traditionally "important" topics – and also because of copyrights towards material that is in the public domain or in other way easily available. Researchers may neglect everything that is not online as non-existent.[89]

---

[85] DPC Strategic plan 2012–2015, 2. http://www.dpconline.org/.

[86] Flanders 2009.

[87] See also Andrew Prescott (2014b) about the digital humanities projects in the UK so far enforcing a traditional and conservative view of humanities.

[88] Shepherd's inaugural lecture, UCL, 20 May 2014.

[89] This is not to say that the situation was different in the analogue era: For example the opening of the Soviet archives in the beginning of the 1990s led to a surge on Soviet Union and Russian history, and many important

Different organisations also have different background political influences and financial resources. For example historical reasons have influenced German institutions to consider it important to digitise material that documents the country's Jewish heritage. In the former socialist states of Europe the digitisation projects have often highlighted recently blossoming nationalism and the historical roots of this patriotism.[90]

Besides digitisation, born-digital material presents huge challenges for archiving. Information is increasingly stored only for example as web pages that constantly change. Archives could be overwhelmed by the mass of born-digital records and outpaced by the technological change if they lack the resources to preserve, catalogue or make the collections accessible.

One of the interviewees noted that it's actually quite easy to appear relevant and to look like staying in the forefront of the digital turn, but behind the surface the challenges are great and new skills are needed. For institutions where the (digital) preservation and access to documents are core tasks it seems to be planned accordingly but if the archiving is just in a support role, it may easily be neglected because of limited funds and skills.

Archiving the UK web is done by the British Library.[91] The situation is similar in Finland where the National Library maintains the Finnish web archive.[92] The Big UK Domain Data for the Arts and Humanities project's objective is to highlight the value of web archives as a source for arts and humanities researchers, and to transform the way in which those researchers interact with the data.[93]

The national libraries' core task is to preserve the published works in the respective countries. However, now in the digital age one could argue that the roles of archives are blending into each other, which is manifested in the terminology: the web archive is run by the library. The British Library also has a project called Endangered

---

colonial archives were previously in East Germany, so some of the best research on German colonial history was done in the GDR until the reunification of Germany.

[90] Zaagsma 2013, 3–29; Jeurgens 2013, 30–54.

[91] UK Web Archive, http://www.webarchive.org.uk/ukwa/

[92] Finnish Web Archive. http://www.nationallibrary.fi/services/digitallegaldepositmaterials/webarchive.html

[93] Big UK Domain Data for the Arts and Humanities. http://buddah.projects.history.ac.uk/

Archives which digitally preserves archival material that is in danger of destruction, neglect or physical deterioration world-wide.[94]

Preserving emails is another crucial matter.[95] When organisations began to use email as the main correspondence tool in the late 1990s, the document management systems were not capable of archiving all this data properly. There has been lots of improvements in the area but not all organisations have implemented sufficient measures to preserve the emails as evidence of their activities. The resulting 'digital black hole' or 'digital dark age' can be seen as a threat to contemporary history research. Are the archives of the beginning of the 21st century empty?

Copyrights are an important issue to be addressed.[96] For digitisation projects it is always easier to use material that is not protected by restrictive copyrights. In some cases the copyrights issues can be really complex and challenging but a lack of clarity combined with lack of specific expertise can result in an organisational policy, which restricts the use of digital objects by incorrectly claiming copyright and imposing restrictive licensing on using images of public cultural heritage works.[97] The rules vary by country but the common EU legislation also opens possibilities for collaboration.

Free and open archives have been the dominant ethos in Western Europe since the French Revolution. In principle the traditional archives are open but access can be difficult because of physical location or skills needed. Now in the digital era access can be much easier but the urge for public institutions to engage in partnerships with private companies may change the ethos of openness. This is addressed in the next subchapter.

Free entrance is also standard practice in all UK National Museums. Free can mean zero-cost but also free as in freedom. Freedom to use the collections in research and creative work is something new, and pushing the archives and collections through the door can be a bold step for the public memory preservers that have traditionally seen

---

[94] Endangered Archives Programme, http://eap.bl.uk/

[95] Prom 2011. See also Pennock 2006.

[96] More about this theme in the UK context: Archives and Copyright. http://www.create.ac.uk/archivesandcopyright/

[97] Pekel 2014a. Examples: British Museum http://www.britishmuseum.org/about_this_site/terms_of_use/free_image_service.aspx, Tate museum http://www.tate.org.uk/about/who-we-are/policies-and-procedures/website-terms-use/copyright-and-permissions

their role as preserving the physical documents and artefacts in their buildings. This requires a change in the mindset of the institutions.

Opening up the collections is a theme that has been addressed by the OpenGLAM network, initiated by the Open Knowledge Foundation and active in different countries (Avoin Glam in Finland). In Ireland the Inspiring Ireland project does similar work.[98] Now there are also new tools to measure the openness of GLAM organisations.[99]

Although open cultural collections are an ideal, simply opening all data or digitising everything is not a practical real-life solution for all archives. Releasing pictures of art collections is completely different than releasing children's personal health information. Government data and cultural heritage are in many ways linked but the picture is not simple. These decisions should be carefully planned.

[98] OpenGLAM. http://openglam.org/; AvoinGLAM. http://avoinglam.fi/; Inspiring Ireland. http://www.inspiring-ireland.ie/

[99] OpenGLAM Benchmark Survey. https://outreach.wikimedia.org/wiki/GLAM/OpenGLAM_Benchmark_Survey

Case example: British Library and a Million Images Online

In 2013 the British Library released over a million images onto Flickr Commons. These images were taken from the pages of 17th, 18th and 19th century books digitised by Microsoft. The images are in public domain for anyone to use, remix and repurpose. The images have already seen creative use for example in the artworks for the popular Burning Man festival in the United States.

