Digitisation and Discoverability of Historical Texts Online Using
Chaucer’s *Canterbury Tales*

A study submitted in partial fulfilment
of the requirements for the degree of
MA Librarianship

at

THE UNIVERSITY OF SHEFFIELD

by

MEAGHAN WRIGHT

September 2016
Abstract

Background

Libraries make their collections accessible by digitising them and making them discoverable online. There are specific challenges with historical collections undergoing these processes. Much of the literature is focused on individual or experimental attempts to address or survey these issues.

Aims

This dissertation aimed to gain a deeper understanding of the challenges surrounding the digitisation and discoverability of historical materials using a practical and holistic approach.

Methods

A case study was built around digitised Canterbury Tales manuscripts. Mixed methods research was conducted. Content analysis of eight digitised manuscripts investigated the features and tools which comprise the end-user interaction with the digital objects. Interviews were conducted with both librarians (eight participants) and researcher users (three participants) of digital historical collections.

Results

The digitisation of historical materials begins with a careful selection of objects. Digitisation by project is most common, although some mass digitisation occurs. The physical condition of the material is integral to the process, and special imaging equipment should be used. The online presentation of historical materials includes high resolution images to accurately reproduce the original object. Interface features are usually dependent on the library’s computer system. Metadata should be detailed and standardised, despite the expense to create and verify. Browse and search both need to be supported for discoverability. Centralisation and interoperability of searching and collections makes it easier to find objects. Users want transparent policies to understand how the digitisation and discovery processes impact resource accessibility. With diverse users and needs to consider, librarians must balance different notions of usefulness in many aspects of the digital library.
Conclusions

It is concluded that historical materials digitisation presents challenges, which libraries overcome by working collaboratively. The main factors which impact the digitisation and discoverability include projects, funding, time, material physical condition, external relations, technology, and user desire for clarity.
Acknowledgments

I would like to thank my supervisor, Paula Goodale, for her support and advice throughout this dissertation, and for making every meeting a positive experience. I would also like to extend my gratitude to the librarians and researchers who were kind enough to take the time to allow me to interview them, and who never failed to be helpful and bring new light to my research. Thank you to my family and friends for their support, especially Ken and my Mom for taking the time to read my draft and for their encouragement and feedback.
# Table of Contents

Abstract ................................................................................................................................. i  
Acknowledgments ................................................................................................................ iii  
List of Tables and Figures ..................................................................................................... vi  
1. **Introduction** .................................................................................................................. 1  
   1.1 Background and Context .............................................................................................. 1  
   1.2 Research Questions .................................................................................................... 2  
   1.3 Approach .................................................................................................................. 2  
2. **Literature Review** ......................................................................................................... 4  
   2.1 Imaging Process .......................................................................................................... 4  
   2.2 Presentation ................................................................................................................. 5  
   2.3 Digitisation User Needs ............................................................................................. 6  
   2.4 Metadata .................................................................................................................... 7  
   2.5 Information Retrieval ................................................................................................. 7  
   2.6 Discoverability User Needs ......................................................................................... 8  
4. **Methodology** .............................................................................................................. 11  
   3.1 Theories .................................................................................................................... 11  
   3.2 Case Study ................................................................................................................ 11  
   3.3 Mixed Methods .......................................................................................................... 12  
   3.4 Content Analysis ....................................................................................................... 13  
   3.5 Interviews ................................................................................................................ 14  
   3.6 Ethics .......................................................................................................................... 16  
5. **Results** ......................................................................................................................... 18  
   4.1 Content Analysis ....................................................................................................... 18  
      4.1.1 Imaging .............................................................................................................. 19  
      4.1.2 Presentation ..................................................................................................... 20  
      4.1.3 Metadata .......................................................................................................... 22  
      4.1.4 Information Retrieval ....................................................................................... 24
4.2 Interviews ......................................................................................................................... 26
  4.2.1 Librarians .................................................................................................................. 26
  4.2.2 Researchers .............................................................................................................. 33

6. Discussion .......................................................................................................................... 37

7. Conclusion .......................................................................................................................... 44

References ................................................................................................................................ 47
  Digital Objects Examined in the Content Analysis ............................................................... 49

Appendices ............................................................................................................................. 51
  Appendix A ............................................................................................................................ 51
  Appendix B ............................................................................................................................ 55
    Appendix B.1 ....................................................................................................................... 55
    Appendix B.2 ....................................................................................................................... 56
  Appendix C ............................................................................................................................ 58
  Appendix D ............................................................................................................................ 59
  Appendix E ............................................................................................................................ 67
  Appendix F ............................................................................................................................ 70

Access to Dissertation .......................................................................................................... 82

Address & First Employment Destination Details ................................................................. 84
List of Tables and Figures

Table 1........................................................................................................12
Table 2........................................................................................................14
Table 3........................................................................................................15
Table 4........................................................................................................18
Table 5........................................................................................................19
Table 6........................................................................................................20
Table 7........................................................................................................22
Table 8........................................................................................................24

Figure 1.....................................................................................................21
Figure 2.....................................................................................................22
Figure 3.....................................................................................................26
1. **Introduction**

1.1 Background and Context

Digitisation efforts in libraries, archives, and other cultural heritage institutions have expanded immensely in the past twenty years. In addition to technological developments, the issue has been propelled by heightened concerns about preservation challenges, access provision, and value demonstration in the Library and Information Services sector. Oftentimes when choosing which resources to prioritise in digitisation, organisations focus on the most used or unique items within their collections. The latter category may include historical items. Preservation considerations, such as reducing the handling of fragile materials, also make historical items favourable candidates for digitisation. Many institutions digitise their historical material for such reasons. There are, however, significant challenges associated with this task.

Digital collections are put online to be made accessible, meaning able to be used, which is achieved by making them available and findable. Digitisation makes collections available, which for historical materials is not an unvarying, simple process. Many materials are too fragile for standard scanning techniques, high-resolution images are required for preservation and to detect material detail, and Optical Character Recognition (OCR) technologies are often ineffective tools when faced with handwritten scripts. Once an object is digitised, it must be made discoverable in order to be truly accessible. Again, this proves difficult with historical materials. Detailed item-level metadata is difficult and expensive to create, matching Information Retrieval (IR) techniques against non-standardised spelling are unproductive, and overall collection organisation can be confusing.

Despite these obstacles, cultural heritage institutions digitise these historical materials. 84% of cultural heritage institutions have digital collections, but only 41% of these institutions have a digitisation policy and on average 23% of heritage collections have been digitised (Nauta & van den Heuvel, 2015, p. 4). In terms of access, libraries report that 68% of their descriptive metadata is online for general use, but only 48% of their digital collections are likewise available, and just 55% of those available collections are on the institution’s website, the most popular access channel (Nauta and van den Heuvel, 2015, p. 5). While cultural heritage institutions keep open communications, some scholars (Berger, 2014, p. 393) have noted that practices are still a chimera, with too many institutions working on problems independently. Literature on the topic often focuses on the efforts of a single project or technology for digital historical items, such as a project digitising a collection of a particular author’s manuscripts (Nicolas, Paquet, & Heutte, 2003) or an...
analysis of software solutions allowing online publication of finding aids (Cornish & Merrill, 2010). It is important for institutions to share their current practices, as well as to aggregate these procedures and assess solutions comparatively. User needs is essential to designing useful digital libraries, as well as for evaluating and justifying them. This includes who will use the digital collections, how they will use them, and what features and tools they may need. It would be beneficial to understand what digital librarians are trying to achieve, and how closely that connects with what users find useful in digital collections.

1.2 Research Questions

The aim of this dissertation is to gain a deeper understanding of the challenges surrounding the digitisation and discoverability of historical materials in order to achieve greater insight into best practice and common goals. It will examine how institutions digitise historical materials, the perceptions of librarians on the processes in practice, and the views of researchers of how effective the end result is. The objectives of the dissertation are to investigate:

1) What processes historical materials undergo in order to be digitised in practice.

2) How the digitised object is presented online and the interface features available.

3) What researchers using these digital materials think of current digitisation practices.

4) What processes are used to make digitised historical materials discoverable in practice.

5) Whether researchers find the discovery approaches useful and valuable.

This dissertation does not aim to criticise libraries or advocate a one-size-fits-all approach, but to evaluate the practices in use by applying a holistic approach to identify recurring themes and developing trends.

1.3 Approach

To investigate these questions, this dissertation will use a case study to concentrate the research and designate historical collections for examination. The focus of medieval manuscripts was selected. This type of historical material is handwritten and old enough to offer digitisation difficulties, without being so ancient as to require the complication of extensive restoration work. It has a more complicated history and structure than that of a book: “a MS [manuscript] codex is a complex artefact, which may be made up of text copied by several different scribes, programmes of
decoration carried out by multiple artists, or distinct parts bound together many years after they were produced” (“Material philology and MS records,” 2014).

As mentioned above, there are many previous studies focused on the digitisation and findability of a specific collection or system. The scope of this dissertation is to examine multiple institutions, which consequently includes different systems and software, to gather a more complete view of current practice. To introduce a controlled variable, this project will survey the digitisation of similar manuscripts from the same time period and place across these institutions. Since multiple institutions do not digitise the same material repeatedly, these digital objects could not be the same manuscript, but similar versions of the same text were selected. The case study of this research is digitised versions of *The Canterbury Tales* medieval manuscripts. There are dozens of these manuscripts, which were handwritten in England in Middle English during the fifteenth century, are held at some of the most notable institutions in the world, and are of high literary significance. These factors increase their likelihood of being digitised and provide an advantageous sample group. The data was collected from analysis of the digitised manuscripts themselves, interviews with digital librarians connected to the manuscripts and their digital collections, and interviews with researcher users of digital historical collections.

The dissertation focuses on two main themes addressed in the aim: digitisation and discoverability. These were broken down into six themes: (under digitisation) the imaging process, online presentation, digitisation user needs, (under discoverability) metadata, IR, and discovery user needs. The Literature Review and Results chapters are structured around these themes, as well as the data collection.

The next chapter will review the relevant literature. The following chapter will provide a more detailed methodology of this research project. The main findings of the three sources of data collection will be then presented. The results will be brought together and analysed in the Discussion chapter, answering the five objectives. A Conclusions chapter will summarise the main points of significance, the limits of the study, and areas for further research.
2. Literature Review

This chapter will provide an overview of the topic of digitisation and discoverability relating to digitised historical materials. The structure of this literature review will follow the six themes of the overall dissertation. First it will examine digitisation of historical materials in terms of the imaging process, the online presentation, and the digitisation user needs. Next, it will investigate the findability of digitised historical materials in terms of metadata, Information Retrieval (IR), and discovery user needs. The subjects discussed are interconnected, and some points are related to multiple themes. These issues will inform and direct the dissertation.

2.1 Imaging Process

The process of imaging historical materials is an area of concern for many libraries. Key sources for understanding imaging practice today are digitisation guides published by leading library organisations. The International Federation of Library Associations and Institutions (IFLA) has useful guidelines for digitising rare books and manuscript collections, acknowledging that special considerations and procedures for these materials are necessary and are separate from typical digitisation. The first step is selecting appropriate materials based on the project goals, intended use, scope, resources, and copyright, and includes a physical assessment of the material (IFLA Rare Books and Special Collections Section, 2014, pp. 8-9). Digitisation of Special Collections is a complex process because it is important to maintain relations within parts of a single object and among parts of a collection as a whole, and to provide contextual and authenticity information, requiring special metadata and platforms (IFLA RBSCS, 2014, p. 5). The maintenance of work and collection relationships and of context is a more common issue in archives than libraries, potentially requiring specialised expertise to achieve. The digitisation process of historical materials requires high resolution image quality to recreate the look of the original work as much as possible, which is significant for some research (IFLA RBSCS, 2014). Unlike modern or electronic materials, historical resources require more than merely legible digital versions, and tend to be difficult to read without clear and close-up images. The fragility of the materials also complicates digitisation, and the right equipment must be chosen to avoid damaging the resources (IFLA RBSCS, 2014, p. 10). These concerns add to the difficulty of digitising historical materials.

In his overview of Special Collections departments, Berger (2014) notes that digitised versions of analogue materials may never be enough; many people simply prefer to handle physical objects and there are some needs that digital objects cannot meet, such as bookbinding research (p. 419). This links to Rekrut’s (2014) point connecting the materiality of historical records with the
circumstances of its creation, which may be lost or disregarded once online (p. 238). These points highlight the importance of preserving the physicality of historical texts as much as possible in digital form. Digitisation can complement the physical preservation of the analogue item, as long as the digital librarian remembers and accounts for diverse needs and technological obsolescence. Such intentions need to be planned for in advance because it will affect aspects such as the format and metadata.

2.2 Presentation

Another key factor within digitisation is the presentation used. Le Bourgeois and Emptoz (2007) recount the multi-disciplinary project DEBORA, which aimed to digitise and make accessible rare sixteenth-century printed books. Most digitised historical materials are only available in image mode because they cannot be automatically processed or users prefer the original typography (p. 193). This project sought to address technological issues including limited access due to image file size, lack of standard file format for progressive transmission, and limited querying possibilities (Le Bourgeois & Emptoz, 2007, p. 194). DEBORA developed a prototype which digitised and retrospectively converted the historical documents to extract the detailed description of the image contents as suited to users’ needs (Le Bourgeois & Emptoz 2007, p. 193). This dissertation investigates whether such problems and solutions are still relevant.

Another related project to this dissertation is e-codices, which intends to digitise and virtually reunify medieval and early modern Swiss manuscripts. Cusimano (2014) attributes part of the effectiveness of e-codices to its graphic user interface which is easy to use, provides users with updates and information on the objectives and progress of digitisation projects, and provides access to digitised manuscripts (p. 125). Other success factors included high-quality digital copies, easy access through a functional online visualisation tool, and the portal’s metadata population, which ensures accessibility, traceability, and interoperability of data (Cusimano, 2014, p. 125). Another e-codices author emphasises the re-shaping of digital images to produce a true digital representation of the physical object, i.e. a medieval manuscript, by creating multiple separate files, then incorporating metadata for each image file and for the whole manuscript in the presentation database (Austenfeld, 2010, p. 149). The way a work is digitally re-structured is especially significant with manuscripts where pages may contain different stories, types of text, or marginalia, or be written by several scribes in different times or places. These materials may not follow modern notions of structure in their physical form.
2.3 Digitisation User Needs

User needs is an essential concern when planning or implementing digitisation of library collections. It is necessary to know why the process is being undertaken, for whom, and how the final outputs will be used. The intended purpose and potential user base need to be considered in order to determine the imaging process and presentation. These are advised by several important funding bodies such as IFLA and the Wellcome Trust (IFLA RBSCS, 2014, p. 7; Wellcome Trust, n.d., p. 2). The importance of user needs is more well-known in the area of IR, but it is also inseparable from digitisation. User needs considerations in the digitisation process are observed when Austenfeld (2010) notes that the application housing digital objects must be constructed to reflect the requirements of humanities scholar users (p. 146). Imaging equipment and practices should be chosen based on materials and user groups. For example, the e-codices project used professional photographers and the most up-to-date imaging equipment because the delicate material could not be opened flat and images needed to offer “palaeographic” quality for scholarly use (Austenfeld, 2010, p. 149). Procedural guidelines were followed to ensure that each digital manuscript was identically sized and oriented with the same colour index and a black background for uniformity, to produce the most aesthetically pleasing and functional online research environment for the benefit of users (Austenfeld, 2010, p. 149). These are obstacles which most historical digitisation projects share.

