Discovering digitized special collections: An investigation of researchers’ practices and priorities

Thesis submitted in partial fulfilment of the requirements for the degree of MSc in Library Science at City University London

Emma Stanford
December 2016
word count: 21,079
Acknowledgments

Thanks to David Bawden, my supervisor; Christine Madsen, who provided valuable advice as well as an admirable example; Judith Siefring and other Bodleian Libraries staff for their understanding and cooperation; and the scholars who generously gave up their time for interviews.
# Table of Contents

1 Introduction ................................................................................................................. 4

2 Aims, Objectives, Scope and Definition ........................................................................ 5
   2.1 Aims and Objectives................................................................................................. 5
   2.2 Scope and Definition............................................................................................... 6

3 Literature review ........................................................................................................... 7
   3.1 Information seeking behavior of humanists.......................................................... 7
   3.2 Resource discovery in special collections.............................................................. 8
   3.3 Digital collections.................................................................................................. 9

4 Methods ....................................................................................................................... 13
   4.1 Questionnaire ........................................................................................................ 14
      4.1.1 Design................................................................................................................. 14
      4.1.2 Dissemination..................................................................................................... 14
      4.1.3 Analysis............................................................................................................... 15
   4.2 Interviews .............................................................................................................. 15
      4.2.1 Recruiting........................................................................................................... 16
      4.2.2 Interview process............................................................................................... 17
      4.2.3 Transcription and analysis............................................................................... 18
      4.2.4 Managing bias.................................................................................................... 19

5 Analysis ..................................................................................................................... 19
   5.1 Questionnaire ........................................................................................................ 19
      5.1.1 Demographic breakdown.................................................................................. 20
      5.1.2 Discovery........................................................................................................... 20
      5.1.3 Opinions on existing resources ........................................................................ 24
      5.1.4 Research functions of digitized special collections ......................................... 26
      5.1.5 Metadata.......................................................................................................... 27
      5.1.6 Important features of digitized items ................................................................. 28
      5.1.7 Downloading...................................................................................................... 29
      5.1.8 Research planning.............................................................................................. 29
      5.1.9 Preferred method for finding new digitized items ............................................ 30
      5.1.10 Citation............................................................................................................. 30
      5.1.11 Questionnaire conclusions................................................................................ 31
5.2 Interviews .................................................................................................................................... 31
  5.2.1 Role of digitization in research ........................................................................................... 31
  5.2.2 Discovery methods .............................................................................................................. 34
  5.2.3 Citation ................................................................................................................................ 41
  5.2.4 Desires ................................................................................................................................. 43

6 Conclusion ........................................................................................................................................... 50
  6.1 Recommendations ...................................................................................................................... 52
  6.1.1 Creating discoverable resources ......................................................................................... 52
  6.1.2 Creating usable, reliable metadata ...................................................................................... 53
  6.1.3 Creating images that facilitate scholarship ......................................................................... 54
  6.1.4 Dissemination and instruction ............................................................................................ 55
  6.2 Limitations of this research ......................................................................................................... 56
  6.3 Future directions ......................................................................................................................... 57

7 Appendices .......................................................................................................................................... 58
  7.1 Appendix A: Questionnaire ......................................................................................................... 58
  7.2 Appendix B: Questionnaire Responses ....................................................................................... 59
  7.3 Appendix C: Interview Transcripts .............................................................................................. 60
    7.3.1 Role of digital surrogates in research ................................................................................. 60
    7.3.2 Resource discovery ............................................................................................................. 62
    7.3.3 Contacting library staff ........................................................................................................ 72
    7.3.4 Advice to students .............................................................................................................. 73
    7.3.5 Opinions of existing interfaces ............................................................................................ 74
    7.3.6 Metadata ............................................................................................................................. 78
    7.3.7 Desires and priorities ........................................................................................................... 81
    7.3.8 Deficiencies of digital surrogates ........................................................................................ 86
    7.3.9 Availability of digital surrogates ........................................................................................ 88
    7.3.10 Citation of digital surrogates ............................................................................................ 89
    7.3.11 Downloading ....................................................................................................................... 91
    7.3.12 Social media ........................................................................................................................ 93
  7.4 Appendix D: List of Transcription Codes ..................................................................................... 93
  7.5 Appendix E: Participant Information Sheet ................................................................................ 95
  7.6 Appendix F: Project Proposal ...................................................................................................... 98

8 Works Cited ......................................................................................................................................... 99
1 INTRODUCTION

In 2010, the Vatican Library\(^1\) announced a plan to digitize its entire collection of 80,000 manuscripts by 2020. Among special collections digitization plans, this one is uniquely ambitious in scope, but in substance it is remarkably unremarkable. The last decade has seen a pronounced escalation of rare book and manuscript digitization programs at the world’s leading libraries, from the Bodleian Libraries’ 500,000-page Polonsky Foundation Digitization Project\(^2\) to the British Library’s Hebrew Manuscripts Digitisation Project\(^3\), which aims to digitize the majority of the Library’s 3000 Hebrew manuscripts by 2019. These recent projects tend to be both ambitious and republican in scope; rather than creating a virtual exhibition of prized manuscripts (along the lines of the Bodleian’s aging Early Manuscripts at Oxford project\(^4\)), they aim to faithfully represent an entire collection.

Only in the last few years have library special collections digitization projects achieved any kind of uniformity of appearance, process or structure. The high performance and lossless compression of JPEG2000, the widespread adoption of metadata standards such as METS and TEI, and the development of content and workflow management systems such as Goobi, ContentDM and iNQUIRE have allowed a kind of consensus, obviating the need for every library to forge its own path. Since 2012, that consensus has been strengthened by the software and standards of the International Image Interoperability Framework (IIIF), which make it possible to view digitized items from a variety of institutions in a single interface, and which may soon allow for cross-repository searching.