The images were scanned in a project called Mechanical Curator. It is an experimental creation from the British Library Labs project, part of the Digital Scholarship department of the British Library. On the hour, it selects and uploads an image taken at random from the pages of digitised works held at the library.
The pictures are chosen at random, selected by a computer algorithm and their selection is in no way meaningful.

According to a British Library statement: "The release of these collections into the public domain represent the Library's desire to improve knowledge of and about them, to enable novel and unexpected ways of using them, and to begin working with researchers to explore and interpret large-scale digital collections."

The releasing of these images can be seen as a bold step into the darkness by the British Library. They decided to publish the pictures without knowing what these automatically scanned images depicted. Some of the material could be overtly offensive, for example displaying 19th century's racist attitudes.

However, what the British Library so far has not done is to publish all its online galleries as public domain. There are of course copyrights issues but the library also has many old master pieces such as Leonardo Da Vinci's notebook that could in principle be released.

More information:
British Library on Flickr: https://www.flickr.com/photos/britishlibrary/
British Library Wiki for curating Public Domain digital content: http://blpublicdomain.wikispaces.com/home

3.3 Costs of Digital Projects

The costs and benefits of digital projects need to be addressed when planning digitisation and long-term digital preservation. We found out that:

- Skills and attitudes have a big influence on digitisation projects.
- The costs of digital projects mainly go to salaries.
- Open source software is free but communication with ITC specialists can be difficult.
- To succeed, digitisation needs to be thought as a long-term investment.
- There are substantial costs in digital preservation.
- The analogue material cannot be forgotten.

The interviews suggest that the costs and benefits of digitisation projects seem to heavily relate to people's skills and competences. There is also an on-going British survey about costing and budgeting digitisation projects conducted by the UCL Centre of Digital Humanities which will hopefully present a clearer picture about this theme.[100]

Nowadays there are lots of open source software that do not impose costs or restricting licenses on the users. Open source software can be tweaked to run optimally in different kinds of projects, although this requires smooth collaboration between the humanities and technology sides of the projects. This kind of communication was seen as difficult to achieve.

The digitisation itself can in many cases be automated to a great extent, but the workload of the projects is still very work hour intensive. Humans are needed for example to make sure that the metadata is good enough for long term use.

The technology used in most digital humanities projects is no cutting edge or very expensive, and the costs of digital projects are mainly going to salaries. The lack of time (or insufficient funding to pay somebody else) was seen as the main obstacle in the projects. Digitisation projects were seen to lack money to properly add metadata, not technology. In the future metadata technology like OCR (optical character recognition) software will become better and better and for example machine reading

[100] Costing and Budgeting Digitization Projects, https://opinio.ucl.ac.uk/s?s=31273

of handwriting could revolutionise digitisation processes. However, this progress still needs time.

For smaller institutions like local archives the initial cost of equipment such as scanners or cameras can also be a burden. This is another point that calls for collaboration possibilities and long term planning. The investment to digital projects should be seen as a long term investment, which calls for development in skills and infrastructure. Digitisation can be costly, opening data requires new skills and also new attitudes, and sometimes there is a fear that opening data leads to lost profits even though the reality can be quite the opposite.[101]

Digitisation projects do not end in the moment when the content is published online. Long term digital preservation includes considerable costs, which should be considered when budgeting the projects. Preserving for example government data just in digital format is of course different than presenting a digitised cultural heritage collection online but once again the challenges are similar throughout the sector. The responsibility about the content should be transmitted also in case of organisational changes.

With all this talk about the digital it would be easy to neglect the analogue material. In the following decades the archives will still have a huge amount of analogue material that cannot be digitised. When digitising content, information can be lost too. For example old documents can be dated by examining the paper quality, handwriting, ink or even the smell of the paper. If digitisation is done poorly, it is also difficult to distinguish the notes in the margins, possibly by several different authors.[102]

As has been to this day, there will be costs in preserving this data. If there is a new gap emerging between the digitised and analogue content is there a risk of researchers not bothering to use the analogue content? In the future we need to think about new ways to link the analogue material to the digitised collections so that they will not be forgotten.

---

[101] In the Netherlands the Rijksmuseum released all its public domain content online with the highest possibly quality without any restrictions. This didn't lead to lost profits: on the contrary, the fact that the Rijksmuseum is so well known for their open access policy has made getting project funding easier. Pekel 2014b.

[102] Jeurgens 2013, 30.

3.4 Partnerships? Engaging Companies, the Public and Other Countries

For publicly funded organisations that face scarce resources, there has been pressure to engage in partnerships with private companies in digital projects.[103] The way the organisations are funded does not leave much room for experimentation outside their core tasks. So sometimes for them it is much easier to pay private companies and their IT experts to build digital archives, usually at a vastly reduced price, than to really think what would be valuable with people from the academia and cultural heritage preservation specialists.

The public-private relationships were seen as controversial or problematic by most interviewees. However, there were also some more optimistic evaluations. The public and private institutions want to solve the same problems but have different reasons. It seems that without public-private partnerships, the digitisation of cultural heritage objects will be a lot slower. Nevertheless, the results of these partnerships can restrict creativity that would be enabled by making cultural heritage openly available online.