One study in New Zealand investigated the information needs of digital cultural heritage users to aid the development and presentation of such resources. Noting that user needs studies tend to focus on the evaluation of particular collections or projects, Dorner, Liew, and Yeo (2007) attempt to avoid this by examining many different cultural heritage organisations (p. 167), similar to this dissertation. They found two main reasons for digitisation activity development: top-down, driven by records management perspective and organisation-wide strategies, mostly in local or regional authorities; or bottom-up, driven by the need to increase access to cultural and heritage resources in response to user demand while ensuring materials are preserved, mostly in libraries and museums (Dorner et al., 2007, p. 169). It is expected that the examined libraries will accordingly have an access-driven approach, but there may be questions of whether digitisation is driven by policy or user demand. Users were found to be from humanities or history-related topics, and used resources mainly for scholarly or teaching purposes in various formats, with different people preferring specific resources: for example, manuscripts were preferred by historical researchers and lexicographers (Dorner et al., 2007, p. 171). Digital cultural heritage materials need to be presented with authenticity, contextualisation, and integration of resources, particularly
through a central portal and integrated with related information (Dorner et al, 2007, pp. 180-2). These are important issues specific to cultural heritage resources, which should be taken into account in order to properly address the needs of intended users.

2.4 Metadata

The creation and selection of metadata to describe digital objects is essential to discoverability. Foulonneau and Riley (2008) describe metadata implementation, design, and interoperability within the overall mission of cultural heritage institutions. Functions performed on these digital resources through metadata include discovery, interpretation, preservation, management, representation, and re-use; however, it is a standardisation challenge that many different people create metadata and do so subjectively (Foulonneau & Riley, 2008, pp. 5-6). Cultural heritage institutions increasingly store both digital objects and metadata records in the same application (Foulonneau & Riley, 2008, pp. 7-8). Metadata and their digital objects are more linked than ever. When designing digital collections, it is important to define current and potential users, and to decide what needs will be actively supported (Foulonneau & Riley, 2008, p. 76). The issues of metadata functions, its connection with the digital object itself, and the consideration of user needs are explored within this study.

Skinner (2014) defines the primary role of metadata standards in order to increase the interoperability of institutionally created metadata. In her literature review focusing on museums and archives, she notes that such institutions historically have not shared finding aids cross-institutionally and the material described is unique (Skinner, 2014, p. 53). These concerns are a driving force behind the need for trustworthiness of metadata in these domains, achieved by preserving form, content, and context (p. 57). Disparate metadata makes it more difficult for institutions to share and users to understand information about digital objects.

2.5 Information Retrieval

The other discoverability issue explored is IR processes. Many of the problems relating to retrieval of historical manuscripts relate to lack of matching. Hauser, Heller, Leiss, Schulz, and Wanzeck (2007) describe the challenges of accessing historical documents in digital form due to non-normalised language and spelling, curtailing the ability to use standard indexing techniques directly (p. 1). They describe a project designing an approximate matching model to historical words (Hauser et al., 2007, p. 7), but this required transcribed or OCR texts to analyse.

More recently, Naji and Savoy (2011) describe and evaluate IR models and search strategies for digitised medieval manuscripts. They note that “performing effective retrieval on
historical manuscripts is still an unsolved issue despite the increasing need of museums, libraries, and even the general public for easy access to historical manuscripts” (Naji & Savoy, 2011, p. 104). This article assessed the retrieval effectiveness of four automatic transcripts, each introducing a form of spelling correction (Naji & Savoy, 2011, p. 103). The best performance was achieved through retrieving possible words with a log-likelihood close to the best alternative, and recommended applying either aggressive (longer queries) or no (short queries) stemming normalisation (Naji & Savoy, 2011, p. 113). Methods such as these demonstrate the development of automatic IR, and approaches to addressing inherent issues within historical texts.

Deal (2014) experiments with data visualisations of digital historical collections, which facilitates browsing IR. She remarks that the next step is to leverage the array of digital content in new experimental ways, but that current text-based digital collections are limited in browsing capabilities (Deal, 2014, p. 15). Visualisations can show basic features of collections through charts and graphs, let users quickly grasp the nature of collections and browse contents, and give a sense of the collection’s scope and context (Deal, 2014, p. 15). Such approaches enable browsing and also address user needs in the context of digital cultural heritage collections by making IR easier.

2.6 Discoverability User Needs

User needs is an essential consideration in the discovery processes. The work of Marchionini on exploratory search is particularly relevant to this project. He defines three types of search activities: lookup, learn, and investigate, the latter two being types of exploratory search (Marchionini, 2006, p. 42). Exploratory search blends browsing and querying strategies, requiring strong human participation in more constant and exploratory processes, and is better supported by interactive IR, annotation, and results manipulation tools than analytical strategies around a specific query, which many search engines favour (Marchionini, 2006, pp. 42-3).

These observations can be matched to similar concepts in other studies investigating cultural heritage users’ information needs. Examining cultural heritage user needs from nine institutions throughout the Netherlands, Amin, van Ossenbruggen, Hardman, and van Nispen (2008) found that the daily search tasks of experts were dominated by a range of different and complex information gathering tasks including topic search, combination, and exploration, although search tools tend to support simpler, fact-finding tasks (p. 39). These information gathering tasks directly correlate to Marchionini’s information model and validate the particular importance of exploratory search in cultural heritage research. As Amin et al. (2008) point out, the resources used in cultural heritage research are rich; heterogeneous; combine highly structured,
semi-structured, and unstructured information; include authorised and unauthorised sources; and
text and other media (p. 39). Given such disparate and unstable information, a browsing and fluid search strategy may be preferred. It was also found that cultural heritage experts value combining searches, direct communication for information transfer, and authenticable trust in sources (Amin et al., 2008, p. 39). These studies, their findings, and their recommendations were influential in this dissertation’s areas of investigation.

Before undertaking a mass digitisation project, the Southern Historical Collection (SHC) (2009) conducted focus groups and interviews with researchers in the humanities to ascertain search needs and preferences. More than one group indicated that researchers favoured finding aids that mirror and cross-reference their physical counterparts, and that familiar, easy-to-use search was a priority (SHC, 2009, pp. 11, 14). This may point to a lack of differentiation between original materials and their digital surrogates. This dissertation will gauge whether this sentiment is still prevalent among researchers, and how mass digitisation connects to the traditional project-based approach to digitising.

In a similar investigation of Special Collection user needs, Gueguen (2010) found a concern that novices may be confused if they expect item-level description and are faced with collection-level, but that researchers familiar with aggregate-level descriptions may be unable to use item-level descriptions efficiently (Gueguen, 2010, p. 97). Considering the research methods of humanities scholars, she notes a tendency to browse large and diverse sets of resources, retrieve a large set of results to sort through, use serendipitous retrieval to meet specific and unarticulated needs, and rely heavily on texts (Gueguen, 2010, p. 97). Research such as this is highly valuable to designing digitised collections that will be functional to users in practice.

This dissertation addresses several gaps in this field as observed in the literature. Many articles describe experimental methods for improving digitisation or discoverability of historical materials. While this research subject is necessary and promotes innovation in the field, such focus tends to be narrow and eclectic, reports on one-off projects, and gives little indication of common practices or themes in actual libraries. Similarly, many articles concentrate on a single digitisation project or technology. There are many different types of collections and technology in practice, and some comparison across these factors could be useful. These are the gaps which the current study will address by examining the actual digitisation and discoverability practices across several institutions. The concentration on Canterbury Tales manuscripts focuses the study across these variables.
The topic of digital historical materials is reviewed in relation to digitisation, discoverability, and user needs. This dissertation will investigate the extent of the application of these solutions in the wider context of multiple institutions.
4. **Methodology**

To investigate these questions, this research focused on resources which bring difficulties specific to historical materials. The intention of this project is to holistically examine digitisation and discoverability of historical materials through practical experiences of librarians and researchers.

3.1 **Theories**

This dissertation takes an interpretivist view of knowledge, that research need not follow the principles of natural sciences and that knowledge does not have to be confirmed by the senses (Bryman, 2012, pp. 27-8). This project accepts the opinions of the examined processes as evidence. Accordingly, this dissertation takes a constructivist perspective of social entities, that they are continually produced by social actors within them (Bryman, 2012, p. 33). People have divergent but equally valid perspectives regarding the subject of this study. This project is focused on professional current practices, therefore these overarching viewpoints are underlying foundations of the research. It is an inductive study, formulating theories based on data collected.

3.2 **Case Study**

This project centres on a small base of data examined from different angles, which supports a case study. Case studies are appropriate for ‘how’ and ‘why’ questions without behavioural control, focusing on contemporary events (Yin, 2014, p. 9). Case study research can manage a variety of evidence (Yin, 2014, p. 12). These features align with the intended investigation. This dissertation is a single, embedded case study. The case discussed is digitised manuscripts of The Canterbury Tales and the units of analysis are the content analysis and the interviews.

The case chosen was of materials old enough to be fragile and contain unfamiliar script and spelling, so the digitisation and discoverability processing would be difficult and require special procedures, which is what this dissertation investigates. The selected case is fifteenth-century Canterbury Tales manuscripts because there are enough of them in different institutions to offer a semi-controlled variable. Although not the same object, the similar structure, context, and content offers some standardisation. Additionally, these manuscripts hold certain value in any library collection as a famous literary work, ensuring digitisation. Digitised Canterbury Tales have been identified in at least seven different institutions (see Table 1 for list of institutions investigated, p. 12), some with multiple manuscripts online.
This selection of institutions, as well as being in-line with the present case study, also represents some of the most renowned and cutting-edge libraries in the world. Their activities in digital collections indicate the best practices and future of the sector, making the information provided by them more generalisable.

3.3 Mixed Methods

This dissertation utilises a mixed methods approach. It is a fixed design, in that the methods have been planned from the start, and follows a convergent parallel design, where the strands of data are analysed individually and then merged for a more comprehensive understanding of the phenomenon (Creswell & Plano Clark, 2011, pp. 54, 69). Priority is given to the qualitative data as it is more detailed, but quantitative data was collected first to provide considerations during the interviews. The point of interface was at interpretation, in the Discussion chapter.

The content analysis data establishes how digitised historical material is presented and made discoverable to the users. Such examination of the end-user interface is valuable to understand the practices put into use by libraries, but it can only reveal information on the top of a very complicated process where much of the activity is conducted below the surface. Moreover, a piecemeal analysis of features represented through one digital object and collection per library does not well represent the holistic and iterative experience of using such resources in research.

To address these gaps, interviews were conducted with librarians and researchers of digital historical collections. Interviews can address specific points of interest through questions, and provide the opportunity to gain from the cumulative knowledge of experts. This form of data collection has its limits: people may not remember all pertinent information in the moment, may venture off the subject, or may not describe the state of digital libraries as they are, but as they want

<table>
<thead>
<tr>
<th>Institutions Holding Digitised Canterbury Tales (Focus of Content Analysis &amp; Expert Interviews)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bodleian Libraries, University of Oxford</td>
</tr>
<tr>
<td>British Library</td>
</tr>
<tr>
<td>E-codices, Virtual Manuscript Library of Switzerland</td>
</tr>
<tr>
<td>Harry Ransom Center, University of Texas at Austin</td>
</tr>
<tr>
<td>Huntington Library, Art Collections, and Botanical Gardens</td>
</tr>
<tr>
<td>John Rylands Library, University of Manchester</td>
</tr>
<tr>
<td>National Library of Wales</td>
</tr>
</tbody>
</table>
them to be. These gaps of keeping the research current and on topic are addressed in the content analysis.

Both methodologies employed were structured around the themes of the dissertation, which can be mapped to the objectives: imaging process, online presentation, digitisation user needs, metadata, IR, and discovery user needs. These themes each address different but important areas within the main themes of digitisation and discoverability, providing a comprehensive perspective of the issues at large. These themes were based on the research question, literature review, and observations from exploring digital historical collections. After independent analysis was conducted on each of the components of data, results were compared and contrasted across key themes. Through the combined case study and cross-institution approach, the study took both a specialised and comprehensive viewpoint.

3.4 Content Analysis

An interface content analysis was conducted on both the digital objects of the digitised Canterbury Tales and the available discoverability features with which to access them (see Table 4 for list of manuscripts and collections, p. 18). To narrow the scope and keep data comparable, only the digital collection hosting the digitised Canterbury Tales was analysed. Content analysis is understood to be the systematic and objective analysis of message characteristics (Neuendorf, 2002, p. 1), where here the message is the digitised manuscript itself and its communicated findability. It incorporates typical or expected tasks and supporting features, and analyses how each digitised object or digital platform compares to the others. It is a transparent type of research which allows extensive analysis relatively easily and without obstruction (Bryman, 2012, p. 304).

Each digital object and library was analysed using a list of twenty-five relevant measurements. Each dimension had a number of potential criteria to grade the measurements (see Appendix A for complete content analysis coding manual). There were some notes taken to clarify or elaborate measurements, for example noting what features had been found under ‘combination’ criteria.

Once the data was collected it was analysed by comparing the results of every library for each criterion to identify common trends and to assess in which areas libraries tended to do the same things and where there was disagreement.
3.5 Interviews

The interviews addressed the objectives of this dissertation by investigating how librarians manage the challenges of digitising historical manuscripts and how well researchers believe they have achieved their goals. Each library under investigation (see Table 1, p. 12) was represented in the participants. Interviewees were usually people managing the digital collections, although one interview included a Photographer and Cataloguer, lending new perspectives to the data (see Table 2 for list of librarian job titles and codes).

<table>
<thead>
<tr>
<th>Job Titles of Librarians Interviewed</th>
<th>L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cataloguer</td>
<td>L2</td>
</tr>
<tr>
<td>Digital Access Section Manager</td>
<td>L3</td>
</tr>
<tr>
<td>Digital Curator, Digitisation</td>
<td>L4</td>
</tr>
<tr>
<td>Head of Digital Collections and Preservation Services</td>
<td>L5</td>
</tr>
<tr>
<td>Head of Digital Collections Services</td>
<td>L6</td>
</tr>
<tr>
<td>Heritage Imaging Manager</td>
<td>L7</td>
</tr>
<tr>
<td>Kemble Digital Projects Librarian</td>
<td>L8</td>
</tr>
<tr>
<td>Photographer</td>
<td>L9</td>
</tr>
<tr>
<td>Project Director</td>
<td>L10</td>
</tr>
<tr>
<td>Project Manager</td>
<td></td>
</tr>
</tbody>
</table>

Identification of librarians was made through the staff pages or library enquires service of the relevant institutions. All but two libraries were represented by one interviewee each. The exceptions were e-codices, with two interviewees, and the Rylands Library, with three interviewees.

The researchers interviewed were academic users of the historical materials in online digital collections. As Gueguen (2010) notes: “one of the strongest indicators of whom the collection may be useful for are the traditional users of the physical resources that are being digitised- researchers in the humanities” (p. 97). Participants were identified through the University of Sheffield's Humanities Research Institute, which specializes in cross-disciplinary digital humanities.
Researchers in the University of Sheffield History and English Departments were also contacted for potential participation, by which an additional interviewee was gained. Although limited to three researcher interviews, the breadth of discipline represented gives value to this data (see Table 3 for list of researchers’ job titles).

<table>
<thead>
<tr>
<th>Job Titles of Researchers Interviewed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Director of HRI Digital</td>
<td>R1</td>
</tr>
<tr>
<td>Research Associate</td>
<td>R2</td>
</tr>
<tr>
<td>English Lecturer</td>
<td>R3</td>
</tr>
</tbody>
</table>

*Table 3*

Participant perspectives included:

1. Digital humanities projects, particularly research which was impossible before the creation of digital collections and tools, such as data mining.