Despite all this progress, however, digital libraries still vary widely. Collections of digitized manuscripts and rare books may or may not be integrated into the main library electronic catalogue. They may contain a single institution’s holdings or those of a network of national libraries. Smaller institutions—or large institutions looking to attract a popular audience—may use social image-sharing software such as Flickr rather than a traditional library content delivery system. Even within a collection, image and metadata standards and discoverability may vary, as items may have been digitized at different times and according to different standards.

\(^1\) http://digi.vatlib.it  
\(^2\) http://bav.bodleian.ox.ac.uk  
\(^3\) https://www.bl.uk/projects/hebrew-manuscripts-digitisation-project  
\(^4\) http://image.ox.ac.uk
Among the higher tier of large, wealthy libraries, Digital.Bodleian, the platform launched by the Bodleian Libraries in 2015, is a representative—but by no means an exemplary—instance, demonstrating many of the problems particular to digitized special collections. Items in Digital.Bodleian cannot be accessed directly from SOLO, the Libraries’ online catalogue, but they can be accessed from some digitization project websites. Digital.Bodleian is searchable by keyword and offers a few refinable search facets (language, genre), but the search algorithm is intolerant and the search facets are not comprehensive (for example, several hundred Hebrew manuscripts are not currently listed under the “Hebrew” language facet). Metadata is displayed alongside each digitized item, but in many cases this metadata has been transferred from an older system and shows keying and transformation errors. All images in Digital.Bodleian are presented as zoomable JPEG2000s, although the zoom function is limited for older, lower-resolution images. Metadata and lower-resolution JPEGs are available for download. The Digital.Bodleian interface links directly to IIIF manifests and to an instance of the IIIF Universal Viewer.

Given such an imperfect system, how might a researcher be expected to discover and use a digitized book or manuscript? Do the necessary steps map onto users’ existing resource discovery skills, or require a new skill set? If new skills are required, how easy are they to learn? Are resources available to help teach them? Is this a usable system, and if not, what would a usable system look like? This research aims to provide a preliminary answer to these questions in order to steer future development and enhancement of digital libraries.

2 AIMS, OBJECTIVES, SCOPE AND DEFINITION

2.1 AIMS AND OBJECTIVES

Through an online questionnaire and a series of one-on-one interviews with users of digitized special collections, this study aimed to create a picture of how researchers learn about the existence of digitized items, how and why they access them, how they would like to be able to access them, and what interface features and functions they need or want. Key research questions included:

- What channels do researchers use to learn about digitized resources (e.g., social media, web forums, library websites, word of mouth)?
- What channels would researchers like to be able to use?

5 http://digital.bodleian.ox.ac.uk
6 http://solo.bodleian.ox.ac.uk
• What metadata and descriptive information is important for resource discovery?
• What research purposes do researchers consider digitized special collections resources to be useful for?
• How much time do researchers spend viewing digitized resources versus the analogue originals?
• What elements of viewing interfaces are most important?

The objectives of this research were to give a broad overview of global digitized special collections discovery preferences, as well as an in-depth study of the preferences and methods of a small but diverse group of scholars; and to develop a list of recommendations to enable digital library curators to design resources that suit their academic users’ needs.

2.2 Scope and Definition

For the purpose of this study, “digitized special collections” are defined as online collections of high-quality images of rare books and manuscripts, provided by libraries and potentially also by museums. All manuscripts qualify as special collections, as they are by nature unique to their holding institution; early and rare printed books, maps and ephemera also qualify. For the purpose of this study, books digitized en masse and in multiple witnesses do not qualify as special collections.

Digitized special collections may take a variety of forms, from searchable databases to Flickr collections. Example collections include Digital.Bodleian, e-codices7, Parker on the Web8, Gallica9, and Qatar Digital Library10. These collections may be maintained by institutions across the globe, as geography is rarely a barrier to use. While this project was primarily focused on high-quality photograph digitization rather than on microfilm or slide scans, large-scale microfilm imaging and transcription projects such as EEBO-TCP11 were included in the scope of the research insofar as they offer digital visual representations of special collections objects. The transcription and text-searching aspects of EEBO-TCP and similar resources were not considered, however, as this project focused specifically on scholarly purposes for which visual fidelity is key and a plain-text transcription or black-and-white low-resolution scan would not be sufficient.

7 http://www.e-codices.unifr.ch/
8 https://parker.stanford.edu/
9 http://gallica.bnf.fr/
10 www.qdl.qa/en
11 http://www.textcreationpartnership.org/tcp-eebo/
This project defines “researcher” loosely, and the terms “researcher”, “scholar” and “academic” are used more or less interchangeably. The questionnaire part of the study was pitched toward all users of digitized special collections, although the responses of self-described academic researchers were analyzed in more detail than those of self-described casual users. The interview part of the study focused more narrowly on researchers at the University of Oxford, whether professors, librarians, postdoctoral researchers, or postgraduate students. While undergraduate and taught graduate students were not intentionally excluded from the study, no suitable candidates for interview were found from these demographics. While every interviewed researcher had some affiliation with the University of Oxford, many hailed from other institutions and countries.

Researchers who do not use digitized special collections at all were outside of the scope of this study. While researchers outside the humanities have not been intentionally excluded from study, the preponderance of humanists among users of special collections resources meant that every participant in the study was currently working as a professional or amateur humanist.