As a broader question it can be argued that the publicly funded collections are a public investment that has accumulated during even hundreds of years. The private companies have not made this investment, so giving the public good in their hands could be seen even unethical. How far can you go with private partnerships without hurting your own agenda?

For example Google Books was seen as a project where the metadata is not as good as it could be, but as Google has already digitised collections from many great libraries such as the Bodleian Library of the University of Oxford, nobody wants to do it again because of the costs.[104]

All private companies cannot be directly labelled as representing the same motives either. Google is different from many other private companies because they want to acquire as much knowledge as possible and then exploit that. The content is seemingly free for the users, whereas companies such as Ancestry want to simply make people pay for content.

---

[103] The European Union *Digital Agenda 2020* promotes public-private partnerships. https://ec.europa.eu/digital-agenda/en/digitisation-digital-preservation

[104] About Google Books metadata see Jon Orwant's comment at http://languagelog.ldc.upenn.edu/nll/?p=1701#comment-41758. Some of the limitations of the metadata are structural but there also thousands of errors in the data. However, mass digitisation always involves a trade-off between quantity and quality. When digitising millions of images, errors will always occur.

The National Archives (TNA), the UK government's official archive, and the publisher Adam Matthew have a joint digital project called Archives Direct. The project has involved the scanning of millions of British government documents from the 19th and 20th centuries, relating to a wide variety of global regions and focusing on British diplomatic history and international relations. Additionally, major projects have been struck between TNA and genealogy sites Ancestry.com and Findmypast.co.uk. These projects have helped TNA to digitise its contents rapidly but the content is available only behind paywalls.[105]

The historical newspaper websites offer a good case for comparison. The Finnish historical newspaper database is free to use for everyone online. The newspapers that are still under copyright are being digitised in collaboration with the publishers.[106] In the UK the British Newspaper Archive is owned and run by brightsolid Newspaper Archive Limited. It is a commercial partnership with the British Library. The material is behind paywalls and accessible only by subscription.[107]

The digital era and networks make the relationship with the public different. The GLAM institutions and public broadcasters like BBC or Yle make the decisions and broadcast the content to passive receptors. The network thinking immediately raises questions about feedback and the public can curate their own interpretations of the media.

Crowdsourcing projects are an important example of reaching the public in new ways. At least in Finland the GLAMs have been suspicious about using crowdsourcing methods in enhancing their metadata, fearing that the quality doesn't meet their standards.[108] However, as already mentioned in chapter 2, Dunn & Hedges (2012) argue that many of the best crowdsourcing projects have in fact been instigated by the GLAM sector because they are naturally public facing compared to the higher education institutions. Also the Finnish historical newspaper metadata project Digitalkoot is a good example of a successful crowdsourcing project. [109]

---

[105] Fogg 2014; Archives Direct. http://www.amdigital.co.uk/m-collections/collection/archives-direct/

[106] DIGI - National Library's Digital Collections. http://digi.lib.helsinki.fi/sanomalehti?language=en

[107] The British Newspaper Archive. http://www.britishnewspaperarchive.co.uk/

[108] Hupaniittu 2012.

[109] Digitalkoot. www.digitalkoot.fi. More examples: Old Weather. http://www.oldweather.org/ ; Operation War Diary. http://www.operationwardiary.org/ More about crowdsourcing in humanities, see Dunn & Hedges 2012.

Besides academic research and semantic metadata creation the archive material can be used in artistic, creative ways.[110] With easy-to-use web tools such as Mozilla's Webmaker, the public can create their own interpretations straight out of the network material too.[111]

[110] For example filmmaker John Akomraf has used archive footage extensively, creating new interpretations, see interview, Salt 2013. The British Library's public domain images were used in the artwork for 2014 Burning Man festival: http://www.davidnormal.com/#!crossroads-of-curiosity/ccw3

[111] Mozilla Webmaker. https://webmaker.org/

4 Conclusions and Suggestions

In the light of the interviews and current research, the main conclusions of this digital humanities survey were:

- Access to digital content online has benefited researchers and can improve geographic and socioeconomic equality.
- Digital history may narrow the gap between academic and popular history, and increase cross-disciplinary and international collaboration.
- New digital tools have not yet changed most historians' research methods substantially but digital tools and methods should be included in history curriculum.
- Libraries, archives and universities see future digital preservation challenges similarly, which encourages further collaboration.
- Digitisation projects are work hour intensive and require new skills and attitudes.
- Private-public partnerships in the culture heritage sector are in many ways controversial but may offer opportunities for collaboration.
- Measuring the societal value of digital humanities is difficult but might be possible in the near future.

The main change for researchers has been the wide availability of digital resources. Whereas the digital era has changed the day-to-day practices of researchers substantially, not many historians still engage in deeper level computational methods and tools. Nevertheless, historians and other researchers and activists have built lots of digital history projects online.

The literature and interviews suggest that the areas of digital humanities, open knowledge and open data are in many ways connected. The main idea is linking the digital humanities scholars and digital preservers with the wider community. The GLAM sector was seen as an important channel for academic research to reach the wider public. The archives and libraries need to deal with similar problems, diminishing the barriers between their fields.

The findings of this survey are in many ways similar to the Finnish research *Tutkijoiden ääni* (Hupaniittu 2012), the British *Reinventing research?* report (2011)

and the American Ithaka S+R reports (Rutner & Schonfeld 2012; Maron & Pickle 2014). They all promote collaboration within the academia and between universities and the GLAM sector.