2. Traditional methods historical research based on digital collections of crime history.

3. Traditional methods English literature research often involving digital collections of medieval material.

The interviews were semi-structured and included a list of prompts and topics to cover in the form of an interview guide. All questions were similarly worded, while allowing for leniency in how the interviewee responded, and the questions could deviate from the schedule (Bryman, 2012, p. 471). The interview guide was structured around the six themes of this dissertation, with one question for each theme and several underlying prompts to explore further if necessary. There were separate questions for librarians and researchers. There was also an introductory question on their organisation and role within it (see Appendix B for full interview guide). The questions and prompts arose from the literature review, content analysis, and research objectives.

A small pilot was conducted with a non-librarian, but the pilot participant had over a decade of working experience and gave helpful advice on the wording and presentation of the questions. Additionally, review and revision of questions were undertaken as necessary during and after each interview according to circumstance. For example, the question to librarians regarding information retrieval of digital collections (see Appendix B.1) was often substituted with the prompt about linear versus interactive search systems, as this seemed to be more in line with how the librarians understood and viewed the discovery process.
The interviews were transcribed and manually coded. When analysing interviews, the emphasis is on what the interviewee views as important in explaining and understanding events and patterns (Bryman, 2012, p. 472). Responses were taken as stated and were not edited or fact-checked as they represent each participant’s understanding of circumstances. During interview coding, one or more keywords are linked to a text segment to identify a statement, which is then more systematically conceptualised into a categorisation (Kvale, 2007, p. 105). These keywords or labels are applied to areas from the transcripts which are of potential theoretical significance and/or particularly striking within the social world under examination (Bryman, 2012, p. 568). This dissertation followed open coding, in order to break down, examine, and conceptualise the data for concepts to be developed into categories (Bryman, 2012, pp. 569-70).

Librarian and researcher interviews were analysed separately, as their separate questions addressed areas from different perspectives. One slight exception was R1, who technically created collections rather than used them. However, the collections and their understanding of the issues were researcher-based rather than librarian, so R1 was asked a mix of librarian and researcher questions. Concepts were determined based on significance, logic, and repetition, and these concepts were compared and categorised. These categories were composed of sub-themes and related topics, organised into tables according to the dissertation’s six-theme structure. The most significant of the categories were then explained in the Discussion chapter.

3.6 Ethics

This research has been approved as ‘low risk’ (see Appendix C for ethics approval and Appendix D for ethics application) because it involves human participants. To preserve context, interviewees were anonymised to job title and institution during transcription. This was stated on the informed consent form (see Appendix E) and participants could choose further anonymisation through a generic job role, and request not to have their data linked to their institution. Participants could read their interview transcript and redact information. Several participants did request a review, and no information was objected to by the participants. The research does not involve personal information.

The interviews were conducted in person or via Skype or phone, and were recorded. Data was stored on a password-protected personal computer. Backups were kept on a personal hard drive, and on the Information School’s research data drive which can be accessed by only by the researcher, the dissertation supervisor, the School’s Examinations Officer, and ICT staff operating the facility. The data will be deleted three months after the dissertation has been completed. The
analysis was conducted by the researcher, who controlled the generated data. This research was not externally funded.
5. Results

The results of this research are presented below. First the quantitative content analysis findings will be presented and then the qualitative interview findings. For clarity and coherence, the results have been organised into the themes of this dissertation structure (see Methodology, p. 13). Several topics may pertain to more than one theme, but these themes are broadly reflected in the findings and the structure gained by such organisation is useful for a more systematic and meaningful understanding of the results. User needs topics cover intended users, intended uses, and tools to aid these users and uses.

4.1 Content Analysis

The content analysis was conducted on the digitised Canterbury Tales in eight digital collections, from the seven institutions (see Table 1, p. 12), one of which (University of Oxford) had manuscripts in two separate digital collections. The analysis was of the specific collection database in which the Canterbury Tales was held (see Table 4 for list of manuscripts and collections examined). Some libraries had multiple collections and/or multiple databases, but these and the library’s catalogue were outside the scope of this project. Analysis was conducted 23-25 June, 2016, and updates or changes may have occurred since then.

<table>
<thead>
<tr>
<th>Digitised Manuscript Examined</th>
<th>Digital Collections Examined</th>
<th>Library of Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus Christi College MS 198</td>
<td>Early Manuscripts at Oxford University</td>
<td>University of Oxford</td>
</tr>
<tr>
<td>Harley MS 7334</td>
<td>Digital Collections: Digitised Manuscripts</td>
<td>British Library</td>
</tr>
<tr>
<td>Cod. Bodmer 48</td>
<td>Foundation Martin Bodmer</td>
<td>E-codices, Virtual Manuscript Library of Switzerland</td>
</tr>
<tr>
<td>HRC 46</td>
<td>Digital Collections: Medieval and Early Modern Manuscripts Collections</td>
<td>Harry Ransom Center, University of Texas at Austin</td>
</tr>
<tr>
<td>MssEL 26 C 9</td>
<td>Digital Library: Manuscripts</td>
<td>Huntington Library, Art Collections, and Botanical Gardens</td>
</tr>
<tr>
<td>English MS 113</td>
<td>Rylands Medieval Collection</td>
<td>John Rylands Library, University of Manchester</td>
</tr>
<tr>
<td>Peniarth MS 392D</td>
<td>Digital Gallery: Manuscripts</td>
<td>National Library of Wales</td>
</tr>
</tbody>
</table>

Table 4
The end-user interfaces which were analysed only displayed public data, so the content analysis portion focuses mainly on the user-facing themes of presentation and Information Retrieval (IR) of digital objects. Results are listed in descending order so it is visually clear which criteria were most or least common (N is number of interfaces found with each criteria, out of eight). The tools and features used to view and manipulate the images were categorised under presentation, and those used to find an image were categorised under IR. While user needs themes are reflected in the user interfaces, such as what features libraries thought users would need, they were not explicitly addressed in the content analysis data.

4.1.1 Imaging

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Criteria</th>
<th>N=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of work digitised</td>
<td>Full</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Partial</td>
<td>1</td>
</tr>
<tr>
<td>Information of digitisation process</td>
<td>None</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Imaging service name</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Description of process</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5

Table 5 shows the results of the imaging measurements (see Appendix A for all possible criteria) of the content analysis. All but one digital collection contained at least one fully digitised *Canterbury Tales* manuscript. The exception was the Bodleian, which had partially digitised multiple *Canterbury Tales* manuscripts. The British Library had also multiple partially digitised manuscripts, but it had one fully digitised version. The Harry Ransom Center’s (HRC) digitised manuscript was a fragment physically, but they had fully digitised that amount. Both of the latter two instances were considered sufficient to fulfil the ‘fully digitised’ criterion. Five of the digital objects did not contain any reference to the imaging process, two only mentioned the name of the imaging service (both were references to the library’s own digitisation service), and one had a brief description of the imaging process.
4.1.2 Presentation

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Criteria</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom</td>
<td>Multiple</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Once</td>
<td>1</td>
</tr>
<tr>
<td>At max zoom</td>
<td>Clear</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Blur</td>
<td>2</td>
</tr>
<tr>
<td>Material detail</td>
<td>Legible with material aspects</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Legible only</td>
<td>1</td>
</tr>
<tr>
<td>View object</td>
<td>Same window</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>New page</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Combination</td>
<td>1</td>
</tr>
<tr>
<td>Parts of individual work</td>
<td>Structured</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Unstructured but linked</td>
<td>2</td>
</tr>
<tr>
<td>View multiple items at once</td>
<td>Unavailable</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Available</td>
<td>1</td>
</tr>
<tr>
<td>Format</td>
<td>Not indicated</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>PDF</td>
<td>1</td>
</tr>
<tr>
<td>User commentary</td>
<td>Unavailable</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Social media element</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Combination</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6

Table 6 shows the results of the presentation measurements (see Appendix A for all possible criteria) of the content analysis. All but one viewer allowed multiple zooming capabilities. The exception was Early Manuscripts Oxford, a legacy site from the late 1990s which has not been updated to reflect modern technology. Most digital objects were clear at maximum zoom. All but one presented enough detail to have the digital object both be legible and reveal material aspects, such as holes and crack in the parchment (see Figure 1 for example of deep zoom, material detail, and thumbnail browse, p. 21). The approaches to viewing an object were mixed, but half of the object viewers were in the same window as the full item. Only one viewer allowed viewing of multiple objects at once (see Figure 2 for example of multiple item view and thumbnail browse, p. 22), but it was only the secondary viewer for that site, and the default primary viewer did not. No libraries indicated the image format, although one which was not explicitly stated to be a PDF in the metadata was revealed when the image had a file name end in “.pdf.” User commentary was included in presentation because it is an interface feature allowing interaction with the digital
object. Three digital objects allowed user commentary of some kind, two of which were links to social media sites.

Figure 1: Example of deep zoom, material detail, and thumbnail browse
Table 7 shows the results of the content analysis in the metadata measurements (see Appendix A for all possible criteria). Six digital items displayed information about the manuscript (MS) history or provenance. All description was at item level rather than collection level, meaning that it described the individual object in the convention of libraries rather than describing the
collection of items, as for archives. This was of interest because unique, historical material could be considered archival. Most metadata was at work level, meaning that it described the text product as a whole rather than each story within it or each page. The metadata comprehensiveness was a combination of three metadata types for six items (all lacked technical metadata) and two items contained all types. Types of metadata included bibliographic/descriptive, structural, administrative, and technical, as identified in IFLA RBSCS (2014, p. 13). However, although Early Manuscripts at Oxford technically had metadata which could be considered to fit all four types, it had by far the least amount of metadata. For example, the bibliographic/descriptive metadata only contained the author, title, date, and a note on the decoration. All but one item had added description. Bibliography was the most common, which five items in total contained. Half the items included content description.
### 4.1.4 Information Retrieval

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Criteria</th>
<th>N=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital collection search system</td>
<td>Advanced search options</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>No search</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Basic search</td>
<td>0</td>
</tr>
<tr>
<td>Search across digital collections</td>
<td>Available (some collections)</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Unavailable</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Available (all collections)</td>
<td>1</td>
</tr>
<tr>
<td>Multiple searches</td>
<td>None</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Combination</td>
<td>2</td>
</tr>
<tr>
<td>Sort by</td>
<td>Combination</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>No search</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Search but no sort, default sort</td>
<td>1</td>
</tr>
<tr>
<td>Browse by pre-made groupings</td>
<td>Available</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Unavailable</td>
<td>3</td>
</tr>
<tr>
<td>Facets</td>
<td>Available</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Unavailable</td>
<td>3</td>
</tr>
<tr>
<td>Relations between works/items</td>
<td>None</td>
<td>8</td>
</tr>
<tr>
<td>Location of collections within library website</td>
<td>Multiple collections in multiple databases</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Multiple collections in one database</td>
<td>3</td>
</tr>
<tr>
<td>Relation to library catalogue</td>
<td>In catalogue and linked to digital object</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>In catalogue and mentions/has images of digital object</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>No online catalogue</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>In catalogue and no mention of digital surrogate</td>
<td>1</td>
</tr>
<tr>
<td>Find database from library homepage</td>
<td>One link</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Multiple links</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 8

Table 8 shows the results of the IR measurements of the content analysis (see Appendix A for all possible criteria). All digital collections which had a search system had advanced search
options. The two libraries without search were the only ones which did not hold their collections in database spaces, but rather in lists of items linked to the digital objects, although one of those libraries expected users to search through the catalogue. There was mixed ability to search across different digital collections, rather than having to conduct a separate search in each collection. Five libraries allowed search across some collections (usually denying cross-collection search for collections in separate databases) and one allowed searching across all collections (which had all digital collections in one database). Six of the databases did not feature multiple searching, which allows various search strategies by saving or combining searches. Two databases had ‘combine search,’ both of which were keyword searches within results. All but one of the six libraries with search features had multiple ‘sort by’ options, the most common including sort by relevance, title, creator, and date. Five libraries allowed browsing by pre-defined groupings or clusters, such as by material type or authors. Five digital libraries contained facets in the results (see Figure 3 for example of facets, ‘cross-collection’ search, ‘sort by’, ‘advanced search’, and ‘combined search,’ p. 26). No digital items showed relations between different works, which may have aided discoverability of similar items serendipitously. All digital libraries had multiple collections, five spread across multiple databases and three in one database. There were two collections whose catalogue was not online, but of the six that were, one catalogue did not mention the digital surrogate (but there was a link to the catalogue from the digital object), and the other five catalogues had a link, images, or mention of the digital copy. Users searching resources in those catalogues would be informed of the existence of the digital copy. The number of links to find the database from the library homepage was evenly split between one and multiple clicks, which could impact ease of finding digital collections.
4.2 Interviews

The digital collections expert interviews were conducted at each of the seven institutions (see Table 1, p. 12), in eight sessions with eleven total participants (see Table 2, p. 14). The digital collections user interviews were conducted at departments with association with the University of Sheffield, with three different researchers (see Table 3, p. 15), each with a different research perspective (see Methodology, p. 15). The focus of both librarian and researcher interviews was digitisation and discoverability of historical materials rather than the Canterbury Tales manuscripts themselves, as this would have been restrictive and not all interviewees had worked with that specific manuscript. The main findings are presented below, and tables detailing the complete findings may be found in Appendix F. The interviews were conducted 27 June-18 July, 2016.

4.2.1 Librarians

The librarian interview questions followed the six themes of this dissertation (see Appendix B.1 for librarian questions), although answers naturally flowed outside this structure.
**Imaging**

Imaging was the first theme explored. When asked about the imaging process, all interviewees described the selection of digitisation material in their answer. Criteria for selection by librarians included digitisation requests, user enquires, usage statistics, research interests, original material (i.e. not digitised elsewhere), preservation considerations, and lack of copyright. Copyright restrictions and material fragility were the most common reason for material to not be selected, although one participant also mentioned information sensitivity as being another issue, such as the portrayal of indigenous people. A few librarians implied an object-based approach to some digitisation, in particular with collection treasures, meaning that such digitisation efforts were focused on selecting individual and unrelated objects rather than a group of objects, as in a collection or a project.

Three of the libraries referenced taking a mass digitisation approach for some material. This was distinguished as digitisation efforts on a broader and wider scale, which seek to digitise entire physical collections or type of material for any potential users, such as the e-codices digitisation of all medieval manuscripts in Switzerland. This digitisation could be externally or internally initiated. Reasons for undertaking this type of selection include increased access and practicality (e.g. collection already partially digitised). However, scalability and limited resources would be an obstacle for most libraries, as well as a practical desire to only digitise material which was sure to be used. One librarian also noted another issue restricting the hosting of mass collections in the main digital library:

> They’re digital images of physical things that we hold in the library, it would promote access, it would encourage discovery, it would tick lots of boxes, but it would swamp [the digital library] … it would change the experience and it would distort the results. (L4)

The main constraints with this broader approach to digitisation were funding difficulties and technological limits.

The majority of digitisation material is selected by researchers, donors, commercial partners, and curators in a project-based approach. Although in a sense all digitisation is project-based, having a beginning and an end, this approach was understood to be more designed for specific research purposes and tended to involve participation external to the library and be more focused on specific intended users. The ubiquitousness of projects is mostly due to funding, and it was noted that most libraries did not have much, or any, budget for digitisation specifically. Commercial digitisation could be problematic because the content often becomes licensed. Another
consequence of project-based digitisation is that, “they get money to deliver in that particular way; but we’re increasingly trying to bring all those collections together, because there are other people who are just … not as discipline or subject specific” (L4). Project-based digital collections are often purpose-built and do not support diverse library users. Also problematic is researchers or funders who are interested in part of a manuscript and will only pay to digitise that section. Several librarians acknowledged that the real cost of digitisation comes from all the procedures to get the item to the imaging studio. Once that has been achieved, it is almost as quick and costly to digitise the whole item.