3 LITERATURE REVIEW

3.1 INFORMATION SEEKING BEHAVIOR OF HUMANISTS

Starting with Stone (1982), research into the resource discovery behavior of humanists has repeatedly demonstrated the importance of browsing and serendipitous discovery (Bates 2002, Gueguen 2010, Knop 2015, Anderson 2004, Barrett 2005, Delgadillo and Lynch 1999, Duff and Johnson 2002). Humanists often start by researching a broad area and a “large and diverse set of resources” (Gueguen 2010, p. 97), which allows them to make unexpected connections across materials. Bates (2002) associated browsing with a need to “find information that we do not know we need to know”, a common feature of the humanities’ relatively open-ended research questions. While browsing may seem haphazard and disorganized, Delgadillo and Lynch, surveying graduate historians in 1999, found that their subjects reported being encouraged or assigned by their professors to browse the library stacks, indicating that browsing is not the result of disorganization or poor resource discovery skills, but rather a consciously cultivated technique. Duff and Johnson (2002), meanwhile, theorized that the browsing behavior of historians in archives in particular was a highly tactical endeavor aimed at adding meaningful context to a research subject.
Humanities scholars also use more focused resource discovery strategies. In particular, students look to advisors and other students for help, and more advanced researchers look to colleagues (Catalano 2013, Delgadillo and Lynch 1999, Duff and Johnson 2002). Another relatively low-cost strategy is monitoring—keeping up with mailing lists, scanning the tables of contents in new journal issues, and relying on colleagues to alert them to new research—to stay abreast of discoveries in a particular area once the initial work of browsing is done. By contrast, citation chaining or chasing—in which researchers follow the citations of particularly compelling references to build a network of interrelated and interdependent scholarship—is a useful but “time consuming and exhausting” strategy (Buchanan et al. 2005, p. 228). Sandstrom (1995, p. 441) distinguished between the “foraging” habits of lone scholars, who were more likely to use labor-intensive tools such as citation indexes and subject headings, and those of scholars in more well-connected research groups, who tended to use monitoring and socially mediated strategies and to avoid exhaustive searching.

3.2 RESOURCE DISCOVERY IN SPECIAL COLLECTIONS

The traditional model for humanities resource discovery—browsing, citation chaining, consulting with colleagues, monitoring from afar—describes the discovery of mostly secondary sources and widely-available primary sources; when it comes to archives and special collections, the model must be reworked. As Yakel (2004) pointed out, browsing is impossible in an archive, because the stacks are closed, and collections are most often organized by size or provenance rather than subject. Researchers must search or browse catalogues and finding aids instead, but these may be poorly integrated and are rarely comprehensive. Berner (1972) surveyed 50 manuscript repositories and found that most did not see the advantage of integrating catalogues and finding aids, as they considered the catalogue to be the primary resource and the finding aids to be “bonus features” for “more important” collections. Berner further argued that many manuscript catalogues sacrifice comprehensiveness for “the descriptive glorification of selected items” (p. 370), and recommended a comprehensive skeleton index that would allow researchers to refer to catalogue entries and repositories for further information.

While some special collections discovery tools have advanced since Berner conducted his research, many libraries’ catalogues are still incomplete and their finding aids difficult to use. Hamburger (2004) found that researchers located manuscripts mostly from footnotes in articles or books, and were unaware of digital discovery tools such as OCLC or ArchivesUSA. Roegiers (2003) advocated the reorganization of manuscript libraries and finding aids, recommending a top-down
hierarchy of holdings much like Berner’s skeleton index, and arguing that the researcher needs “something about everything” rather than “everything about something”. “Something about everything” may not be enough, however; Duff, Craig and Cherry (2004) surveyed historians on their use of archives and found that while finding aids were the most important resource discovery tool, researchers often perceived them as lacking in detail. Duff and Johnson (2002) found that historians who used archives often needed to consult archivists due to the complexity of finding aids; that finding aids often required multiple passes in order to find all the relevant material; and that historians consulted with archivists in order to find new or less high-profile material that was not included in finding aids. Duff and Johnson characterized these open-ended, informal consultations as a form of browsing, as researchers who didn’t know exactly what they were looking for solicited recommendations from archivists in order to approximate the serendipity of browsing open stacks.

In general, it appears that the more difficult or obscure a collection is perceived to be, the more willing humanities researchers are to ask library staff for help. Delgadillo and Lynch found that researchers were more likely to trust and seek help from subject specialists than reference librarians, and Wiberley and Jones (1989, 1991) found that researchers were very likely to seek help from archivists, as it was commonly perceived that archives are very difficult to navigate without staff help. Where resources are readily available, then, researchers may be likely to work independently and to engage in satisficing and lower-effort foraging behaviors (see Connaway, Dickey and Radford 2011 for more on these). Where finding aids are inadequate or difficult to use, researchers may be more likely to turn to experts for help.

3.3 DIGITAL COLLECTIONS

Analyzing the resource discovery methods of humanists among digital collections is a difficult task, as the available digital collections—and the most advantageous or popular methods of accessing them—are constantly changing. In a 1998 study of humanities researchers, Wiberley and Jones (2000) found that only two of the 13 studied used electronic surrogates of primary sources (one on CD-ROM and one online). More recent studies have found that users of digitized special collections prefer the digitization of entire physical collections, and want the digitized collections to closely match the traditional organization of the originals (Gueguen 2010, University of North Carolina at Chapel Hill 2009). Digitized primary sources and special collections are clearly becoming more popular and more ambitious, and despite early concerns about the reliability of digital collections (Duff, Craig and Cherry 2004, Ellis and Oldman 2005, Prom 2004, Kelley and Orr
Meyer et al. (2009) found that researchers were very willing to trust digital resources affiliated with trusted institutions such as the British Library.