What we would like to add is the role of non-profit organisations such as Open Knowledge or Wikimedia and the important contribution that grassroots activists have in creating and disseminating digital content. These organisations and individuals could offer important information eg. about copyrights for the GLAMs and researchers. The copyrights rules vary by country but common EU legislation also opens possibilities for collaboration.

From a Finnish perspective international collaboration and networking are especially vital to get a critical mass of ideas flowing. In Finland there are fewer universities and the government and GLAM sector organisations are smaller than in the United Kingdom. The situation calls for even more collaboration to make most use of the resources available. The Open Cultural Data and Open Government Data courses are some promising projects.[112]

The small organisations in Finland could be an asset too. In huge organisations like the British Library changing a strategic course may not be so agile. Collaboration in the field is a good way to make sure that the different organisations' metadata can communicate with each other, and organisations do not end up as silos.

Publishing cultural heritage online without restrictions helps researchers, the public and the cultural heritage organisations themselves. There are promising examples and these policies are emphasised in European Union strategies.[113] Europeana is an important portal to digitised material but not a solution to all needs in its present form.

The new possible gap between online material and analogue content that is not digitised needs to be addressed too. There is a concern that the analogue material will become invisible for the researchers and the public, thus distorting research interests and processes.

[112] Avoin kulttuuridata -mestarikurssi. http://fi.okfn.org/koulutus/avoin_kulttuuridata/, Avoin julkishallinnon data -mestarikurssi http://fi.okfn.org/koulutus/avoin-julkishallinnon-data/ .

[113] Pekel, Fallon & Kamenov 2014.

Politics will always influence cultural heritage preservation. Archivists have so far been somewhat invisible in public discussion. They should speak up and demonstrate how archives are valuable in preserving societal memory.

There is a certain fear for private sector domination in public-private partnerships, but collaboration between the two can be fruitful too. It is worth asking, how far can you go with these private partnerships without hurting your own agenda? The private companies may want to put content behind pay walls or exploit it otherwise. Open and honest discussion between public and private organisations is needed.

Data sharing and open access publishing changes academic publishing. How this new nature of scholarly publishing affects the knowledge gap between socioeconomic groups would be an interesting area for further research.

Measuring the societal value of digital humanities projects has proven to be very difficult and direct comparisons between the impacts of projects in different countries and languages would be challenging. Some interviewees argued that quantitative measurement and comparison would be quite pointless. However, with new frameworks and toolkits along with more accurate statistics coming out in the near future, further investigation in that area should also be done.

This report presented the best and worst case scenarios for digital preservation in the year 2030. What kind of policies lead to the utopia or dystopia situations is also an important topic for further research and discussion.

This paper has presented some of the questions on a more philosophical level. Trying to provide concrete answers for practical questions would not be feasible in a report like this one, but we hope that the presented ideas, referenced literature and projects circulate good ideas and encourage discussion and collaboration in and between Finland, the United Kingdom and Ireland.

References

Literature

Armitage, David & Guldi, Jo (2014), "The Return of the Longue Durée: An Anglo-American Perspective". In Annales. Histoire, Sciences sociales 69. Available online: http://scholar.harvard.edu/files/armitage/files/rld_annales_revised_0.pdf (Retrieved 29 July 2014).

Blaney, Jonathan (2014), " The Problem of Citation in the Digital Humanities". Proceedings of the Digital Humanities Congress 2012. Available online: http://www.hrionline.ac.uk/openbook/chapter/dhc2012-blaney (Retrieved 21 August 2014).

Burdick Anne, Drucker Johanna et al. (2012). Digital_Humanities (2012). MIT Press. Available online: https://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf (Retrieved 31 July 2014).

Campbell, Stephen W. (2014), "Improving Wikipedia: Notes from an Informed Skeptic". Perspectives on History, May 2014. American Historical Association. http://www.historians.org/publications-and-directories/perspectives-on-history/may-2014/improving-wikipedia (Retrieved 21 August 2014).

Cohen, Daniel & Rosenzweig, Roy (2005) Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web. Available online at: http://chnm.gmu.edu/digitalhistory/ (Retrieved 1 August 2014).

Cohen, Daniel (2010), "Is Google Good for History?" http://www.dancohen.org/2010/01/07/is-google-good-for-history/ (Retrieved 20 August 2014).

Cohen, Daniel (2011), " The Ivory Tower and the Open Web: Introduction: Burritos, Browsers, and Books [Draft]". http://www.dancohen.org/2011/07/26/the-ivory-tower-and-the-open-web-introduction-burritos-browsers-and-books-draft/ (Retrieved 21 August 2014).

Colvile, Robert (2007), "How to stave off a digital 'dark age'". In the Telegraph: http://www.telegraph.co.uk/technology/3353826/How-to-stave-off-a-digital-dark-age.html (Retrieved 22 August 2014).

A Companion to Digital Humanities (2004). Siemens, Ray, Schreibman, Susan, Unsworth, John (eds.) Oxford: Blackwell Publishing. Available online at: http://digitalhumanities.org/companion/ (Retrieved 20 August 2014).

Cronon, William (2012), "Scholarly Authority in a Wikified World". Perspectives on History, February 2012. American Historical Association. http://www.historians.org/publications-and-directories/perspectives-on-history/february-2012/scholarly-authority-in-a-wikified-world (Retrieved 21 August 2014).

Davidson, J. (2006). "Persistent Identifiers". DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. Handle: 1842/3368. Available online: http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation (Retrieved 31 July 2014).