Another key topic was the condition of historical materials. Most interviewees mentioned condition and/or conservation as a major concern during imaging, and they brought up liaison with a conservation department. Before imaging, the condition of material must be assessed to determine whether it could be imaged, and repair if necessary and within budget. One librarian noted that delicate condition could also be a reason to digitise, it in order to preserve resources digitally. The physical condition of historical materials, often fragile and faded, makes the digitisation process more expensive and time-consuming, as special allowances need to be made. Examples librarians stated include photographing material instead of scanning, dealing with tightly bound manuscripts which cannot be fully opened, and investing in an air table which sucks crumpled paper flat without damaging it. It used to be taken as a reason for digitisation that having a digital version meant less people would handle the fragile original and it would assist in preserving it. Conversely, as one librarian pointed out, “a lot of people, they get excited and they still want to come and see the original object” (L4). Some libraries now realise that digital versions of beautiful works may actually increase awareness and handling of historical materials.

**Presentation**

Several subjects stood out regarding online presentation. The structure of digital objects was an important consideration. It was generally agreed that digitally structuring an object so that it corresponded with its physical counterpart was the best representation. This involves keeping all pages in one object and in order with page numbers and perhaps titles of works within. This could only be achieved with available structural metadata and an interface capable of such digital structure. Optical Character Recognition (OCR) of handwritten historical materials was universally acknowledged as a technology continuously in development with no end in sight; however, one librarian mentioned progress in automatic identification of the layout of manuscripts. This could assist in structuring digital objects.
Most librarians mentioned high resolution and quality images as a priority. It is a particular requirement for historical material because people may want to examine the physicality of the object as well as its content. Most libraries held a policy of digitising a true and accurate representation of the physical object, which involves digitising cover to cover, not re-touching the image, and heightened concerns over colour management and size representation. This method was the official policy of most libraries, although one mentioned a perspective change from regarding each digital image as an original object in its own right to as a surrogate. Another librarian mentioned the question of enhancing an image to “optimise the resource so that it’s as useful as possible to the end-user,” (L2) or combining both approaches. These decisions are complicated by the effect of technology on presentation, as the photographer explained: “depending what monitors you’re looking at or computer system you’re using, it can completely change everything … you don’t know that this is exactly what you’re looking at, could look completely different” (L8). Also related to quality is the development or consideration in some libraries to change from TIFF formatted images to JPEG2000. Research into the presentation and preservation of the latter is still underway, and even libraries making the switch still save their TIFF files. However, this decision may be made for libraries depending on their viewer’s supported formats.

The end-user interface has a major impact on the presentation of digital objects. When asked about the approach to presenting objects, a librarian admitted, “we’re kind of governed by what [the image host] can do. It’s not bad, but it’s not perfect,” (L6) which was a sentiment several librarians shared. Some libraries were in the process of changing interface software for more modern features. Libraries choose their systems, but have little control over how they actually function. Although not specifically asked, no librarians mentioned liaison with Information Technology (IT) departments or staff. Some librarians implied that the viewer was a tool, something “just on the top, it’s the end” (L9). These interfaces and viewers are currently developing more interactive features and tools, such as deep zoom, view multiple items, annotate, download, print, share, and save items in a digital workspace. Whether the digital library can offer these features usually depends on what the software comes with, which could prevent the delivery of features even though libraries know users want them, such as download. Half the libraries also discussed experiments with multispectral imaging for new examinations of historical objects, with potential to one day present these and other versions of the same digital object. One librarian also mentioned investigating the implementation of different interfaces for various user groups. User evaluation was mainly found in the larger institutions, although the smaller ones were interested in doing so, given the appropriate resources.
Metadata

The next theme which was explored is metadata. One basic problem is that metadata creation is a time-consuming and expensive process. Most metadata is created by librarians and cataloguers, but this can be difficult when they are not an expert on the material or its content, and thinking of the right keywords can be difficult, which one librarian described. Another librarian pointed out: “no one person is an expert … is going to recognise all of these texts and what they are, especially when you’ve got many composite manuscripts [in a collection]” (L6). Some metadata may be created by academic experts of the material. Several librarians pointed out that this can create issues if the metadata becomes biased toward that academic’s research interests, which is particularly pronounced in project-based metadata. Metadata creation by crowdsourcing or automatic processes were also reported. Some structural and technical metadata is created by the imaging technicians and their equipment.

Five librarians specifically discussed the reuse of existing metadata, pulled or reproduced from the catalogue, finding aid, or content management system, as a common practice in digital libraries. This is complicated when, as is a particular issue with historical materials, catalogues are not online, are far too detailed, or records simply do not exist for those items. However created, metadata then needs to be verified and, if possible, updated as necessary to keep the records ‘living.’ As one librarian stated:

Our metadata is static, is like a snapshot of what we knew at the time, so for the research or even … authorities making changes to the way they would like metadata to be put in, means that ours is just outdated almost as soon as those changes happen. (L1)

Even if reusing metadata saves time, updating and verifying it is also a time and labour intensive process.

The level of detail in metadata is another question which libraries must face. A few mentioned a debate between quality and quantity. Most manage to find an appropriate balance, and although some librarians describe their metadata as “fairly robust and standardised” (L5) and “quite extensive” (L10), two mention a “push back” on “how much is good enough, rather than we have to so everything perfectly [sic] to make it discoverable” (L6). Going above and beyond in the metadata is simply not a viable option for most libraries given their limited staff and funding. However, the importance of metadata and the role it serves is not lost on digital librarians either:
If you have not described the item well enough and used enough terms, it will not be found … the key to finding … I keep drilling this in, is metadata. You need to have good metadata, descriptive metadata, that will allow you to bring you to the object. (L7)

When asked about the IR processes and discoverability, most answers from librarians involved metadata. This indicates that the topic is at the forefront of their minds and one of the most important steps in the digitisation process.

**Information Retrieval**

On the theme of IR, there were several other topics discussed. One was the understanding of how users search and browse, and how the digital library accommodates them. Browse was understood to mean non-specific searching, and could involve browsing a collection or within a specific digital object. The most common feature mentioned supporting browse was thumbnails, which was seen as time-saving for the user. Only the two interviewees from e-codices mentioned facets. Browsing could also be connected to specific spaces for viewing the library’s treasure objects. Another form of browsing, mentioned by one librarian, is making available the digitised archival material that has only been catalogued at series level and requires tradition sifting through digital boxes to find what could be useful.

When discussing IR, most interviewees focused on search, particularly keyword query search. It was recognised by some as a new method of conducting research and a time-saver. However, given the aforesaid limits on OCR, search ability is limited to what is contained in the metadata. Some librarians mentioned the need to think specifically about how people search for things, not just how to catalogue them. One of the biggest challenges is the likelihood that user expectations exceeds the library’s capabilities:

It’s just one search box there, which people kind of interpret as a kind of Google search, but it isn’t, really, it’s a database search … we’re of a generation now where I guess you see a search box and you think you can put anything in, and it’s gonna give you a raft of results. (L1)

Half the librarians mentioned this type of ‘Google confusion’ when discussing user expectations as problematic to the library’s discovery process. As with interfaces, library search capabilities are largely confined to what their discovery systems are capable of, which libraries cannot control, only take into consideration when choosing systems. Librarians expect that users find digital objects through the digital library itself, the main catalogue or finding aids, Google, other third party sites such as meta-catalogues, or bibliographies. The last source is particularly relevant to
humanities and historical material, and implies the continuation of traditional research methods among academics.

The development of centralisation is a key topic in several areas, but particularly relevant to IR. Libraries want to reduce the number of places where users have to search for what they need. Disparate content spread throughout unconnected databases and websites makes this difficult, particularly commercial and legacy websites. Problems were reported with ‘floating websites,’ set up to house a collection as part of a project but may not be maintained, linked, or preserved once the project ends. These sites are difficult to link to the digital library retrospectively, and could easily be lost in the digital realm. Every librarian mentioned linked data. They mostly referred to linking metadata with images, the interface, discovery system, and catalogue, but also to directly linking the catalogue record and the digitised item.

All of the libraries mentioned interoperability as a major trend. As one librarian stated: “interoperability, for us is one of the big concerns because … you don’t want to be reinventing the wheel” (L1). Five of the librarians mentioned it in conjunction with the International Image Interoperability Framework (IIIF), which they have or are in the processing of incorporating into their digital structure. A librarian explained that this would allow users to cull items from other IIIF-compliant servers and view them in the browser they are currently in, facilitating centralised IR. Interoperable metadata, especially IIIF-compliant and Text Encoding Initiative (TEI) standards, were also common. Interoperability was a frequent answer in response to the question (see Appendix B.1) about ensuring future development of digital collections.

User Needs

Concern for the needs of users and what those needs were thought to be dominated most of the librarian interviews. They believe historical collections online are used for research, of both traditional and new methods, teaching, learning, and general interest, such as with collection highlights. Libraries were aware that users want to reuse digital content for various reasons, but sometimes struggled to deliver this capability because of technology, funding, or copyright. Users may need support, particularly those without expertise, but most do not want the library to control content delivery. As one librarian explained:

People nowadays, they tend to want to create subsets of things that they find interesting, and they’re not interested in being directed. That said … there will be a direct[ory] using subject headings … so it’ll be a mixture. (L7)
Librarians think users expect to find what they search for but lack knowledge about search systems, and some lack IT skills in general.

4.2.2 Researchers

The findings of the researcher interviews are presented below. The interviews were organised according to the dissertation’s themes (see Appendix B.2 for researcher questions). These interviews addressed user needs directly, so separate user needs sections were unnecessary. The users had limited knowledge of behind-the-scenes themes imaging and metadata, although R1 had experience creating digital collections from a researcher-specific perspective.

Imaging

The first theme explored was the imaging process. The question of project-based versus mass digitisation was discussed. It was generally agreed that project-based digitisation was an overall positive because at least it makes content available given funding constraints, although the limits in broader applicability were acknowledged. One researcher noted a limitation of mass digitisation was that the tools introduced may not be relevant to actual users. They stated:

The advantage of project-based digitisation is that it’s more likely to be focused on specific needs of groups and anchored around particular research questions […] but certainly not without its limits … there needs to be an attempt to do the more project-based stuff but make sure that it’s expanding the range of stuff that we’re interested in, or not becoming too focused on one thing at the expense of another. (R2)

Essentially, project-based digital collections are acceptable as long as they consider the possibility of other users and do not limit further potential application. All academics interviewed were involved in a digital collections project of some type, suggesting increased awareness of issues involved with historical collections online, and perhaps a new method of conducting research in general. A new consideration for this dissertation revealed in the user interviews was of the differences and impact of researcher-created digital collections of historical material, which may use library material but not be part of the institution’s digital library.

Users acknowledged that the imaging process may have serious implications for research. One researcher discussed how the creation of a digital copy of an object, “usually depends on the nature of the source, the size of it, and what it is needed for” (R1). But as another researcher explained:

People don’t always recognise how things have been digitised. And the ways in which that then shapes how they can do their research and the conclusions that they’ll reach … it’s not
always made clear on websites how they’ve actually gone through the process of digitising. (R2)

This lack of user awareness of the digitisation process was seen as potentially problematic, and something libraries could address by clarifying their policies.

**Presentation**

Presentation was the next theme addressed. Both the History and English researchers indicated a desire for more “readable” images rather than completely accurate. As one researcher noted:

I think more legible’s always good … if you can enhance that preservation at the same time, I don’t see why [you wouldn’t]. The key point about that kind of thing is that it’s transparent … the risk there is that scholars end up saying things about manuscripts which … are only true in the context of general realisation. (R3)

Researchers also thought it was a problem if libraries only present poor-quality images. It was suggested that such practice was due to download and reuse, because libraries are afraid people may misuse high-quality images, for example by selling them. One researcher stated that most humanities researchers use digital resources “as a convenient way of accessing something, and then applying traditional methodological approaches … rather than necessarily wanting high-end tools to do stuff” (R2). Another researcher also mentioned traditional research methods as more common in using digital resources. Data visualisations were appreciated for assisting in the management and understanding of mass amounts of data and search results.

Issues of digital structure were revealed in the user interviews. One researcher talked of particular difficulties in searching and using digital objects which had been broken up from their physical structure. Another researcher also mentioned the specific challenges with medieval manuscripts:

It’s a three-dimensional complex object … it’s really complex, the ways in which things were written and put together … all this stuff is really important for researchers to understand … you just can’t convey that in a digitised image. (R1)

Digital object structure, or lack thereof, can have a major impact on the research and use of historical collections online.
**Metadata**

The theme of metadata was touched upon. A key point was that usefulness of metadata often depends on the user. One researcher drew a comparison of librarian-generated metadata versus the type of metadata created by and for research projects. The librarian metadata was seen to be more concerned with curation and collection management requirements, although they acknowledged the purpose bibliographical metadata served. They mentioned MARC records where it was clear a lot of time and detail went into creating records that are still inconsistent and not very useful, which is often determined by how the metadata is connected to the user interface. By contrast, they viewed researcher-created metadata as more concerned with how the data was meant to be used, and was usually descriptive, for example creating headers and tagging in the text, then “generating aspects of the search interface based on the tagging” (R1). This disparity of understanding between librarians and users influences the usefulness of metadata, and therefore the discoverability of the content. Researcher participation in digital collections may contribute to better user understanding of metadata creation, observed in how one researcher described the difficulty of thinking of terms by which people would search the collections and future usefulness.

**Information Retrieval**

IR was a theme on which users could comment more comprehensively. The decision to search or browse usually depends on the task and collection at hand for the researchers. Archives material was particularly noted by both the History and English researcher as requiring browsing because of traditional search methods. One researcher suggested that browsing is still the most common form of research. It is particularly notable for historical material that:

Certainly historians have like a real fear of missing something. A lot of history research is based on being comprehensive … they’re just used to browsing through the entire box of something … so I think they probably take that approach and that methodology and just apply it to digitised versions of things. (R2)

They also specifically stated that people are often irritated when digitised sources are not made easily browsable for this reason, but that digital humanities scholars are also increasingly engage in searching.

IR of digital historical collections through search was a common topic in the interviews. A main point was the way researchers find digitised resources. The primary approach was by asking colleagues where to find resources, which all researchers mentioned as being central and the first step in finding digital items. They all referred to general searches through Google, although they
were wary about how comprehensive those searches were. Bibliographies, library catalogues and third party websites, such as collective source pages or digital projects sites, were also mentioned as discovery methods. One researcher noted the need to regularly re-check for updates on third-party sites. When asked about finding resources without already knowing where they were, one researcher stated: “I will admit to moments of aimless searching on the Internet” (R3). This vague, “aimless” searching does not imply a great deal of confidence in their own search skills, or in the search features provided by digital collections. Keyword search was seen as time- and effort-saving, usually involving searching by basic keywords, such as authors, titles, and topics. Challenges in search involved the diversity of uses for resources and how that effects user search techniques. A desire for more centralised access and linked data, particularly metadata and user interfaces, was observed as well. One researcher summed up the problem of discovering historical resources online:

    We are kind of spoiled. But at the same time [it] can also be quite frustrating because you don’t really know how to access it or what is available, so just having that be a bit more transparent [would be a good thing]. (R3)

Ultimately, users simply want to access the resources they need in a quick, easy, and convenient manner, and of course for free.
6. Discussion

This section discusses the results which have been laid out in detail, and brings together the different sets of data, making connections with the literature. The aim of this research is to gain a deeper understanding of the challenges surrounding the digitisation and discoverability of historical materials in order to achieve greater insight into best practice and common goals. The six subthemes organised under the main themes of digitisation and discoverability were kept in place throughout the research process. It was sometimes difficult to pull apart the topic of imaging from metadata, and presentation from IR. These themes are on different sides of the digitisation and discoverability divide, but are often more connected to each other in practice because the former are back-end processes, and the latter are front-end ones. Although the themes of the planned structure occasionally overlapped, it was still useful to frame and organise the copious amount of data collected. Factors which influence digital historical libraries include projects, funding, time, material condition, liaison relations, technology, and users’ desire for clarity, particularly in how digitisation and discoverability processes may influence their research. These factors were found throughout the themes, which also had more specific concerns. This dissertation provides a comprehensive investigation of the practices and challenges in digitising and making discoverable historical material online, and addresses key issues regarding library limitations and user expectations.