In order to trust or welcome a digital resource, however, scholars must know that it exists. Matusiak (2012) found that the perception of libraries as physical repositories of books was harming the uptake of digital collections; students and teachers were unaware of the digital image collections they had been asked by Matusiak to evaluate, but once they were made aware of them they perceived them as useful and said that they would use them again. A similar phenomenon was noted by Archer et al. and Meyer et al. (both 2009) and Tibbo (2003). Buchanan et al. (2005, p. 222) found that users were not even “explicitly aware of the concept of digital libraries”, but referred to them as “online databases”. DeRidder and Matheny (2014) found a similar lack of awareness; the faculty members they interviewed tended not to distinguish between primary and secondary sources, citing the online journal database JSTOR, for example, as a digital primary source. Prom (2004), meanwhile, found that researchers who had been asked to evaluate online archive finding aids were unclear on what the finding aids’ function was; they expected links to digitized primary sources.

The lack of awareness of digital collections can partly be traced to libraries' minimal efforts to advertise and disseminate their digital resources. Libraries tend not to use search engine optimization techniques to bring users to their digital collections (Hudson 2012, Matusiak 2012), or even to choose memorable and search-friendly collection names (Meyer et al. 2009). This may be due to an ingrained “build it and they will come” philosophy (Harley 2007), or it may be because libraries are conscious of the risk of creating a resource that seems too commercial or insufficiently scholarly. Matusiak (2012) noted that the reluctance of libraries to promote their digital resources is compounded by the tendency of scholars to view libraries as brick-and-mortar institutions holding print resources. There is also the problem of evaluating and assessing digital resources, which may discourage scholars from using them; although humanists may not be as distrusting of digital resources as they once were, the lack of context and established indicators of reliability for digital resources means that researchers may place undue weight on a plainly descriptive name or a clear attachment to a reputable institution, thereby ignoring many useful resources (Warwick et al. 2007).

Beyond the initial problem of finding a digital collection, there is the additional difficulty of finding what one needs within that collection. Bates (1996, 2002) raised concerns about the applicability of traditional electronic resource discovery tools to humanities resources. Boolean
searching, for example, is not only alien to humanists (who are used to nuance rather than black-and-white); it is also ill-suited to humanities resources, which are often indexed by subject heading with only one heading per item, so that a Boolean search for several related concepts or headings will not turn up any results at all. The difficulty of locating individual resources in digital collections is a common thread in digital library research. Harley (2007) found that academics in California cited both the difficulty of finding and managing digital resources and a lack of time as discouraging factors—which is particularly disappointing as Barrett (2005) found that humanities graduate students appreciated the time-saving potential of digital resources; it seems that that potential has in at least some cases failed to deliver. Buchanan et al. (2005) found that while some users expressed satisfaction with digital resources, this was strongly correlated with an expressed willingness to learn different searching systems and experiment with search terms. This willingness is rare, according to Matusiak (2006), who found that users of digital libraries tended to use either browsing or keyword searching and to stick with whatever they first had success with. Matusiak also found that while older members of the community tended to prefer browsing digital collections, students preferred keyword searching, suggesting that they were reluctant to give up on their expectation of a Google-style search option.

In recent years, as search algorithms have improved and users accustomed to Google have grown less willing to learn complex search commands (Faiks et al. 2007, Nicholas et al. 2011, Connaway, Dickey and Radford 2011), digital library collections, like online public access catalogues (OPACs), have largely adopted keyword search interfaces. Recent research into digital libraries has tended to emphasize the technical side of things, evaluating users’ opinions about the collection interface rather than the effectiveness of the collection in facilitating research (and indeed, Warwick et al. (2007) found that users had very high standards for the interfaces of digital resources they had been asked to evaluate).

Inadequate metadata hampers discovery of digitized special collections much as it does their analogue counterparts, and digital collections often do little to facilitate the browsing behavior that is key to humanities research (Meyer et al. 2009). Getting from one resource to another within a digital collection is generally much more difficult than simply scanning a shelf, and unfocused searches are rarely rewarded. Rowlands et al. (2008) theorized that users who bounced quickly away from collection sites without visiting many pages might be engaging in a form of “power browsing”—another conjectured strategy for simulating the experience of scanning shelves—but this type of behavior is likely to be hampered by the fact that customary physical clues that help users to assess a collection (e.g., the length of shelves containing a particular category of resource)
are absent from digital resources (Warwick et al. 2007). The faculty members interviewed by DeRidder and Matheny (2014) largely found digital library resources to be unintuitive and confusing; more than half got lost at least once during usability testing, and they were unfamiliar with resource-specific searching rules but were also generally unwilling to ask for assistance—suggesting that the willingness Wiberley found among users of archives and special collections does not translate to digital resources. DeRidder and Matheny’s research suggests that digital library resources would benefit from scholarly input both in crafting scope indicators (such as visualizations and lists of new additions) and in structuring metadata, as the importance of certain metadata fields varies from one discipline to another. The focus groups studied by the Southern Historical Collection digitization project at the University of North Carolina similarly indicated the importance of mimicking familiar special collections search and browse options. An interface that works smoothly and intuitively by a UX developer’s standards, as rare as it might be within the world of digital libraries, will not satisfy humanities researchers if it has not been developed with expert scholarly input. As Dussert (2010, p. 219) concluded, librarians who talk to staff and students about digital resources “may find out that what [they] thought was wanted is not what is needed”.