Debates in the Digital Humanities (2012). Gold, Matthew K. (ed.) University of Minnesota Press. Available online at: http://dhdebates.gc.cuny.edu/ (Retrieved 20 August 2014).

Defining Digital Humanities. A Reader (2013). Terras, Melissa, Nyhan, Julianne & Vanhoutte, Edward (eds.) Farnham: Ashgate.

Digital Preservation Handbook (2008), Digital Preservation Coalition. Available online: http://www.dpconline.org/advice/preservationhandbook (Retrieved 21 August 2014).

Dunn, Stuart & Hedges, Mark (2012), "Crowd-Sourcing Scoping Study. Engaging the Crowd with Humanities Research." Centre for e-Research, Department of Digital Humanities, King's College London. Available online: http://crowds.cerch.kcl.ac.uk/wp-content/uploads/2012/12/Crowdsourcing-connected-communities.pdf (Retrieved 18 August 2014).

Eijnatten, Joris van, Pieters, Toine & Verheul, Jaap (2013), "Big Data for Global History: The Transformative Promise of Digital Humanities". BMGN - Low Countries Historical Review, Vol 128, No 4, 55–77.

Establishing Incentives and Changing Cultures to Support Data Access (2014). Expert Advisory Group on Data Access. Available online: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtp056495.pdf (Retrieved 22 August 2014).

Exceptions to Copyright: Research (2014). Intellectual Property Office. Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/315014/copyright-guidance-research.pdf (Retrieved 20 August 2014).

Flanders, Julia (2009), "The Productive Unease of 21st-century Digital Scholarship." In Digital Humanities Quarterly, Volume 3 Number 3. Available online: http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html (Retrieved 5 August 2014).

Fogg, Martha (2014), "Digitization: a successful collaborative partnership between library and publisher." Insights – 27(2), July 2014. Available online: http://uksg.metapress.com/content/u8337705854v2lq9/?genre=article&id=doi%3a10.1629%2f2048-7754.97 (Retrieved 7 August 2014).

Gold, Matthew (2012), "The Digital Humanities Moment". In: Debates in the Digital Humanities.

Halonen, Antti (2012), Being Open About Data. Analysis of the UK open data policies and applicability of open data. Finnish Institute in London. Available online: http://assets.finnishinstitute.studiocoucou.com/media/W1siZiIsIjIwMTMvMDEvMzEvMTNfMDNfMTNfNjM1X2JlaW5nXzIwb3Blbl8yMGFib3V0XzIwZGF0YS5wZGYiX

V0/being%2520open%2520about%2520data.pdf?sha=9b1040ca (Retrieved 20 August 2014).

Hupaniittu, Outi (2012), Tutkijoiden ääni ja sähköiset aineistot. Selvitys muistiorganisaatioiden asiakkaitten digitoitujen aineistojen tarpeista ja saatavuudesta. Svenska litteratursällskapet i Finland. Available online: http://www.sls.fi/media/pdf/forskarnas_rost_slutrapport_28112011_fi.pdf (Retrieved 21 August 2014).

Ihalainen, Pasi (2013), ”Vertaileva Euroopan historian tutkimus parlamenttihistorian näkökulmasta”. Tieteessä tapahtuu 4/2013, 30–35.

Jeurgens, Charles (2013), ”The Scent of the Digital Archive: Dilemmas with Archive Digitisation”. BMGN - Low Countries Historical Review, vol. 128, no. 4, 30–54.

Jäntti, Kristofer (2014), Knowledge Gap and the Information Environment. More informed citizens or a growing divide? Finnish Institute in London. Available online: http://assets.finnishinstitute.studiocoucou.com/media/W1siZiIsIjIwMTQvMDQvMTUvMTQvNDUvNTYvMjc1L0tub3dsZWRnZUdhcC5wZGYiXV0/KnowledgeGap.pdf?sha=9de7c4de (Retrieved 21 August 2014).

Kirschenbaum, Matthew (2010), "What Is Digital Humanities and What’s It Doing in English Departments?" ADE Bulletin, Number 150. Available online: http://mkirschenbaum.files.wordpress.com/2011/01/kirschenbaum_ade150.pdf (Retrieved 6 August 2014).

Kirschenbaum, Matthew (2014), "What Is “Digital Humanities,” and Why Are They Saying Such Terrible Things about It?" differences 25.1. Available online: http://mkirschenbaum.files.wordpress.com/2014/04/dhterriblethingskirschenbaum.pdf (Retrieved 6 August 2014).

Koehl, Sean (2013), "We Need to Act to Prevent a Digital ‘Dark Age’" In Wired: http://www.wired.com/2013/05/we-need-to-act-to-prevent-a-digital-dark-age/ (Retrieved 22 August 2014).

Kroes, Neelie (2014), "Embracing the open opportunity." Speech at Open Knowledge Festival 2014, 17 July 2014. http://europa.eu/rapid/press-release_SPEECH-14-556_en.htm (Retrieved 21 August 2014).

Lahtinen, Anu (2013), "PID - Minkä teen, sen ymmärrän". Presentation in the National Archives Service of Finland seminar 8 November 2013. Available online: https://www.youtube.com/watch?v=hOLbX3_CZWs (Retrieved 31 July 2014).

Maron, Nancy & Pickle, Sarah (2014), Sustaining the Digital Humanities. Host Institution Support beyond the Start-Up Phase. Ithaka S+R. Available online: http://www.sr.ithaka.org/sites/default/files/SR_Supporting_Digital_Humanities_20140618f.pdf (Retrieved 19 August 2014).