The first objective of this dissertation was to understand what processes historical materials undergo in order to be digitised in practice. The two major subjects which arose from the data were selection and imaging. Selection was a key process which all experts focused on, and was also a significant step in the IFLA guidelines for digitising manuscripts (IFLA RBSCS, 2014). Librarians detailed various criteria for selection, which depends on the goal of the digitisation at hand. Concern for the condition of objects is especially important for historical materials, both because of their fragility and rarity, if not singularity. For this reason, digital historical libraries must have close relations with a conservation department. An assessment is undertaken prior to digitisation which informs the selection process according to whether some materials are able to be digitised based on their physical condition.

The approach to most digitisation is project-based, which largely directs the selection and intended use of digitised objects. IFLA RBSCS (2014) reflect this by framing all digitisation in terms of projects. It does distinguish, however, between projects digitising an existing collection, and those which create a new digital collection (IFLA RBSCS, 2014, p. 8), the former of which this
dissertation generally considered a type of mass digitisation (depending on the collection size). Practical reasons, particularly research methods and funding concerns, indicate that project-based digitisation will continue to be the normal approach for the foreseeable future, although several libraries attempt some mass digitisation. The need to accommodate diverse users online and to centralise access were major areas of discussion throughout the interviews, which conflict with piecemeal and disparate digitisation projects. Dorner et al. (2007) found that most libraries have an access-driven, bottom-up approach to digitisation, which was corroborated by libraries’ focus on user demand and research needs, but these reasons also intersected and mixed with top-down, institution strategy digitisation. Also relevant is the practice of partially digitising objects, which makes them visible but limits their actual use. The content analysis found that most libraries contained a fully digitised version of *The Canterbury Tales*, suggesting that libraries are aware of this problem. Librarians cannot always make this decision if the funder decides to only pay for partial digitisation.

The imaging process itself is clearly a vital step in the digitisation process, and in practice varies by institution, project, and object. The condition of historical materials is an important concern for librarians, and special equipment and precautions have to be taken so that no damage is incurred to the objects. This agrees with standard best practice by IFLA RBSCS (2014, pp. 11-12). France, Emery, and Toth (2010) describe the need for fragile items to have a high level of accountability in techniques for a non-destructive, non-invasive, and enduring method (p. 34). These added concerns for historical materials are beneficial for their preservation, but increase the time and money spent on digitisation. Another increase in imaging cost is the need for high resolution images which capture materiality of the items in the digital realm, the importance of which was stressed by Rekrut (2014). The findings support the advice of Austenfeld (2010) for imaging to avoid damage and maximise usefulness to users. Berger’s (2014) remark that digital images alone will never be sufficient connects to the remarks by several librarians and one researcher that seeing digital historical objects makes people want to handle the physical item more. This contradicts IFLA RBSCS’ (2014) hopeful note that a digital copy may serve as a surrogate and help ensure preservation of the original (p. 11). This way of thinking may still be promoted by guides and authorities, but in practice the opposite effect often occurs.

The second objective of this dissertation was to understand how the digitised object is presented online and the interface features available. The findings usually centred around the digital object itself or the user interface. The importance of re-creating the object’s physical structure digitally was reflected in librarian responses and the fact that six of the manuscripts were
structured as one digital object, and the last two were unstructured but still linked so the structure could be recreated. Structured digital objects are recommended by IFLA RBSCS (2014, p. 5), and their importance was explained by Austenfeld (2010, p. 149). This issue is especially important with historical materials, which may have complicated, non-modern structures to begin with.

High quality digital images are particularly important in historical collections online, shown by librarians’ emphasis in high resolution, and willingness to spend limited resources on special imaging equipment. Additionally, the content analysis found most of the digital images were clear at deep zoom and had discernible material aspects. Besides addressing the needs of material researchers, the importance of high quality images could also be connected to Le Bourgeois and Emptoz’s (2007) note that historical materials are usually only available in image mode. Librarians repeatedly stated that historical handwritten OCR is not yet technologically viable and that frequent transcription is not financially viable. The literature and librarians described ample experimentation in OCR, but successful widespread implementation has not been found in practice. Without the availability of machine-readable versions of these materials, it is all the more important that the digital versions be human-readable by being high-quality images. This has to be balanced against the standard recommendation for images not to be re-touched or enhanced, but to accurately represent the object (IFLA BSCS, 2014), which librarians supported.

A development in formats from TIFF to JPEG2000 was found in some libraries. IFLA RBSCS (2014) recommends either (p. 14), but the Wellcome Trust (n.d.) favours JPEG2000 (p. 2). The ongoing debate between cultural heritage institutions can be observed in this thread from the Digital Preservation Coalition (Parry, 2010), which suggests that libraries are slowly considering a switch to JPEG2000. The content analysis revealed that libraries do not provide format information in the public-facing metadata.

The end user interface could offer opportunities or limitations depending on with what it is equipped. The content analysis revealed that deep zoom is the most prevalent feature adopted, but most did not have the other features reviewed such as multiple item view or user commentary. However, several libraries were changing interfaces for more modern features, implying an awareness of the current limitations. This connects with librarian comments that the interface was simply a tool which came with certain features and lacked others. Similarly, when Cusimano (2014) attributes part of the digital library’s success to the interface, he focuses on its features (p. 95). These issues also arose when discussing the IR process, which is similarly tied to the interface functionality. In the digital realm, presentation and IR are the chief means of user experience. Librarians had less control or engagement in those areas, despite an overall preoccupation with user.
needs. The absence of specified collaboration with IT departments is in stark contrast to the widespread cooperation with conservation departments, which was revealed without prompting either. Digital libraries must depend on IT staff expertise, but the lack of mention may suggest that these relations are not as close or that their contribution is not viewed as central to the process.

The third objective was to examine **what researchers using these digital materials think of current digitisation practice**. With regards to imaging, users accepted project-based digitisation but hoped that creators would make accommodations for broader users. Researchers expressed a frustration with the lack of information about digital library selection because of its effect on research. Selection determines what materials are made accessible, which influences what type of research is able to be conducted on the whole. Mills (2015) examines this connection between user impact and digital Special Collections, noting that many institutions have increased user involvement in selection. She concludes that a balanced approach needs to be taken to satisfy both users and library strategy to offset the limitations of both sides (Mills, 2015, p. 167). It was most important for users that materials were made available as high resolution images online. Many researchers are at some point involved in a digital project with historical materials, so they may understand the restrictions and practicalities better than previously. The lack of awareness of imaging processes is supported with content analysis findings that information of the digitisation process and technical metadata were lacking on the public interface. Since the librarians made it clear that they collect this information, it appears they do not consider it relevant for users, a point on which researchers disagree.

Presentation of digital objects was an essential topic for users because it represents their main interaction with the digital resources. One surprising finding was the response from researchers that enhanced and more readable images may be preferred as long as it is clear what changes have been made. This contradicted the librarians’ true surrogate policies and recommendations of IFLA RBSCS (2014) and the Wellcome Trust (n.d.). While the sample size was too small for statistically significant findings and more clarification would be necessary, this at least indicates a desire by some researchers for edited digital objects to make them more usable. This relates to some libraries investigating the potential for offering different versions of digital items or different interfaces to address various user needs more specifically rather than attempting to make one general image and interface to please everyone.

Various uses of historical materials online were found, especially teaching and research by various disciplines, which was supported by librarians’ expectations and Dorner et al. (2007, p. 171), and supports the argument for different presentation based on different users. It was found
that most researchers use the same research methods on digital objects as their physical counterparts, supporting the earlier suggestion based on the findings of SHC (2009) that researchers may not greatly differentiate the two types of resources, as long as the digital copy allows them to achieve their research aims.

Digitally structured objects were important for users, as it also impacts their understanding of the digital object and the way research is conducted. Users expressed a desire for interfaces to support traditional research methods before all else, and then adding tools, particularly download. This may contradict much of the exciting, experimental research about new functionalities and interactions, but it connects with the lack of those features in actual libraries. Apart from technological restraint, some may feel the investment unnecessary.

The fourth objective was to investigate what processes are used to make digitised historical materials discoverable in practice. The processes promoting discoverability of digital objects was broken down into metadata and IR. Metadata was acknowledged as the key to discovery by both the librarians and the literature, particularly given that full-text search is not currently an option. The various functions that metadata performed were described by Foulonneau and Riley (2008), and support a need for abundant metadata which can serve all these purposes. Therefore, detailed metadata with added description is vital to allow maximum access. Most librarians had a positive outlook on their metadata quality, although some were more realistic about achieving perfection and the cost of producing it. Content analysis found that all digital objects had at least three types of metadata, most had some added description, and all had work- or page-level metadata rather than vaguer, archival collection-level. Conversely, the archives principle of provenance could be detected in that most items described the manuscript’s history. This supports Amin et al.’s (2008) point that cultural heritage experts value authenticable trust in sources (p. 39).

Centralisation and interoperability are themes which extend beyond IR, but they are particularly pertinent because they make it easier to search digital collections. Foulonneau and Riley (2008) mentioned the increasing linkage of data, particularly metadata, a trend which has continued in the various types linked data which librarians discussed. As one librarian had mentioned, it is more difficult to put into practice. Dorner et al. (2007) stressed the importance of centralised access and integrating resources with related information (pp. 180-2). Half the libraries examined had direct links from the catalogue to the digital collection or vice versa, but no libraries demonstrated relations between works. Given the limited resources available to digitisation teams and the struggle just to create appropriate metadata, this last point is not very surprising; however, it contravenes the suggestion that historical items should be connected within the collection as a
whole (IFLA RBSCS, 2014, p. 5). Centralised access is also made more difficult because five of the surveyed libraries spread their collection across multiple databases, although five allowed searching across digital collections in the same database. The project-based nature of digitisation is likely a contributing factor to the disconnection between collections, as well as other factors such as ‘floating websites.’ Interoperability is a key step to allowing separated collections to be searched, which Skinner (2014) identifies as one of the primary roles of metadata today. This is reflected in the rise of interoperable standards, particularly IIIF and TEI. Interoperability and collaboration were the key strategies libraries used to face challenges and develop together.

The IR results mainly focused on search and browse. Browse was represented, as all but Early Manuscripts Oxford allowed either faceted results or browse by groupings. Most the libraries allowed sorting of results by several different criteria, which could also assist browsing large sets of results. The most commonly mentioned browsing functionality by librarians was thumbnails, an approach often lacking formal organisation. The literature revealed a tendency for humanities researchers to prefer browsing, such as SHC (2009), so it was expected that this function would be better supported.

Search was more discussed, and content analysis found all digital collections which were held in a database had ‘advanced search’ options. Librarians acknowledged the lack of full text search caused by technological constraints, such as poor OCR, which contributed to user frustration with searching digital libraries due to ‘Google confusion.’ User misunderstanding of database vs search engine search has not gone unnoticed in IR literature (e.g. Lown, Sierra, & Boyer, 2013). While the literature recognises the difficulties of searching historical materials directly, such as non-standardised spelling, much of the research is focused on experimentally searching historical digitised materials, such as Naji and Savoy’s (2011) automatic search method and Hauser et al.’s (2007) matching procedure. While this is probably more exciting to research and fund, it does not reflect the current situation in cultural heritage institutions where most historical, particularly handwritten, material must still be searched solely by metadata.

The last objective was to examine **whether researchers find the discovery approaches useful and valuable.** Metadata usefulness depended on the user and their search techniques. One user’s differentiation between ‘librarian metadata’ and content-based metadata generated for use is at a disconnect with the prevalence of added value metadata and of librarian descriptions of thinking of appropriate ways to describe metadata according to user needs, implying that the librarians do consider these issues when creating metadata. The user focus on browsing and traditional research methods matched findings of SHC (2009) and Amin et al. (2008). However,
users also demonstrated a desire for simple keyword searching as well, preferring to mix their methods to find resources. The literature seemed to imply fact-finding searches were less valuable to humanities researchers, but data showed many researchers already know what resources they are looking for when they search digital collections, from consulting colleagues, catalogues, and bibliographies. Marchionini’s (2006) exploratory search blends query and browsing techniques, which match user description mixed methods discovery of new resources. Gueguen (2010) implied historical researchers would prefer collection-level description mirroring archives, which was not found to be the case in the content analysis, although one researcher suggested historians would be amendable to this broader search technique for appropriate archival material. Researchers agreed that most digital humanities research is based on traditional methods, the digital component chiefly serving as access, but that there is growing interest in new tools and research methods based on the digital realm. Although visualisations and search tools such as those described by Deal (2014) were not found in any libraries yet, the innovative drive and attention to user needs found in library processes and future plans suggests they may be developed in the future.

This study revealed a holistic view of the practises which historical materials undergo to be digitised and made discoverable, and the challenges faced. Most, although not all, of the processes aligned with best practice and recommendations from the literature. Main elements which affect digitisation and discoverability included projects, funding, time, material condition, liaison relations, technology, and user desire for clarity. A greater understanding of digitisation and discoverability of historical material was achieved, in particular the continuous tension between what users expect the digital library to deliver and what digital libraries are able to actually deliver. Libraries confront these challenges by developing advances collaboratively.
7. **Conclusion**

This dissertation examined the digitisation and discoverability of historical materials in order to better understand the challenges involved and ways they are overcome. The *Canterbury Tales* case study was mainly used in the content analysis and as a general focal point to compare digitisation of similar handwritten historical materials between different organisations and computer systems. Although the interviews did not discuss that manuscript in great length particularly, the focus helped identify an adequate number of exemplary libraries and digital collections of the numerous in existence. Major factors which affected the decisions made about digitisation and discoverability of historical material include projects, funding, time, material condition, liaison relations, technology, and user desire for clarity.

In order to be digitised, historical materials undergo a rigorous selection process based on the digitisation goal, after which they are imaged with special equipment as appropriate for the materials. The condition of the often fragile material is a major concern in both cases. Most digitisation is focused around a specific project although some libraries are starting to incorporate mass digitisation. Both approaches have benefits and restrictions, but for practical reasons projects are likely to continue to be the dominant method.

It was found that the presentation of the digital objects online should be structured according to their physical counterparts and high quality images should be made available without re-touching. These factors contribute to making the digital manuscript a true surrogate so that it can be understood virtually in structure and detail. Some libraries are moving to favour JPEG2000 as a format. The user interface often determines the viewing, manipulation, and search features available for interaction with digital objects. A non-central relationship with IT department was detected, which may contribute to a lack of control over the technological aspects of digital libraries.