Moreover, given the importance of browsing and monitoring in humanities resource discovery, and the customary reliance of special collections users on library staff to help them locate resources, it is clear that even the most finely engineered metadata and the smoothest interface are not enough to support users of digitized special collections. Anderson (2004, p. 115) emphasized the importance of developing features that emulate “the context and mediation provided by leads in print and informal methods”, suggesting an Amazon-like review and rating system. As of writing, no library seems to have implemented such a system, but it may be that humanists are turning to external tools such as Twitter, Delicious and Mendeley to approximate the serendipity of browsing. These tools incorporate a social aspect that mirrors the importance of input from other academics in resource discovery. Furthermore, Quan-Haase et al. (2015, p. 7) found that “serendipitous, unexpected experiences” and “making random and important connections” were key reasons for Twitter use among digital humanists. These connections are most likely between people, however, rather than between resources; study participants used Twitter to communicate with colleagues, follow conferences and keep up with new projects, but not, apparently, to share links to digitized resources. Twitter may be usable for resource discovery—and specifically for browsing, rather than for crowdsourcing access to known resources—but the practice is not yet mainstream.
Finally, there is the nature of the digitized special collections resource itself. Is the end object sufficiently useful to researchers to justify the trouble of finding it? Kachaluba, Brady and Critten (2014) found that surveyed humanists used electronic journals and reference materials but preferred hard copies of primary sources, although they appreciated the searchability and portability of Early English Books Online\(^\text{12}\) (EEBO) (a large and high-profile subscription database of scanned digital reproductions of microfilm copies of early printed books). Rose (2002) and Ellis and Oldman (2005) found that researchers in the fields of art history and English literature respectively were willing to use digitized resources for the early stages of their work, but relied on the originals for advanced research. Searchable text is an important advantage of digitization, but while improvements in optical character recognition software are making it easier for libraries to provide full-text transcriptions of their digitized resources, these transcriptions are usually confined to 18\(^{\text{th}}\)- and 19\(^{\text{th}}\)-century resources that are easy to OCR, or to a particular high-profile category of resources (for example, the National Library of Wales’s Welsh Newspapers Online\(^\text{13}\), or the Archaeology of Reading\(^\text{14}\) project’s selection of Gabriel Harvey’s annotated books). Much like Berner’s lopsided manuscript catalogues, this unevenness of quality hampers resource discovery and discourages study of the lower-profile resources that Duff and Johnson’s historians relied upon archivists to alert them to.

4 \textbf{METHODS}

In order to form a picture of global use of digitized special collections across a variety of backgrounds, while also obtaining a nuanced understanding of the methods and needs of individual scholars, a mixed-methods approach was taken. Data gathering consisted of two elements: an anonymous web questionnaire and a series of in-person interviews. These elements were combined for two reasons. First, in-person interviews would have to be undertaken in or near Oxford, meaning that any abnormalities or biases in Oxford scholars would be reflected in the data. Second, it was known from conducting library user surveys in the past that it is difficult to get professors to fill them in. Combining interviews and a questionnaire made it possible to balance the Oxford-focused interviews with a more global perspective (even if only at the superficial level of an

\(^{12}\) http://eebo.chadwyck.com/
\(^{13}\) http://newspapers.library.wales/
\(^{14}\) http://archaeologyofreading.org/
online questionnaire), while ensuring a decent number of high-quality responses from faculty members. The timing of these two elements was slightly staggered, with the intention that early answers to the questionnaire would guide the design of the interviews.

4.1 QUESTIONNAIRE

4.1.1 Design

The questionnaire was created using Google Forms, a free online tool. Google Forms was chosen because of previous favorable experience, and because, compared to the free options of other online questionnaire tools, it offered the highest degree of direct access to questionnaire answers. Google Forms offers two main ways of viewing and analyzing results: automatically generated charts, which can be viewed directly in the web browser, and downloadable CSVs. In each case, the results are anonymous. Responses are identified by timestamp only, rather than by anonymized IP address, making it impossible even to tell if someone has taken the questionnaire more than once.

The design and administration of the questionnaire were modeled partly on questionnaires described in the secondary literature (Dobreva et al. 2012, Knop 2015). The questionnaire (Appendix A) consisted of 19 questions, plus a space for any additional comments. The questions were a mix of multiple-choice, yes-or-no, and short-answer questions. All multiple-choice questions included an “Other” option among the possible answers. In order to make data analysis as efficient as possible, and to keep the questionnaire from seeming too laborious to complete, the number of free-text questions was kept to a minimum. Responses were monitored at regular intervals, with the idea that tweaks could be made to the questionnaire if responders appeared to be finding particular questions difficult or if additional options needed to be provided for any of the multiple-choice questions. In the event, however, only one question was tweaked: two institutions were added to the possible answers to the question, “Do you use any of the following for access to digitized rare books or manuscripts?” after several respondents had listed them under the “Other” option.

4.1.2 Dissemination

After receiving permission from the heads of the department to do so, the questionnaire was advertised via the Bodleian Digital Library Systems and Services Twitter account on two different occasions, as well as via the researcher’s personal Twitter. The questionnaire was also tweeted from the Bodleian Department of Rare Books Twitter account, linked to from the Oxford
Research Centre in the Humanities website, and advertised on several humanities mailing lists at the University of Oxford after interviewees offered to set this up. Links to the questionnaire were also included in every email sent out to potential interviewees, with the idea that recipients would share the link with other researchers, and complete the questionnaire themselves if they did not have time for an interview. Links to the questionnaire were also emailed to a list of 20 scholars around the world who had previously contacted the researcher for digitized images during the Bodleian Libraries’ hardware failure several months prior.

The tweet from Rare Books appeared to generate the highest number of responses, with the possible consequence that users of rare books may have been slightly overrepresented in the questionnaire responses. However, the proportion of self-identified rare book vs. manuscript users in the responses was more or less even, indicating that this was not the case.

4.1.3 Analysis

Once the survey period was over, the results were analyzed for spam, incorrectly-filled-out responses, and other problems. It appeared that the vast majority of respondents had filled out the entire questionnaire, and no spam responses were noted. However, in one case, identical responses had been submitted 23 times, with the timestamps showing a gap of about 30 seconds in between responses. It is unclear how this happened, whether due to a glitch in the software, an elaborate hack, or simply a person or several people submitting the same set of responses many times in quick succession. In any case, it seemed likely that these responses represented only one person’s actual experience of using digitized materials. Consequently, and because the 23 responses represented almost a fifth of the total responses to the questionnaire, it was decided to remove all but one of these responses from the final results, in order to keep them from skewing the data.