McCarty, William (2012). "Introduction". In Collaborative Research in the Digital Humanities: A volume in honour of Harold Short on the occasion of his 65th birthday, September 2010, ed. Marilyn Deegan and Willard McCarty. London: Ashgate. Available online:

http://www.mccarty.org.uk/essays/McCarty,%20Into%20to%20Collaborative%20research%2020111204.pdf (Retrieved 21 August 2014).

McCarty, William (2014), "Getting there from here. Remembering the future of digital humanities." Literary and Linguistic Computing Advance Access published May 19, 2014. Available online: http://llc.oxfordjournals.org/content/early/2014/05/18/llc.fqu022.full.pdf?keytype=ref&ijkey=w2NFEEIl7G6Qli6 (Retrieved 21 August 2014).

Murray-Rost, Peter (2014), "ContentMine and Wikidata". Presentation at the Wikimania 2014 conference 8 August 2014. Slides available: http://www.slideshare.net/petermurrayrust/contentmine-and-wikidata (Retrieved 21 August 2014).

Nyhan, Julianne & Duke-Williams, Oliver (2014), "Joint and multi-authored publication patterns in the Digital Humanities". In Literary and Linguistic Computing Advance Access, published June 28, 2014. Available online: http://llc.oxfordjournals.org/content/early/2014/06/28/llc.fqu018.full.pdf+html (Retrieved 1 August 2014).

Onion, Rebecca (2014), "Snapshots of History". Slate. http://www.slate.com/articles/life/history/2014/02/_historyinpics_historicalpics_history_pics_why_the_wildly_popular_twitter.html (Retrieved 21 August 2014).

The Open Book (2013). Braybrooke, Kaitlyn & Nissilä, Jussi (eds.). The Finnish Institute in London. Available online: http://issuu.com/finnish-institute/docs/theopenbook_issuu_final (Retrieved 20 August 2014).

Parland-von Essen, Jessica & Nyberg, Kenneth (2014), Historia i en digital värld. Version 1.0.1. Available online: http://digihist.se/ (retrieved 1 August 2014).http://digihist.se/

Pekel, Joris, Fallon, Julia & Kamenov, Lyubomir (2014), Public Sector Information in Cultural Heritage Institutions. European Public Sector Information Platform Topic Report No. 2014 / 06.

Pekel, Joris (2014a), "Maintaining a healthy and thriving Public Domain". Europeana professional blog. http://pro.europeana.eu/pro-blog/-/blogs/maintaining-a-healthy-and-thriving-public-domain (Retrieved 7 August 2014).

Pekel, Joris (2014b), "Democratising the Rijksmuseum. Why did the Rijksmuseum make available their highest quality material without restrictions, and what are the results?" Available online: http://pro.europeana.eu/documents/858566/858665/Democratising+the+Rijksmuseum?version=1.1 (Retrieved 13 August 2014).

Pennock, Maureen (2006), "Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages". The Digital Curation Centre. Available online: http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf (Retrieved 21 August 2014).

Pollock, Rufus (2011), "Open Data: A Means to an End, not an End in Itself". Open Knowledge Foundation Blog. http://blog.okfn.org/2011/09/15/open-data-a-means-to-an-end-not-an-end-in-itself/ (Retrieved 20 August 2014).

Prescott, Andrew (2014a), "Dennis the Paywall Menace Stalks the Archives". Available online: http://digitalriffs.blogspot.co.uk/2014/02/dennis-paywall-menace-stalks-archives.html (Retrieved 4 August 2014).

Prescott, Andrew (2014b), "Digital Humanities and the Quest for Academic Respectability". Available online: http://digitalriffs.blogspot.co.uk/2014/05/digital-humanities-and-quest-for.html (Retrieved 21 August 2014).

Prom, Christopher J. (2011), Preserving Email. Technology Watch Report 11-01. Digital Preservation Coalition. Available online: http://www.dpconline.org/component/docman/doc_download/739-dpctw11-01pdf (Retrieved 21 August 2014).

Reinventing Research? Information Practices in the Humanities (2011), A Research Information Network Report April 2011. Available online: http://www.rin.ac.uk/system/files/attachments/Humanities_Case_Studies_for_screen_2_0.pdf (Retrieved 19 August 2014).

Rutner, Jennifer & Schonfeld, Roger C. (2012), Supporting the Changing Research Practices of Historians. Ithaca S+R. Available online: http://www.sr.ithaka.org/sites/default/files/reports/supporting-the-changing-research-practices-of-historians.pdf (Retrieved 1 August 2014).

Salt, Limara (2013), "Filmmaker John Akomraf on working with archives". Ideas Tap. http://www.ideastap.com/IdeasMag/the-knowledge/filmmaker-john-akomfrah-archive-interview

Scheinfeldt, Tom (2012), ”Why Digital Humanities is ’Nice’” . http://www.foundhistory.org/2010/05/26/why-digital-humanities-is-%E2%80%9Cnice%E2%80%9D/ (Retrieved 21 August 2014).

Shepherd, Elizabeth (2014), "Past, Present and Future: the Archive, the Academy and the Community". Inaugural lecture at University College London, 20 May 2014.

Spiro, Lisa (2011), "Computing and Communicating Knowledge: Collaborative Approaches to Digital Humanities Projects." In: Collaborative Approaches to the Digital in English Studies. Computers and Composition Digital Press. Available online: http://ccdigitalpress.org/cad/Ch2_Spiro.pdf (Retrieved 21 August 2014).