The main problems researchers had with digitisation were less based on the methods used and more on the lack of transparency about those methods. It is easy to distort perceptions of objects and collections in the digital realm, and users want to know how this affects their research. They are more concerned with the availability of the materials online, so they generally accept project digitisation, in which they increasingly take part themselves. Although there was some interest indicated for optimised digital objects, researchers mainly used them as traditional resources, so interfaces which supported these uses were most essential. However, the many diverse user needs in the widely accessible digital realm should be addressed.
Full-text search of historical resources is not viable, so discoverability depends on metadata. This makes metadata detail and standardisation imperative, which is limited by funding and expertise. Libraries are moving towards making their collections centralised and interoperable to improve search, and for collaborative development. Basic browse and keyword search are supported, and although there is ample experimentation, advanced discovery features such as visualisations have not generally transitioned to library interfaces.

Users may not always understand the difference between databases and search engines, which can lead to frustration. Many users’ discoverability expectations are based on familiar but advanced search engines. Some find librarian-generated metadata discordant with what is useful to researchers. Humanities researchers mainly utilise browse and basic keyword searches, but certain digital needs would require advanced features.

This dissertation contributes to research because the perspective is across institutions, collections, and systems, based on what is being done in practice, not just experimentally or singularly. In this way, the research provides a comprehensive and practical analysis of the subject. The inclusion of user perspectives confirms the value of user needs in information studies, and revealed aspects of what users want or find useful which librarians had not considered.

There are several limits with this research. As noted, many librarians had limited knowledge about technology-centred topics, particularly the user interface and its features. The researcher interviews did not have the numbers or range of participants to provide abundant and diverse data for more generalisable conclusions. Also, the Canterbury Tales focus of this dissertation missed some of the newer developments in digital libraries because sometimes the manuscript was digitised earlier and held in an older database. However, the approach taken was still useful because it revealed what is actually available in digital libraries and highlighted the issue of updating or migrating older digitisation efforts.

Further study could integrate the perspective of IT professionals or digital library designers, or survey the impact of IT departments on digital libraries. Focus on most recent digitisation efforts would better explore the state of newer digitisation capabilities. The incorporation of more users, including non-researchers, would provide more perspectives and contribute knowledge about different types of user needs. A comparison against the digitisation and discoverability of non-historical documents could also lend different perspectives. These types of resources are generally easier to work with and are more suitable to automated methods, so they may have more advanced
features associated with them. Modern materials are also more likely to have a wider audience of users, so these projects may receive more funding.

A new concern introduced was the issue of digital historical collections which are not within digital libraries, usually which have been created by researchers or organisations in response to a particular research interest or question. These collections are often on disparate websites and are considered useful resources by researchers, but it is unclear how they will integrate with centralisation trends or how they will be preserved once the main project leaders leave the site. The issue of ‘floating websites’ which may continue or disappear from the Internet due to luck more than human intervention was a concern voiced by both librarian and researcher interviewees, and could constitute further research as well.

Despite the thematic structure employed throughout the research, this dissertation required flexibility and responsiveness as assumptions changed and the gap between experimentation or theory and actual practices was uncovered. New understandings were revealed and analysed about how digital historical collections are currently endeavoured and managed in practice, and in what direction the field is moving. It is hoped that this research may lead to greater awareness of the interrelation between the digitisation and discoverability processes, deeper recognition of the need for digital libraries to collaborate in historical materials digitisation, and heightened perception of diverse user needs in future digital library experimentation.

Word Count: 14,998
References


**Digital Objects Examined in the Content Analysis**


Appendices

Appendix A

Content Analysis Coding Manual

Digitisation

Quality

Amount of work digitised
1. full
2. partial

At max zoom
1. blurry
2. clear

Material detail
1. legible only
2. legible & some material aspects apparent (including holes/cracks in parchment/bindings)
3. illegible

Format
1. TIFF
2. JPEG2000
3. JPEG
4. PDF
5. other
6. not indicated

Information of digitisation process in metadata
1. description of process
2. imaging service name
3. digitization project name
4. none

Manipulation

User commentary
1. annotation
2. tagging
3. social media element
4. combination
5. other
6. unavailable

View multiple items at once
1. available
2. unavailable
Parts of individual work
1. structured (e.g. in one item)
2. unstructured (e.g. pages each own item) but linked
3. unstructured

Zoom
1. once
2. multiple
3. none

Discovery

Lookup/fact-finding

Digital collection search system
1. keyword query only
2. advanced options
3. none

Sort by
1. Default/Sort unavailable
2. Date
3. Title
4. Creator
5. Shelf mark
6. Subject
7. Other
8. Combination
9. No Search

Relation to library catalogue
1. in catalogue no mention of digital surrogate
2. in catalogue & linked to database
3. in catalogue & mentioned digital surrogate &/or contains pictures of it
4. not in catalogue
5. no catalogue online

Exploratory search/info gathering

Broad/large results

Search across digital collections
1. available- all digital collections
2. available- some digital collections
3. unavailable

Multiple searches
1. save
2. combine
3. save and combine
4. none

Browse by pre-made clusters/groupings
Facets

1. available
2. unavailable

Interpretation

Relations between different works/items

1. networks/hierarchy/tree
2. linked data
3. suggested/similar items
4. combination
5. none

Trust in source

1. provenance or MS history
2. no information

Description level

1. item level (library-based)
2. collection level (archives based)
3. combination
4. none

Metadata Level

1. work-level only
2. digital item-level (e.g. page) only
3. combination
4. other
5. none

Metadata Comprehensiveness

1. bibliographic/descriptive (describes physical object & intellectual content)
2. structural (recreate physical item from digital images)
3. image/technical (often captured by camera/scanner)
4. administrative (assist managing access to digital file)
5. combination of 2 types
6. combination of 3 types
7. all types of metadata utilised
8. none

Added description

1. content description
2. user annotation or tagging
3. bibliography
4. combination
5. other
6. none
Centralisation/ease of access

Location of collections within library website

1. 1 collection
2. multiple collections in one database
3. multiple collections in multiple databases

Find database from library homepage

1. click one link
2. click multiple links

View object

1. same window
2. popup box
3. new page
4. combination

Take note on: what metadata not included, what combinations, if no search
Appendix B

Interview Guide

Appendix B.1

Interview Guide (librarians)

Note that in practice wifi could cut out, so will call back if does, being recorded

[Reminder: topic- focus is digitized Canterbury Tales MS (remind their MS ref name & its digital collection), but questions can apply to other digitized historical materials, especially medieval MSs- control focus as much as can]

What does organization do & what do you do within it?

Imaging

1) Describe the digitization process for this object/similar historical materials
   - Goals? Drivers? Challenges of digitizing historical materials & how overcome?
     Outsource or in-house? Why not full source (hrc/bodley)?

Presentation

2) Describe library’s approach to presenting this digital object/similar historical materials online?
   - Editing or enhancement done? How decide manipulation tools offered?

Digitization User Needs

3) How did consideration of expected user needs influence digitization process?
   - Any user studies/evaluations conducted? What does the library think user needs for these kinds of digital materials are?

Metadata

4) Describe approach to metadata creation for this object/similar historical materials
   - Standards? How determine level of detail presented on user interface?

Information Retrieval

5) Describe library’s approach to enabling information retrieval of this similar historical materials
   - Focus linear search process or more interactive? Considerations for integrating enhanced search features? Challenges of discovering historical materials & how overcome?
Discovery User Needs

6) How did consideration of expected user needs influence discovery process?
   - Any user studies/evaluations conducted? What does the library think user/information needs are for finding these kinds of digital materials?

Future Plans

7) Does the library have plans regarding the maintenance of historical digital collections as digitization techniques/digital object interfaces/discovery methods become more advanced- for updating digitized materials/collections under today’s/yesterday’s technology. (not preservation question- out of scope)

Appendix B.2

Interview Guide (researchers)

Note that in practice wifi could cut out, so will call back if does, being recorded

What research currently engaged in & what digital resources do you use?

Lead in how looking at user needs, historical digital material & how it’s used/made usable online.

Information Retrieval

1. How do researchers tend to find historical digitized materials?
   - Browsing vs searching? Use digital collection databases or come in through search engine?

IR User Needs

2. How well do you believe current digital resources available fulfill user needs in the search and discovery of digitized historical materials?
   - what level of description expect or would like to see more of-item/work/collection-level? Usefulness of visualizations, facets, timelines, and other advanced search features?

Presentation/Interface

3. How do you believe researchers tend to use historical digitized materials?
   - enhanced imaging? Features find particularly useful/well designed? How compares to physical document? Feel can conduct serious research using digital surrogate?
Interface User Needs

4. How well do you believe current digital resources fulfill user needs in the presentation of digitized historical materials?
   - Different researcher user needs according to subject or faculty? Different presentation practices in different types or eras of historical materials? ‘Boutique’ or project-based digitization efforts useful or too limiting?

Criticisms

5. What do you believe are some of the greatest issues and problems on the user end in the digitization and discovery of historical materials?
   - Illegibility? Materiality?

Improvements

6. Would you suggest any improvements for the digitization and discovery of historical materials to better address user needs?
Appendix C

Ethics Approval Letter

Downloaded: 27/06/2016
Approved: 23/06/2016

Meaghan Wright
Registration number: 150112651
Information School
Programme: MA Librarianship

Dear Meaghan

PROJECT TITLE: Digitisation and Discovery of Historical Texts Using Chaucer’s Canterbury Tales
APPLICATION: Reference Number 009041

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 23/06/2016 the above-named project was approved on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 009041 (dated 10/06/2016).
- Participant information sheet 1019189 version 2 (10/06/2016).
- Participant consent form 1019190 version 1 (06/06/2016).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Yours sincerely

Matt Jones
Ethics Administrator
Information School
## Application 009041

### Section A: Applicant details

<table>
<thead>
<tr>
<th>Created:</th>
<th>Tue 17 May 2016 at 14:18</th>
</tr>
</thead>
<tbody>
<tr>
<td>First name:</td>
<td>Meaghan</td>
</tr>
<tr>
<td>Last name:</td>
<td>Wright</td>
</tr>
<tr>
<td>Email:</td>
<td><a href="mailto:mwright9@sheffield.ac.uk">mwright9@sheffield.ac.uk</a></td>
</tr>
<tr>
<td>Programme name:</td>
<td>MA Librarianship</td>
</tr>
<tr>
<td>Module name:</td>
<td>INF6000</td>
</tr>
<tr>
<td>Dissertation Last updated:</td>
<td>23/06/2016</td>
</tr>
<tr>
<td>Department:</td>
<td>Information School</td>
</tr>
<tr>
<td>Date application started:</td>
<td>Tue 17 May 2016 at 14:18</td>
</tr>
<tr>
<td>Applying as:</td>
<td>Undergraduate / Postgraduate taught</td>
</tr>
<tr>
<td>Research project title:</td>
<td>Digitisation and Discovery of Historical Texts Using Chaucer's Canterbury Tales</td>
</tr>
</tbody>
</table>
## Section B: Basic information

### 1. Supervisor(s)

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paula Goodale</td>
<td><a href="mailto:p.goodale@sheffield.ac.uk">p.goodale@sheffield.ac.uk</a></td>
</tr>
</tbody>
</table>
2: Proposed project duration

Proposed start date:
Sat 4 June 2016

Proposed end date:
Wed 31 August 2016

3: URMS number (where applicable)

URMS number
- not entered -

4: Suitability

Takes place outside UK?
No

Involves NHS?
No

Healthcare research?
No

ESRC funded?
No

Involves adults who lack the capacity to consent?
No

Led by another UK institution?
No

Involves human tissue?
No

Clinical trial?
No

Social care research?
No

5: Vulnerabilities

Involves potentially vulnerable participants?
No

Involves potentially highly sensitive topics?
No
Section C: Summary of research

1. Aims & Objectives

Digitisation efforts in libraries, archives, and other cultural heritage institutions have expanded tremendously recently. Heightened concerns about preservation challenges, access provision, and value demonstration have propelled the issue in addition to technological developments. The digitisation of historical materials is not a standardised, simple process. Once the object is digitised, it must be discoverable in order to be truly accessible. This proves difficult with historical materials. Despite such obstacles, these items are digitised. Although experts try to communicate, practices are still a chimera and too many institutions work on problems independently.

The project aim will be: to gain a deeper understanding of the challenges surrounding the digitization and discovery of historical materials in order to achieve a greater insight into best practice and common goals.

The project objectives are:
- Digitization-
  - What processes historical materials undergo to be digitized in practice
  - How the digitized object is presented online and interface features
  - What researchers using these digital materials think of current practice
- Discovery-
  - What processes are used to make digitized historical materials discoverable in practice
  - Whether researchers find the discovery approaches useful and valuable

2. Methodology

This study will take an inductive approach following interpretist and constructivist perspectives. Data will be collected in the form of content analysis and interviews.

The interface content analysis portion of the dissertation will be aimed at both the digital objects of the manuscripts and the available discovery features with which to access them. It will be conducted by myself and there are no participants. It will incorporate typical or expected tasks and supporting features and analyse how well each digitized object/digital platform performs using mixed methods. These measurements will be based on criteria from the research question and literature review.

The interviews will investigate how digital collections experts manage the challenges of digitizing historical manuscripts and how well researchers believe they have achieved their goals. They will be semi-structured interviews of about 6-8 questions. There are expected to be 7-12 interviews of experts and researchers. There may be one or more people per interview, in which case an individual informed consent will be signed by each. The experts will be those associated with the digitized manuscripts examined in the content analysis.
section, ideally people involved in the digitizing project or currently managing the collections. The researchers interviewed will be academic users of the examined materials. It is hoped that a range of disciplines may be gathered, but expected that History and English scholars will be the main study group. Thematic analysis will be conducted on the interview data.

3. Personal Safety

Raises personal safety issues? No

Personal safety management
- not entered -

Section D: About the participants

1. Potential Participants

Identification of digital library/digitization experts will be attempted through staff pages of the websites of libraries containing digitized Canterbury Tales, but some websites provide more detailed information than others. Once contact has been made, librarians can direct queries to the appropriate person. Researchers will be identified through the University of Sheffield's Humanities Research Institute, which specializes in cross-disciplinary digital humanities. If insufficient participants are found, I will also contact the UoS History and English Departments.

2. Recruiting Potential Participants

If an appropriate expert for interview can be identified a contact through the staff page, I will send an email requesting participation. Sometimes this will not be possible, in which case I will contact the closest person, e.g. head of digitisation or the librarian enquiry service, who will be able to direct me to potential interviewees. I will contact the HRI digital staff via email and request if they can recommend potential participants. If I cannot recruit an sufficient number of participants (e.g. 4-6) through this means, I will use the History & English department volunteer staff/PGR students lists.

My email will contain a quick explanation of who I am and that I am conducting a MA Dissertation. I will explain my aims and use of Canterbury Tales as a case study. For libraries I will explain why I am interested in their institution and identify the MS name and digital collection of the object I am interested in. For researchers I will explain the type of research material I will discuss (i.e. digitized historical materials, especially medieval manuscripts) and indicate I am interested in multi-disciplinary perspectives. I will explain what type and length interviews I am conducting, and provisional dates for interviewing.

2.1 Advertising methods
Will the study be advertised using the volunteer lists for staff or students maintained by CiCS? Yes

I am first contacting HRI due to their specialisation in digital humanities, but if I cannot recruit enough participants from them, the Sheffield History and English departments would be the next best pool of potential participants.

3. Consent

Will informed consent be obtained from the participants? (i.e. the proposed process) Yes

I will create an informed consent form based on the iSchool's template. All interviewees will be emailed this form in advance of the scheduled interview. The interview will not take place unless the consent form has been signed and returned.

4. Payment

Will financial/in kind payments be offered to participants? No

- not entered -

5. Potential Harm to Participants

What is the potential for physical and/or psychological harm/distress to the participants?