4.2 Interviews

Parallel to the questionnaire, a series of interviews were conducted with University of Oxford scholars. The target number was between 10 and 20 interviews; the exact number was not decided in advance, as difficulties were anticipated recruiting willing interviewees, especially as the interview period coincided with the beginning of the academic year, an especially busy time for scholars. Scheduling did prove to be difficult, but 13 interviews were conducted over the course of six weeks.

The recruitment and interviewing strategy was informed by the advice of a colleague, Christine Madsen, who had previously conducted research into scholarly use of digitized resources
The subsequent transcription and encoding of the interviews was also modeled on similar studies described in the secondary literature (Madsen 2010, Barrett 2005, Buchanan et al. 2005). The key points of the strategy were to contact potential interviewees personally by email; to arrange interviews in the interviewees’ offices where possible; to record all interviews; and to follow a semi-structured interview model, with a few standard questions customized to each participant, and follow-up questions determined by the participants’ responses.

4.2.1 Recruiting

In all but one case, initial contact with the interviewee was made via individual email, as this has proven more effective than mailing-list posts or social media promotion (Madsen 2010). The exception was a scholar whom the researcher met at a conference on digitized manuscripts and approached in person. Interviews were initially solicited from scholars with whom the researcher already had a working relationship and who were known to be working with digitized materials. If no scholars were known personally within a particular field, colleague subject librarians were asked to recommend some. In order to recruit further participants, the initial interviewees were also asked to recommend colleagues who might be willing to be interviewed. This method meant that every initial email to a potential interviewee included at least the mention of a mutual acquaintance, in order to reduce the feeling of being cold-called. However, it is possible that this method resulted in the biased selection of a pool of participants, artificially overrepresenting those who are already known to librarians and other scholars and underrepresenting more solitary, independent or nontraditional researchers. Consequently, while the selected participants largely did not report having much contact with librarians or other researchers in their search for digitized materials (see 5.2.2.1 below), the average degree of contact across all researchers may be even lower.

In almost every case, the responses from scholars were enthusiastic and helpful. Only two scholars declined to be interviewed due to being particularly busy, and these two still completed the questionnaire (see above) and recommended colleagues who might be more available.

An effort was made to recruit participants from across a number of academic disciplines within the humanities, and from a variety of stages in their academic careers. In the end, the set of participants comprised:

- two scholars of printed books; four scholars of Oriental manuscripts; seven scholars of western manuscripts
- paleographers, art historians, musicologists, textual analysts, linguists, bibliographers
- seven later-career professors; five early-career postdoctoral researchers or research fellows; one subject librarian

All participants were employed by the University of Oxford.

This sample is not intended as a proportional representation of Oxford scholars in the various areas of manuscript and rare book study, and some areas (for example, the study of East Asian manuscripts, and the study of post-1800 manuscripts) are not represented at all. However, the sample achieved the research goal of gathering in-depth responses from across a range of disciplines and backgrounds in order to flesh out the wider responses gathered from the online questionnaire.

4.2.2 Interview process

Where possible, interviews were conducted in the participants’ faculty or college offices, in order to minimize their inconvenience, place them at ease, and allow for spontaneous reference to digital or analogue materials during the interview. This approach was largely successful. Seven participants were interviewed in their faculty offices, four were interviewed in library cafes and two were interviewed in a library seminar room. Those interviewed in cafes tended to be the earlier-career academics, who did not have offices large enough to conduct interviews in. As expected, conducting interviews in offices proved fruitful; most of the participants interviewed in their office made reference on at least one occasion to materials they had to hand.

All interviews were recorded on the researcher’s phone, using the free app Recorder. During the first two interviews, typed notes were also taken, but this belt-and-braces approach was discarded for subsequent interviews, as the recordings were sufficiently high-quality to be fully transcribed, and the act of typing notes tended to distract from the interview process. All interviewees were asked if they would prefer not to be recorded, and had any of them said yes, typed notes would have been taken instead.

A basic set of questions guided each interview. These questions were:

- Tell me about your research.
- What role do digitized materials play in your research?
- How would you go about finding a particular digitized item?
- What online resources do you use for this, and what are your opinions of them?
- Do you ever use Google/social media/direct contact with colleagues/direct contact with librarians to find resources?
- Tell me about a time you struggled to find a digitized item.
- How do you think your discovery strategy compares to others' in your field?
- What metadata fields are important to you for the purposes of discovery and analysis? What additional fields would you like to see included?
- Is there anything you would like to be able to do with a digitized item that you can’t do?
- Do you like to be able to download images, and if so, what do you do with them once you've downloaded them?
- What is your opinion of the accuracy of digital metadata and digitized images?
- How would you go about citing an item you’d consulted digitally?

The phrasing of these questions depended on the background of the participant. For example, metadata was referred to as “catalogue descriptions” or “descriptive information” if it seemed that the participant might be uncomfortable with the term “metadata”. After the first few interviews, the question about comparison to colleagues was omitted, as the first interviewees were reluctant to make generalizations about others' behavior. Each interview took between 35 minutes and an hour, except for one interview, which was limited to half an hour due to constraints on the participant’s time.

4.2.3 Transcription and analysis

Constraints on the researcher’s time meant that it was not possible to write up notes about the interviews immediately after conducting them. Instead, each interview was played back at a later time (the same evening when possible), and a transcription was made of every response that a) was relevant to the research and b) did not betray the participant’s identity. Some references to specific projects were cut. These partial transcriptions, divided by topic rather than participant in order to preserve the anonymity of participants as much as possible, are attached as Appendix C.