Spiro, Lisa (2014), "Defining Digital Social Sciences". Blog post in dh+lib. Available online: http://acrl.ala.org/dh/2014/04/09/defining-digital-social-sciences/?utm_content=buffer360a4 (Retrieved 1 August 2014).

Terras, Melissa (2013). "I'm not going to edit your £10,000 pay-to-open-access-publish monograph series for you." http://melissaterras.blogspot.co.uk/2013/11/im-not-going-to-edit-your-10000-pay-to.html (Retrieved 21 August 2014).

Terras, Melissa (2014), "A Decade in Digital Humanities". Inaugural lecture at University College London, 27 May 2014. Text and video available online:

http://melissaterras.blogspot.co.uk/2014/05/inaugural-lecture-decade-in-digital.html (Retrieved 11 August 2014).

Toivonen, Tuuli & Lehtomäki, Joona (2013), "Tutkimusaineistojen avaaminen on taitolaji. Helsingin Sanomat 6 December 2013. Available online: http://www.hs.fi/mielipide/a1386227812491 (Retrieved 21 August 2014).

Vanhoutte, Edward (2013), "The Gates of Hell: History and Definition of Digital | Humanities | Computing". In Defining Digital Humanities. A Reader Terras, Melissa, Nyhan, Julianne & Vanhoutte, Edward (eds.) Farnham: Ashgate. 125–127.

Vuorinen, Pentti (2014), Läpidigitalisoitunut maailma. Virtuaalinen tulevaisuus keskuudessamme. TrendWikiä hyödyntävä raportti. TEM-analyyseja 58/2014. Ministry of Employment and the Economy (Finland). Available online: http://www.tem.fi/files/39700/Lapidigitalisoitunut_maailma.pdf (Retrieved 7 August 2014).

Wilhelmsson, Thomas & Keinonen, Juhani (2013), "Tiedon avoimuus mullistaa myös yliopistomaailman." Helsingin Sanomat 23 November 2013.

Williford, Christa & Henry, Charles (2012), One Culture. Computationally Intensive Research in the Humanities and Social Sciences. A Report on the Experiences of First Respondents to the Digging Into Data Challenge. Council on Library and Information Resources. Available online: http://www.clir.org/pubs/reports/pub151 (Retrieved 1 August 2014).

Woollard, Matthew (2012), "Persistent Identifiers for the UK: Social and Economic Data". Metadata and Persistent Identifiers for Social and Economic Data workshop, 7-8 May 2012, Berlin. Available online at: http://www.ratswd.de/ver/docs_PID_2012/Woollard_PID2012.pdf (Retrieved 31 July 2014).

Zaagsma, Gerben (2013), ”On Digital History”. BMGN - Low Countries Historical Review, vol. 128, no. 4, 3–29.

Websites

AddressingHistory. http://addressinghistory.edina.ac.uk/

Ancestry. www.ancestry.com

Archives and Copyright. http://www.create.ac.uk/archivesandcopyright/

Archives Direct. http://www.amdigital.co.uk/m-collections/collection/archives-direct/

AvoinGLAM. http://avoinglam.fi/

Avoin julkishallinnon data -mestarikurssi. http://fi.okfn.org/koulutus/avoin-julkishallinnon-data/

Avoin kulttuuridata -mestarikurssi. http://fi.okfn.org/koulutus/avoin_kulttuuridata/

Big UK Domain Data for the Arts and Humanities. http://buddah.projects.history.ac.uk/

Bomb Sight. http://bombsight.org/

British Library Wiki for curating Public Domain digital content. http://blpublicdomain.wikispaces.com/home

British Library on Flickr. https://www.flickr.com/photos/britishlibrary/

British Museum. http://www.britishmuseum.org/

The British Newspaper Archive. http://www.britishnewspaperarchive.co.uk/

CENDARI. http://www.cendari.eu/

centerNet. http://digitalhumanities.org/centernet/

ContentMine. http://contentmine.org/

Costing and Budgeting Digitization Projects. https://opinio.ucl.ac.uk/s?s=31273

Data.Gov.Uk. http://data.gov.uk/

DataShine. http://datashine.org.uk

David Normal's website. http://www.davidnormal.com/#!crossroads-of-curiosity/ccw3

Digging Into Parliamentary Data. http://dilipad.history.ac.uk/

DIGI - National Library's Digital Collections. http://digi.lib.helsinki.fi

Digital Agenda for Europe. A Europe 2020 Initiative. https://ec.europa.eu/digital-agenda/en/

Digital and Information Literacy Framework. Open University. http://www.open.ac.uk/libraryservices/subsites/dilframework/

Digital Preservation Coalition. http://www.dpconline.org/

Digital resources. Institute of Historical Research. http://www.history.ac.uk/digital/resources

Digitalkoot. www.digitalkoot.fi

Endangered Archives Programme. http://eap.bl.uk/

Europeana. http://www.europeana.eu/

Find My Past. http://www.findmypast.co.uk/

Finnish Social Science Data Archive. http://www.fsd.uta.fi/fi/

Finnish Web Archive.
http://www.nationallibrary.fi/services/digitallegaldepositmaterials/webarchive.html

First World War Centenary. http://www.1914.org/

Historypin. www.historypin.com

Histropedia. http://www.histropedia.com/

The Humanities Matter. By UCL Digital Humanities Centre and 4Humanities
http://4humanities.org/wp-content/uploads/2013/07/humanitiesmatter300.pdf