There is very little potential harm to participants. Perhaps if the project we are discussing was particularly stressful or frustrating, it may remind them of that.

How will this be managed to ensure appropriate protection and well-being of the participants?

The interview will be short and does not focus on the feelings of participants at the time. They will have the option not to discuss any questions they feel uncomfortable about.

Section E: About the data

1. Data Confidentiality Measures

For context, I would like to anonymise only to job title and institution. I will state this clearly on the informed consent form and allow participants to further anonymisation if desired. I will also offer participants the opportunity to read the transcripts and will redact and information they do not wish to be used in the report. The research does not involve personal information.
2. Data Storage

The interviews will be recorded on equipment rented from the iSchool or my iPhone if they are unavailable. The data will be moved and stored on my personal computer in an encrypted folder. Transcripts of interviews will be typed in Word and the files will be password protected. Backups will be kept on my personal hard drive. The data will also be stored on the Information School's research data drive which can be accessed by only by me, my supervisor, and the School's Examinations Officer and ICT staff operating the facility. This data will be deleted 3 months after the dissertation has been completed. I will have control over the data generated. The analysis will be conducted by myself. I will anonymise to job title during transcription. I will destroy the audio from the recording device once it has been moved to the computer and backed up. I will destroy the remaining data once the dissertation has been marked. The research is not externally funded.

Section F: Supporting documentation

Information & Consent

Participant information sheets relevant to project? Yes

Participant Information Sheets

Ethics_info_consent.docx (Document 022955)

Consent forms relevant to project? No

Consent Forms

Ethics_info_consent.docx (Document 022956)

Additional Documentation

None

External Documentation

- not entered -
Section G: Declaration

Signed by:
Meaghan Wright
Date signed:
Fri 10 June 2016 at 16:15
Appendix E

Informed Consent Form

The University of Sheffield Information School

MA Dissertation:
Digitisation and Discovery of Historical Texts Using Chaucer’s *Canterbury Tales*

---

Researchers
Meaghan Wright
mwright9@sheffield.ac.uk

Purpose of the research
The purpose of this research is to gain a deeper understanding of the challenges surrounding the digitization and discovery of historical materials in order to achieve a greater insight into best practice and common goals. It investigates the processes undertaken and user perceptions in both digitization and discovery of these materials.

Who will be participating?
Librarians and digitization experts from relevant digital libraries containing the case study material.
Researchers from relevant disciplines who use similar digitized material in their work.

What will you be asked to do?
An interview of 30-60 minutes about the challenges faced, how they are overcome, and whether participants feel this could be improved.

What are the potential risks of participating?
The risks of participating are the same as those experienced in everyday life.

What data will we collect?
Interviews will be audio recorded.

What will we do with the data?
I will be analyzing the data for inclusion in my Masters dissertation. The data will be stored on the Information School's research data drive which can be accessed by only by me, my supervisor, and the School’s Examinations Officer and ICT staff operating the facility. This data will be deleted 3 months after the dissertation has been completed. I will also store a password protected back up copy on my personal laptop and hard drive. I will destroy data once the dissertation has been marked.
Will my participation be confidential?
I would like to anonymise to job title and institution to provide context for the data. If participants would like further anonymization, they may indicate so below. If participants would like to review the interview transcripts before analysis is conducted, they may also indicate so below.

What will happen to the results of the research project?
The results of this study will be included in my master’s dissertation which will be publicly available. Please contact the School in six months.

I authorize my data to be linked to:

1) ☐ My institution
2) Either: ☐ My job title OR ☐ A generic role

Which shall be described as:

☐ I would like to review my interview transcripts before analysis

- I confirm that I have read and understand the description of the research project, and that I have had an opportunity to ask questions about the project.
- I understand that my participation is voluntary and that I am free to withdraw at any time without any negative consequences.
- I understand that if I withdraw I can request for the data I have already provided to be deleted, however this might not be possible if the data has already been anonymised or findings published.
- I understand that I may decline to answer any particular question or questions, or to do any of the activities.
- I understand that my responses will be kept strictly confidential, that my name or identity will not be linked to any research materials, and that I will not be identified or identifiable in any report or reports that result from the research, unless I have agreed otherwise.
- I give permission for all the research team members to have access to my responses.
- I give permission for the research team to re-use my data for future research as specified above.
- I agree to take part in the research project as described above.

Participant Name (Please print) ____________________________ Participant Signature ____________________________
Researcher Name (Please print)    Researcher Signature

Date

Note: If you have any difficulties with, or wish to voice concern about, any aspect of your participation in this study, please contact Dr Jo Bates, Research Ethics Coordinator, Information School, The University of Sheffield (ischool_ethics@sheffield.ac.uk), or the University Registrar and Secretary.
## Appendix F

### Complete Findings

#### Librarians

<table>
<thead>
<tr>
<th>Imaging</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conservation Considerations</strong></td>
<td><strong>Digital Structure</strong></td>
</tr>
<tr>
<td>▪ Assess object condition before digitise</td>
<td>▪ Structured digital object</td>
</tr>
<tr>
<td>o Usually first step in digitisation of historical materials; need conservator approval</td>
<td>o Development- full work in one record, based on physical structure</td>
</tr>
<tr>
<td>o May require treatment/repair</td>
<td>o Development in progress- automatically recognise text layout</td>
</tr>
<tr>
<td>o May require item deselection if cannot treat</td>
<td>▪ Unstructured object</td>
</tr>
<tr>
<td>o Fragile material- may digitise to preserve digitally or not digitise to preserve physically</td>
<td>o Each page/image of work separate record, unordered</td>
</tr>
<tr>
<td>o Constraint- whether have budget for conservation</td>
<td>o May be linked</td>
</tr>
<tr>
<td>▪ Liaison with conservators</td>
<td>o Constraint- time consuming to create record for each page/image</td>
</tr>
<tr>
<td>o Department often works closely with digitisation team</td>
<td>▪ Structured digital collection</td>
</tr>
<tr>
<td>o Digitisation may inform conservation work- e.g. multispectral imaging</td>
<td>o Group by themes, period, or types of materials</td>
</tr>
<tr>
<td>▪ Conservation good for materials</td>
<td>o Development- some grouped closer to physical collections</td>
</tr>
<tr>
<td>▪ Digital surrogate means less handling of original object by users</td>
<td>o Representative of physical collections</td>
</tr>
<tr>
<td>▪ Or potentially the opposite effect- raises awareness</td>
<td>o Constraint- users may not know where to look</td>
</tr>
<tr>
<td>▪ Development- awareness of conservation issues</td>
<td></td>
</tr>
<tr>
<td>▪ Constraint- time consuming and expensive</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selection</th>
<th>Features and Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Influences research- what is available and accessible online</td>
<td>▪ Development- interactive features</td>
</tr>
<tr>
<td>▪ Select based on digitisation requests, enquires, usage, research interests, original material (i.e. not already digitised elsewhere), preservation considerations, public domain</td>
<td>o Annotate or tag object</td>
</tr>
<tr>
<td>▪ Reactive vs proactive approach</td>
<td>o Download, print, share items- better for library if links/pulls digital object instead of download so can prove value (funding)</td>
</tr>
<tr>
<td>▪ Project based digitisation</td>
<td>o Zoom- in and out</td>
</tr>
<tr>
<td>o Liaison with researchers</td>
<td>o Compare images/work/pages</td>
</tr>
<tr>
<td>o Liaison with commercial or third party partners- may lead to licensing and paywalls</td>
<td>o User workspace or saved items</td>
</tr>
<tr>
<td>o Funding- commercial digitisation orders generate small income</td>
<td>▪ View different versions of digital item</td>
</tr>
<tr>
<td>o Constraint- digitise and present in very specific way</td>
<td>o Black and white, greyscale, multispectral</td>
</tr>
<tr>
<td>o Liaison with curators</td>
<td>▪ Constraint- depends on what comes with viewer or interface</td>
</tr>
<tr>
<td>▪ Collection based digitisation</td>
<td></td>
</tr>
</tbody>
</table>
- Liaison with curators
- Mass digitisation- scalability
- Finite amount of material- possible to digitise while collections with time
- Access does not mean will be used
- May be worthwhile if have already digitised proportion of collection
- Constraint- may disproportionately overwhelm digital library
  - Constraint- libraries may favour non-copyrighted content to digitise

### Actual Imaging Process

- Time consuming and expensive
- Liaison with imaging technicians
  - Dependant on resources and staff expertise; may need to outsource
  - Constraint- lack expertise on materials themselves; may not know what item is
- Photography
  - Development- quick, high quality cameras
  - Better for historical material- especially manuscripts, e.g. with tight binding or fragile pages
  - Constraint- more labour intensive
- Scanning

### Technology

- Viewer interface
  - Just a tool, on top
  - Development- several libraries switching to Mirador or Universal Viewer
  - Constraint- libraries choose but cannot control
<table>
<thead>
<tr>
<th>Quality</th>
<th>True Digital Surrogate</th>
</tr>
</thead>
<tbody>
<tr>
<td>- High resolution images</td>
<td>- Development- more common policy of faithful representation of original item</td>
</tr>
<tr>
<td>- Necessary for digital historical materials</td>
<td>- Do not touch up</td>
</tr>
<tr>
<td>- Quality vs quantity</td>
<td>- Development- digital object not as important to preserve as physical item</td>
</tr>
<tr>
<td>- More items accessible or less items more usable</td>
<td>- Colour and size representation</td>
</tr>
<tr>
<td>- Constraint- high quality images are larger</td>
<td>- Cropping, ruler, or colour card</td>
</tr>
<tr>
<td>- Development- abundant digital storage and fast networks</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standards</th>
<th>Enhanced Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Development in progress- standardise image quality throughout</td>
<td>- Optimise and make more attractive and/or readable</td>
</tr>
<tr>
<td>collections</td>
<td>- May be more useful to end user</td>
</tr>
<tr>
<td>- Development</td>
<td>- Constraint- time consuming</td>
</tr>
<tr>
<td>- Colour and size representation</td>
<td>- Constraint- not how real item looks; users may misinterpret</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Technology</th>
<th>May Influence Academic Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Need proper technology for full imaging process</td>
<td>- How see the image</td>
</tr>
<tr>
<td>- E.g. cameras, scanners, lighting, cradles, climate control, ingest</td>
<td>- May depend on monitor, browse, editing of image, etc.</td>
</tr>
<tr>
<td>system, etc.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Need high quality and preservation level standards for historical</td>
</tr>
<tr>
<td>materials</td>
</tr>
<tr>
<td>- May differ for various purposes</td>
</tr>
<tr>
<td>- Preservation, delivery, download</td>
</tr>
<tr>
<td>- TIFF, common standard</td>
</tr>
<tr>
<td>- Long recognized, high quality, preservation level standard</td>
</tr>
<tr>
<td>- Development in progress- JPEG2000, common standard</td>
</tr>
<tr>
<td>- High-quality, preservation standard, flexible</td>
</tr>
<tr>
<td>- Newer- not universally accepted as preservation level</td>
</tr>
<tr>
<td>- Some libraries switching to (without abandoning TIFF files)</td>
</tr>
<tr>
<td>- Constraint- maybe some differences, e.g. colour representation,</td>
</tr>
<tr>
<td>small data loss</td>
</tr>
<tr>
<td>- Constraint- image viewer may only support certain formats</td>
</tr>
</tbody>
</table>

- Different types- e.g. flatbed or overhead
- Depends on nature of item- e.g. how it opens, size
- Create technical metadata during process
- Digitise original or surrogate
  - Digitise microfilm- cheaper, quicker, and easier but not as high quality
- Colour management
  - Colour truly represents original
  - May present in colour card
  - Development- encode in technical metadata
- Digitise full or partial item
  - True surrogate- need to digitise full item, including cover, margins, binding, etc.
  - Research interests- may only want or be willing to pay for partial item digitisation
  - Funding- Real cost in getting item to imaging studio, from there may as well digitise whole item
- Constraint- funding
  - Good equipment expensive; cheap equipment poor quality
- Constraint- conditions of imaging studio may not be ideal for conservation of materials
<table>
<thead>
<tr>
<th>Multi-/Hyperspectral imaging</th>
<th>How expected to use digital object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraint- technology advances rapidly</td>
<td>Digital libraries try to allow broad access</td>
</tr>
<tr>
<td>o Want to digitise for once, but digital resources quickly outdated</td>
<td>Digital item out of context</td>
</tr>
<tr>
<td>o Question of re-digitising for better quality images</td>
<td></td>
</tr>
<tr>
<td>o Multiple versions of image</td>
<td></td>
</tr>
</tbody>
</table>

**Funding**
- Constraint- may not have much budget for digitisation
- Funded projects are the main way to pay for digitising resources
- Liaison with external funders or donors
- Liaison with commercial partners

**Multilingual**
- Development in progress- interface allow search and provide information for multilingual audiences
- May depend multilingualism of host site’s country
- Constraint- staff expertise; lack funding to outsource

**Workflow**
- Ensure capacity and avoid bottlenecking along process
- Track images as ingested into library systems
  - o Start with metadata
  - o Constraint- system or human error
- Linked data- image with metadata record
- Unique identifiers
  - o Constraint- not human-friendly
- Development in progress- review and revise workflow infrastructure

**Interoperability**
- Development in progress- IIIF standard
  - o Allows display content from other IIIF-compliant collection
- Development- from silos to dynamic and interactive digital libraries
- Display other formats/types of collections in own browser
- Centralise- fewer access points
- Potential future proofing
  - o Common systems, formats, standards, etc.
  - o Libraries move forward together

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Information Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating Records</td>
<td>Search</td>
</tr>
<tr>
<td>o Time consuming and expensive</td>
<td>o Find specific item(s)</td>
</tr>
<tr>
<td>o Finding- takes up most of digitisation budget</td>
<td>o Keyword search</td>
</tr>
<tr>
<td>o Discovery dependant on quality and reliability of metadata</td>
<td>o New type of research from traditional</td>
</tr>
<tr>
<td>o Reuse information- from catalogue, finding aids, collection management system</td>
<td>o Known item lookup- researchers often know what looking for; search by shelfmark</td>
</tr>
<tr>
<td>o Constraint- if catalogue not online or items uncatalogued</td>
<td>o Constraint- manuscripts can only search by metadata</td>
</tr>
<tr>
<td>o Constraint- catalogue record may not be compatible with digital object record- e.g. too detailed (especially manuscript catalogues)</td>
<td>o Constraint- Google confusion; people expect database to work like Google</td>
</tr>
<tr>
<td>o Constraint- older (embedded or out-of-date) metadata</td>
<td>o Search within item</td>
</tr>
<tr>
<td>Development in progress- pulling information rather than reproducing it</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Metadata creators</td>
<td></td>
</tr>
<tr>
<td>o Specific digital project- biased towards that research interest or intended use</td>
<td></td>
</tr>
<tr>
<td>o Liaison with academics- expertise in content; biased toward own research interests</td>
<td></td>
</tr>
<tr>
<td>o Librarians/cataloguers- limited content expertise</td>
<td></td>
</tr>
<tr>
<td>o Automatic- often technical metadata</td>
<td></td>
</tr>
<tr>
<td>o Imaging technicians- technical and structural metadata; image identifiers</td>
<td></td>
</tr>
<tr>
<td>o Other libraries- e.g. if hold physical content; may be limited by resources and staff expertise</td>
<td></td>
</tr>
<tr>
<td>o Development in progress- users in value added features, e.g. annotation or tagging</td>
<td></td>
</tr>
<tr>
<td>o Constraint- after download object, users may change metadata and share</td>
<td></td>
</tr>
</tbody>
</table>

| Not full text but limited capabilities with correct metadata- e.g. search for illustrations |
|--------------------------|--------------------------|
| Across collections |
| Search via- |
| o Digital collection database |
| o Main library catalogue/find aids |
| o Google- no control over what Google retrieves or how prioritises |
| o Bibliographies |
| o Meta-catalogues and third party sites- no control over; mixed success |
| Development in progress: interactive search |
| o Facets and drop-down menus |
| Constraint- digital library systems do not work like Google, e.g. correct misspellings or use clever algorithms (development in progress) |

| Multilingual |
|--------------------------|--------------------------|
| Language of historical material |
| o Constraint- cataloguer linguistic expertise |
| Development in progress- allow multilingual search and information provision |
| More likely if have multilingual partner in project |

| Mass Content |
|--------------------------|--------------------------|
| More items in collections- harder to keep simple and easy to use |
| o Constraint- can swamp retrieval system, makes it hard to find what want |

| Level of Detail |
|--------------------------|--------------------------|
| Detailed or basic metadata |
| o Detail enhanced discoverability |
| o Mass digitisation- may only be able to do basic |
| o Funding- may vary within collections depending on individual project budgets |
| o Constraint- detailed and slow or minimal but available, may depend on library policy |
| Specialists want very detailed metadata |
| Keywords |
| o How expect people to search items |
| Item level metadata most common |
| o Page level- not sustainable, very time consuming |
| o Collection level- archive material, may be only information available |