Once the transcriptions were complete, QDA Miner Lite, a free qualitative data analysis tool, was used to lightly encode them in order to more easily track parallels across interviews. Codes were used to flag passages that answered particular questions, but also to track references to certain subjects—shelfmarks, for example, and Google—that had come up frequently across several interviews and that consequently seemed key to resource discovery. A code was also used to mark passages that indicated confidence or lack of confidence in the participant’s own resource discovery
skills. QDA Miner Lite did not make it possible to collate the tagged responses, so collation was done by hand, pasting passages marked with the same tag into a new document. The complete list of codes is attached as Appendix D.

4.2.4 Managing bias

The greatest danger in conducting these interviews, particularly as a member of Oxford library staff and with Oxford scholars, was that the participants’ responses would be inaccurate or tailored to please the interviewer. Before each interview began, the interviewee was provided with an account of the measures being taken to ensure confidentiality (see Appendix E), with the hope that this would set them at ease and allow them to speak more candidly. In addition, care was taken to phrase questions as disinterestedly and non-judgmentally as possible, so that interviewees were not made uncomfortable about their own practice regarding image downloads, citation or other somewhat contentious subjects. Only two or three interviewees, talking about these issues, self-deprecated or offered disclaimers, suggesting that this effort was at least somewhat effective.

In the early stages of project planning, it was anticipated that the interviews would include a component of testing and feedback on Digital.Bodleian. Once recruiting began, however, it was decided that, given the researcher’s role in creating and maintaining Digital.Bodleian, such a component would be unhelpful. It would confuse the distinction between externally affiliated researcher and Bodleian staff member; it would potentially put interviewees on edge, as they would most likely be reluctant to speak unfavorably about the pet project of the person they were speaking to; and it would tether responses too much to the specific inadequacies of a particular system, rather than providing a broader picture of user requirements and desires. Interviewees were not discouraged from talking about Digital.Bodleian if they wanted to, but the interview questions were designed to attempt to shift emphasis away from it.

5 ANALYSIS

5.1 QUESTIONNAIRE

After removing 22 duplicate responses (see 4.1.3 above), the total number of responses collected in six weeks was 93. These responses are attached as Appendix B. The substantial majority of respondents filled out the entire questionnaire. Even the questions that required long-
form text input were mostly completed; for example, the question about preferred digital resources received 68 responses.

5.1.1 Demographic breakdown

According to the respondents' self-identification, 41 were postgraduate students, 21 were alt-ac or independent researchers, eight each were lecturers or undergraduates, and six were professors. A further nine identified as “other” and specified roles including “retired museum director”, “research fellow”, “library curator”, and “HS student”. History was the most popular field of research for respondents, followed by literature and “other”. “Other” included oceanography, digital humanities, culinary arts, critical race studies, musicology and more, but most respondents who chose “other” did so in addition to one of the given options, the exceptions being art history and philology. In terms of material specialty, approximately equal numbers of participants worked with printed books and Western manuscripts (67 and 71 respectively), followed by archival materials (45), Eastern manuscripts (24) and maps (20). A majority of respondents (85) consulted digitized rare books or manuscripts several times a year or more, with 58 consulting them at least once a week. Similarly, a majority of respondents (68) rated digitized rare books and manuscripts as “extremely important” to their research, and 0 respondents rated them as “not important at all”. These results are unsurprising, as the questionnaire was marketed toward humanists who already use digitized books and manuscripts in their research. Because everyone who answered the questionnaire already did use them, it was unnecessary to filter out any responses.

Respondents were not asked about their geographic location. In retrospect this would have been interesting information to gather, but the number of responses was small enough that it would have been unlikely in any case to achieve a proportional or representative response from a number of locations.

5.1.2 Discovery

A substantial majority of respondents (71) reported learning about digitized items via a library website. The next most common sources were a colleague or advisor (49) and an online list or directory (46). Published research, social media and librarians were all also relatively popular, with fewer respondents reporting finding items via mailing lists or web forums and department websites (14 and 8 respectively). Among the write-in responses, 10 reported using Google for either a manual search or a web alert, one reported trusting to “luck”, and one reported having “digitized ’em myself”.

These responses contrasted with the responses to the next question: "How would you prefer to find out about digitized rare books and manuscripts?" Here, there was little consensus. 39 wanted to find out via an online list or directory, and 22 wanted to find out via a library website. Fewer than 10 each wanted to find out via social media, a mailing list or web forum, a librarian, department website, or a monograph or article, and—most strikingly—only one wanted to find out via a colleague or advisor. Thus, the second-most-common discovery method—colleague or advisor reference—is the very least preferable method. Several write-in responses, however, indicated that the method doesn’t matter as long as it works: “I don’t mind so long as I can access it” said one, and “any means by which I can re-find it easily” said another.

In explaining their answers to these questions, the respondents revealed their priorities in digitized special collections discovery. Of the respondents who preferred to discover via an online directory, most mentioned convenience or ease of access and the value of “immediately” seeing whether a given item has been digitized. One said, “It means I am free to look for things when the library is closed and I don’t have to bother busy staff.” Others noted that searching a central directory would be preferable to checking library websites repeatedly for updates or to checking multiple library websites each time. For example, as one respondent elaborated:

A concentrated single catalogue-- an online version of Ker, for instance--with links to existing digitized manuscripts would mean that links could be added to digitized manuscripts as they became available, and that rather than a librarian or researcher constantly checking all the different library sites (BL, Parker, Bodleian, etc.) to see what has recently been made available, or relying on social media to hear about progress in new manuscript digitization programmes, there would be a single directory where librarians could point researchers which would always be up to date.