Inspiring Ireland. http://www.inspiring-ireland.ie/

Institute of Historical Research. http://www.history.ac.uk/

The International Directory of Digital Humanities Centers.
http://digitalhumanities.org/centernet/centers

Jisc. http://www.jisc.ac.uk

Kansan Muisti. http://www.kansanmuisti.fi

Kuka oikein olet? Yle. http://yle.fi/ohjelmat/1325513

Language Log. Google Books: A Metadata Train Wreck.
http://languagelog.ldc.upenn.edu/nll/?p=1701#comment-41758

Making Digital History. http://makingdigitalhistory.co.uk/

Mozilla Webmaker. https://webmaker.org/

The National Archives. http://www.nationalarchives.gov.uk

National Co-ordinating Centre for Public Engagement.
http://www.publicengagement.ac.uk/

The National Digital Library. http://www.kdk.fi/

Old Weather. http://www.oldweather.org

Open Access Working Group of Open Knowledge. http://access.okfn.org

Open Definition. http://opendefinition.org/

Open Data Index. https://index.okfn.org/country/

OpenGLAM. http://openglam.org/

OpenGLAM Benchmark Survey.
https://outreach.wikimedia.org/wiki/GLAM/OpenGLAM_Benchmark_Survey

Open Science and Research Initiative, http://openscience.fi/

Operation War Diary. http://www.operationwardiary.org/

The Programming Historian. http://programminghistorian.org/

Public Domain Review. http://publicdomainreview.org

School of Data. http://schoolofdata.org/

Sodan jäljet. http://www.sodanjaljet.fi/

Tate. http://www.tate.org.uk

TheyWorkForYou. http://www.theyworkforyou.com

TIDSR: Toolkit for the Impact of Digitised Scholarly Resources. http://microsites.oii.ox.ac.uk/tidsr/welcome

UK Data Archive, http://www.data-archive.ac.uk/

UK Web Archive. http://www.webarchive.org.uk/ukwa/

What's That Picture? http://www.whatsthatpicture.com/

Who Do You Think You Are? BBC. http://www.bbc.co.uk/programmes/b007t575

Wikidata. www.wikidata.org

Wikimania London 2014 videos. http://new.livestream.com/wikimania/

Wikimedia blog. https://blog.wikimedia.org/

Wikiproject Open Access. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Open_Access

William J. Turkel's website. http://williamjturkel.net/

Appendix A: List of Interviewed People

- **James Baker**, Digital Curator, British Library
- **Jonathan Blaney**, Project Editor, British History Online, Institute of Historical Research
- **Adrian Brown**, Head of Preservation & Access, Parliamentary Archives
- **Stuart Dunn**, Lecturer, Centre for e-Research, Department of Digital Humanities, King's College London
- **Andrew Flinn**, Reader in Archival Studies and Oral History, Director of the Archives and Records Management Programme, Department of Information Studies, University College London
- **Adam Green**, Editor, The Public Domain Review
- **Pasi Ihalainen**, Professor of General History, University of Jyväskylä
- **Raine Koskimaa**, Professor of Contemporary Culture, University of Jyväskylä
- **Paula Le Dieu**, Freelance Digital Director
- **Julianne Nyhan**, Lecturer in Digital Information Studies, Department of Information Studies, University College London
- **Jessica Parland-von Essen**, Archive Director, Brages Pressarkiv

Appendix B: List of Digital Humanities Networks, Centres and Study Programmes in the United Kingdom, Ireland and Finland

The list is not complete, it only acts as a further reference aid.

- International Networks and Associations
    - CenterNet, international network of digital humanities centershttp://digitalhumanities.org/centernet/
    - The Alliance of Digital Humanities Organizations (ADHO) http://adho.org/
    - The European Association for Digital Humanities (EADH) http://eadh.org/
- United Kingdom
    - York http://www.york.ac.uk/digital-heritage/
    - King's College London http://www.kcl.ac.uk/artshums/depts/ddh/index.aspx
    - University College London http://www.ucl.ac.uk/dh/
    - Nottingham http://www.nottingham.ac.uk/digital-humanities-centre/index.aspx
    - Open University http://www.open.ac.uk/Arts/digital-humanities/projects.shtml
    - Oxford http://digital.humanities.ox.ac.uk/Default.aspx
    - Cambridge http://www.digitalhumanities.cam.ac.uk/
    - Glasgow http://www.gla.ac.uk/subjects/informationstudies/
    - Edinburgh: http://www.digital.hss.ed.ac.uk/
    - Sheffield http://www.sheffield.ac.uk/hri
    - Queen's University Belfast http://www.qub.ac.uk/cdda/CDDA_2011/Welcome/
- Ireland
    - Trinity College Dublin https://www.tcd.ie/trinitylongroomhub/themes/digital-humanities/
    - An Foras Feasa http://www.forasfeasa.ie/
    - Digital Humanities Observatory http://dho.ie/ (closed 2013)
    - http://www.learndigitalhumanities.ie/

- Finland:
    - Åbo Akademi (minor subject): http://www.abo.fi/institution/en/News/Item/item/7499
    - Aalto University, Media Lab Helsinki https://medialab.aalto.fi/
    - Professorship in Digital Resources (jointly University of Helsinki and Centre for Preservation and Digitisation, National Library of Finland) http://www.helsinki.fi/hum/english/news/140210.htm
    - Digital Humanities Finland symposiums http://matsfridlund.com/digital-humanities-finland/