<p>| Browse |
|--------------------------|--------------------------|
| Find non-specific item(s) |
| Browse digital item collections |
| Browse digital item structure |
| Thumbnails- most common method of browsing |
| o Time saving- for user |
| Space to browse library highlights/treasures |
| May parallel tradition research methods |
| o E.g. hierarchy and series ordering |</p>
<table>
<thead>
<tr>
<th>Standards</th>
<th>Availability of Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Follow existing metadata standards</td>
<td>▪ Online content</td>
</tr>
<tr>
<td>o Various kinds- Dublin Core most common for historical material</td>
<td>▪ Constraint- user awareness</td>
</tr>
<tr>
<td>o Development in progress- TEI for detail and tagging</td>
<td>o Social media and communications efforts</td>
</tr>
<tr>
<td>o Development in progress- IIIF for interoperability</td>
<td>▪ Constraint- copyright and permissions</td>
</tr>
<tr>
<td>▪ Custom metadata schema based on standards</td>
<td>▪ Constraint- licensed and paywalled content</td>
</tr>
<tr>
<td>▪ Constraint- digital collections made up of diverse types of material</td>
<td>o Development- more flexible contracts with commercial partners for limited reuse</td>
</tr>
<tr>
<td>o Some standards better suited for certain types</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correct/Verify Records</th>
<th>Centralise</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ When reuse metadata from other sources- ensure smooth migration and database compatibility</td>
<td>▪ Interoperability- IIIF</td>
</tr>
<tr>
<td>▪ If use automated metadata creation</td>
<td>▪ Linked data</td>
</tr>
<tr>
<td>▪ If crowdsource/outsource metadata creation- still need to verify quality</td>
<td>o Main catalogue record and digitised item</td>
</tr>
<tr>
<td>▪ Keep records ‘living’</td>
<td>o Shared digital objects- pull or directly link item</td>
</tr>
<tr>
<td>o Authorities change standards</td>
<td></td>
</tr>
<tr>
<td>o New research changing content information</td>
<td>▪ Development in progress- search multiple silos to search central search platform</td>
</tr>
<tr>
<td></td>
<td>▪ Downloaded digital object to originating digital library</td>
</tr>
<tr>
<td></td>
<td>▪ Limited access points</td>
</tr>
<tr>
<td></td>
<td>o May direct people to catalogue</td>
</tr>
<tr>
<td></td>
<td>o Constraint- some content only available on third party, paywalled, or older websites</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Centralise</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Linked data</td>
<td>▪ Interface serves diverse audiences</td>
</tr>
<tr>
<td>o Metadata with: image, interface, discovery system, catalogue</td>
<td>▪ Constraint- search depends on discovery system capabilities</td>
</tr>
<tr>
<td>o Metadata sources</td>
<td>o Libraries choose but cannot control</td>
</tr>
<tr>
<td>o Constraint- difficult to do in practice</td>
<td>o May have limited functionality</td>
</tr>
<tr>
<td>▪ Name identity recognition (NER); author and (development in progress) related people index</td>
<td>o May be proprietary</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Types of metadata</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ Technical</td>
<td></td>
</tr>
<tr>
<td>▪ Structural</td>
<td></td>
</tr>
<tr>
<td>▪ Descriptive</td>
<td></td>
</tr>
<tr>
<td>▪ Submission information package</td>
<td></td>
</tr>
<tr>
<td>▪ Preservation</td>
<td></td>
</tr>
</tbody>
</table>

| Funding-                                                                 |                                                                                          |
- Metadata expensive and labour consuming to create
- Constraint- historical materials structure and demand
  - More resources to cataloguing other library departments- e.g. textbooks and articles

<table>
<thead>
<tr>
<th>User needs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Uses</strong></td>
</tr>
</tbody>
</table>
| - Research  
  - Traditional and new methods  
- Reuse  
  - Data and images  
  - Purpose: academic, public, and commercial  
- Public interest and engagement  
  - Especially collection highlights and local history  
- Teach and learn  
- Library monitors digital use |
| **User Expectations** |
| - Material available online  
- Free  
- Easy to use  
  - May lack IT skills  
- Support and information  
  - Want a mix, not too directive  
- Clear library policies  
- To find what they search for  
  - Google confusion- may think all search boxes work like Google  
- Learn through user feedback and/or study |
| **Diverse Users** |
| - Researchers  
- Diverse Disciplines  
  - Medieval manuscripts: especially historians, palaeographers, and digital humanities researchers  
- Students  
  - University and schools  
- Multilingual and multinational  
- Attached to institution or digital-only users  
  - Hierarchy of priority users |

**Researchers**
<table>
<thead>
<tr>
<th>Imaging</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selection of materials</strong></td>
<td><strong>Structured Digital Object</strong></td>
</tr>
</tbody>
</table>
| ▪ Based on research interests  
  ○ Constraint- Only useful for that research interest | ▪ Digital vs physical object structure  
  ○ Constraint- difficult to use or search when digitally broken up  
  ○ Constraint- medieval manuscripts have layers and versions of text  
  ○ Influences research- important to understand how object created  
  ○ Constraint- cannot convey in digital image |
| ▪ Project-based digitisation/Boutique digital collections  
  ○ Positives- well matched for specific users; makes material available  
  ○ Negatives- limit broader users; material only available in specific context | |
| ▪ Mass digitisation  
  ○ More content available  
  ○ Constraint- Irrelevant tools | |
| **Imaging Process** | **Quality** |
| ▪ Digitisation process unclear in digital collections  
  ○ Users lack awareness of limitations  
  ○ Constraint- Influences research and findings | ▪ Poor quality digitisation-  
  ○ Smaller libraries cannot afford better  
  ○ Constraint- less discoverability & inferior user experience  
  ▪ True digital surrogate  
  ○ Eliminate need to see originals  
  ○ Potentially opposite effect  
  ▪ Enhanced & optimal  
  ○ Influences research- describe item based on how represented onscreen rather than how actually is  
  ○ More readable image is more useful- as long as made clear to researchers |
| ▪ Constraint- technology changes | |
| **Availability** | **Transcription** |
| ▪ Influences research- what has been digitised | ▪ OCR of historical materials- poor  
  ▪ Constraint- medieval manuscripts lack standardised spelling  
  ▪ Influences research- how content is transcribed  
  ○ Interpretive process  
  ○ Automatic or manual- different results |
| **Multiple Interfaces** | **Multiple Interfaces** |
| ▪ Based on type of user  
  ○ How metadata is made useful to end-user in interface |
Features and Tools

- Influences research
  - How search and use digital collections or items
- Constraint- depends on the website collections are hosted within

Data Visualisations

- In progress development
- Manage and understand mass data and search results
- Useful for research publication
- Influences research
  - Expect researchers will rely on
  - Constraint- usefulness depends on how use and what look for
  - Constraint- accuracy depends on data itself
  - User awareness of limitation

Technology

- Many standards
  - Open/proprietary, formats, databases, programming languages
  - Depends on type of resource and intended use
- Older or poorly designed technology
  - Constraint- cannot do as much with those resources
- OCR on historical sources still poor
  - Development in progress- but first need to advance OCR of printed materials before handwritten can improve

<table>
<thead>
<tr>
<th>Metadata</th>
<th>Information Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Detail</strong></td>
<td><strong>Browse</strong></td>
</tr>
</tbody>
</table>
| - Add value
  - Tags, semantic markup, added structure
  - Content of digital item
  - Important for how people will search
  - Include physical item information
  - Usefulness- can be very detailed but not useful to users
  - Constraint- inconsistencies
  - Time consuming | - Still most common technique by digital humanities researchers
  - Most mix methods- search and browse
  - Historians particularly comprehensive in research
  - Constraint- some digital resources difficult to browse through
  - Traditional research methods in a digital setting |
<table>
<thead>
<tr>
<th>Usefulness depends on user</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Librarian</td>
<td></td>
</tr>
<tr>
<td>o Bibliographic metadata</td>
<td>Increasingly used technique by digital humanities researchers</td>
</tr>
<tr>
<td>o Use for classification, archiving, curation management</td>
<td>o Users increasingly aware of searching constraints</td>
</tr>
<tr>
<td>o How item fits in wider collection</td>
<td>o Keyword search</td>
</tr>
<tr>
<td>o Constraint- users unaware how librarians catalogue</td>
<td>o Search by basic terms or topic of interest</td>
</tr>
<tr>
<td>User</td>
<td>o Saves time and effort</td>
</tr>
<tr>
<td>o Descriptive and tagging metadata</td>
<td>o Constraint- may not return all relevant results</td>
</tr>
<tr>
<td>o Use for search and research</td>
<td>o Constraint- poor OCR or metadata</td>
</tr>
<tr>
<td>o Not necessarily interested in wider collection</td>
<td>o Constraint- historical spelling</td>
</tr>
<tr>
<td>o If metadata created for research project (i.e. by main user)- detailed and specifically useful; focused only on those research interests and not generally useful</td>
<td>o Constraint- cannot search within manuscript images</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standards</th>
<th>Search Via</th>
</tr>
</thead>
<tbody>
<tr>
<td>In particular- TEI and XML</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Colleagues</td>
</tr>
<tr>
<td></td>
<td>o Very common method of finding new materials, especially in humanities</td>
</tr>
<tr>
<td></td>
<td>o Constraint- may not know everything that is available</td>
</tr>
<tr>
<td></td>
<td>Google</td>
</tr>
<tr>
<td></td>
<td>o General search</td>
</tr>
<tr>
<td></td>
<td>o Resources do not already know about</td>
</tr>
<tr>
<td></td>
<td>o Hope results on first page</td>
</tr>
<tr>
<td></td>
<td>o Constraint- Google does not find or prioritise what search for</td>
</tr>
<tr>
<td></td>
<td>Library catalogue</td>
</tr>
<tr>
<td></td>
<td>o Own university/research library</td>
</tr>
<tr>
<td></td>
<td>o May have search features- e.g. search only digital collections</td>
</tr>
<tr>
<td></td>
<td>Third party websites</td>
</tr>
<tr>
<td></td>
<td>o Union catalogue, databases, or collective websites/projects</td>
</tr>
<tr>
<td></td>
<td>o More reliable than Google</td>
</tr>
<tr>
<td></td>
<td>o Re-check if updated content</td>
</tr>
<tr>
<td></td>
<td>o Constraint- if does not find what search for</td>
</tr>
<tr>
<td></td>
<td>o Constraint- too many portals, not all are comprehensive</td>
</tr>
<tr>
<td>Quality</td>
<td>Centralise</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------------------------</td>
</tr>
</tbody>
</table>
| ▪ Mechanical means improve searching  
  o Constraint- not very good with historical materials  
  ▪ Influences research  
  o Unreliable search means cannot make statistical conclusions | ▪ Access points  
  o Easier to search if many resources in one place  
  ▪ Linked data  
  o In particular- metadata and search interface | ▪ Mass information- difficult to access or know what is available  
 ▪ Constraint- copyright and licensed/paywalled resources |

**User Needs**

**Uses**
- Research  
  o Not possible/easy before digital collections- large-scale data analysis, and quantitative and statistical research  
  o Tradition methods- most common digital collections research; may not want or need high-end skills/features  
- Teaching and Learning  
  o In particular for manuscripts- palaeography and material studies  
- General Interest

**User Expectations**
- Access to resources  
  o Free  
  o Convenient and easy  
  o Want more materials digitised  
  o Most digital humanities projects deal with non-digital documentary evidence- need to digitise first  
- Browseable collections and resources  
- Search and find what need  
  o Quickly
- Recall - humanities want to find everything relevant
- Google confusion - may not understand that not all search boxes work like Google
- Aware databases more reliable, want improvements
  - High quality digital objects
  - Download and reuse resources

### Diverse Users

- Collections designed for narrow purpose less useful to other users
- Academic researchers
  - Different needs based on discipline and research
  - General academics vs those involved in digital project
- Teachers and students
  - University and school
Access to Dissertation

A Dissertation submitted to the University may be held by the Department (or School) within which the Dissertation was undertaken and made available for borrowing or consultation in accordance with University Regulations.

Requests for the loan of dissertations may be received from libraries in the UK and overseas. The Department may also receive requests from other organisations, as well as individuals. The conservation of the original dissertation is better assured if the Department and/or Library can fulfill such requests by sending a copy. The Department may also make your dissertation available via its web pages.

In certain cases where confidentiality of information is concerned, if either the author or the supervisor so requests, the Department will withhold the dissertation from loan or consultation for the period specified below. Where no such restriction is in force, the Department may also deposit the Dissertation in the University of Sheffield Library.

To be completed by the Author – Select (a) or (b) by placing a tick in the appropriate box

If you are willing to give permission for the Information School to make your dissertation available in these ways, please complete the following:

- X (a) Subject to the General Regulation on Intellectual Property, I, the author, agree to this dissertation being made immediately available through the Department and/or University Library for consultation, and for the Department and/or Library to reproduce this dissertation in whole or part in order to supply single copies for the purpose of research or private study

- (b) Subject to the General Regulation on Intellectual Property, I, the author, request that this dissertation be withheld from loan, consultation or reproduction for a period of [ ] years from the date of its submission. Subsequent to this period, I agree to this dissertation being made available through the Department and/or University Library for consultation, and for the Department and/or Library to reproduce this dissertation in whole or part in order to supply single copies for the purpose of research or private study

Name Meaghan Wright

Department Information School

Signed Meaghan Wright Date 29 August 2016
To be completed by the Supervisor – Select (a) or (b) by placing a tick in the appropriate box

(a) I, the supervisor, agree to this dissertation being made immediately available through the Department and/or University Library for loan or consultation, subject to any special restrictions (*) agreed with external organisations as part of a collaborative project.

(b) I, the supervisor, request that this dissertation be withheld from loan, consultation or reproduction for a period of [ ] years from the date of its submission. Subsequent to this period, I, agree to this dissertation being made available through the Department and/or University Library for loan or consultation, subject to any special restrictions (*) agreed with external organisations as part of a collaborative project

Name

Department

Signed Date

THIS SHEET MUST BE SUBMITTED WITH DISSERTATIONS BY DEPARTMENTAL REQUIREMENTS.