Several others also commented on the advantage that a centralized index could be easily kept up to date. Another respondent reported building a central index themselves “because one did not exist”. Separating the directory from any particular academic institution also appears to be valuable. One respondent wrote that a “universal forum” would “objectify” the process of resource discovery. Another said that “I’m an independent researcher with no ties to academia, being able to access whenever and wherever is very valuable”, and another reported that academic libraries with subscriptions to digitized materials “are wholly unhelpful to non-student folks from the local community. I find their snootiness especially sharp toward black and brown patrons like myself.”

These responses from independent researchers are particularly important to consider, as they highlight the frequent failure of libraries to adequately consider the needs of users outside a specific academic community.
Of those who indicated a preference for library websites—the second most common preference—several explained that because they are already used to visiting library websites to find resources, they are a “natural”, “familiar” and “efficient” location (“It’s the first place I would think to look”). One respondent noted that library websites would be likely to have dependable metadata; “I’d hope for good provenance and other standards from a library.” Two respondents who chose “library website” did so with the caveat that a central repository or directory would also be necessary in order to provide up-to-date information efficiently.

Those who preferred social media mentioned accessibility, currency, and the fact that (according to one undergraduate responder) “I’m more tuned in social media nowadays than with my librarian”. Those who preferred a mailing list or web forum mentioned currency (“It would be interesting to have a system that shows you which manuscripts have recently been digitised”) and simplicity (“Comes to me so I don’t have to seek it out, can ideally specialize info for my needs”). Those who preferred librarians noted the advantage of human interaction (“I deem the direct communication with the library holding the manuscripts as of great importance” and stressed the importance of subject librarians, “as they can not only locate sources but direct you to others”, although one gave the caveat that librarians may not have time to help. Those who preferred department websites noted convenience and specialization. One who preferred monographs or articles said “It would help with seeing how other people are using the resources”.

Several respondents said that the mode of discovery didn’t matter as long as it was successful. One wrote in that their preferred method was “Googling shelfmarks”, as “This is the most efficient way to get directly to the manuscripts. The diversity of different library websites' designs makes it extremely tedious to try to negotiate, especially since so few of them are organized by shelfmark, which is the ONLY way that scholars think about manuscripts”. Finally, one respondent, a library curator, indicated no preference and said, “I don’t think it’s possible for there to be one single place to find out about digitized manuscripts - I’m rather happy having a happy constellation”.

Reasons given for preferred discovery method were divided into several categories to form Figure 1.
Figure 1: Reasons given for preferred discovery methods

<table>
<thead>
<tr>
<th></th>
<th>accessibility</th>
<th>comprehensiveness</th>
<th>convenience</th>
<th>currency</th>
<th>human interaction</th>
<th>quality</th>
<th>specialization</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>colleague or advisor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>department website</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>librarian</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>library website</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>mailing list or web forum</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>monograph or article</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>online list or directory</td>
<td>10</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>social media</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>total</td>
<td>14</td>
<td>8</td>
<td>22</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

The distribution of responses indicates that convenience and accessibility are the most important factors in choosing a discovery method, followed by currency and comprehensiveness. Quality, specialization and human interaction are comparatively unimportant, suggesting that scholars are willing to sift through irrelevant or poor-quality results as long as they can do so efficiently.

There was no obvious correlation between career stage or discipline and discovery preference.
5.1.3 Opinions on existing resources

Respondents were provided with a list of existing digitized special collections resources and asked to select each one they had used. The list was made up of resources encountered either during work at Bodleian Digital Library Systems and Services or during the literature review, and subsequently edited to include additional resources named by respondents. The British Library, Google Books, the Bodleian Library and the Bibliothèque nationale de France were the most popular resources (see fig. 2).

In addition to the provided options, there were 15 write-in responses. Among these, the most popular was the Internet Archive, which was listed by four respondents.

Respondents were also asked to name a few resources they especially preferred and to elaborate on why. The intention here was to discover what resources scholars would name unprompted, and to learn what they valued. The resources most frequently named were Gallica/the
BnF (20 mentions), the British Library (15), the Bodleian (12), the Internet Archive (9), e-codices (6), and Pinakes (5). Five respondents mentioned social media (Twitter, Tumblr, or social media in general) rather than, or in addition to, institutional websites or public directories.

In explaining their preferences, respondents echoed the findings in 5.1.2 above. Comprehensive lists, rich metadata, searchability, high-quality images and a usable viewing interface emerged as the top priorities. For example:

- “E-codices: All...Swiss mss in one place; great search facilities, viewer and metadata”
- “Bibliotheque nationale de France is the best re: organization and ease of access; Munich is a nightmare”
- “once I know what’s there, [Digital.Bodleian] is invaluable for access (however there should be a special forum for listing and updating what’s there)”
- “All four give excellent manuscript images, are easy to use and give also further explaining texts. Especially laudable is that the Heidelberg Library permits scholars to use their digitized images also in scholarly publications without further fees.”
- “E-codices: straightforward catalogue, allows download of high-resolution images, provides DOIs for citing manuscripts in a stable manner”
- “BSB particularly useful because of (a.) volume of materials available, and (b.) option to either flip through page by page or download the entire text and read on computer. Option (b.) is particularly important to me because most of my working life is in Egypt, and so slow download speeds mean that flipping through page by page takes too long to be practical, whereas I can start something downloading and come back to it later (or download during periods when I am traveling and have access to faster broadband speeds), then work through more carefully at leisure on my computer.”
- “The British Library is particularly good because the digitisation is integrated with the normal search catalogue, so it means that you find out about the digitisation when you actually want it.”
- “e-codices: Shelfmarks!”
- “all are high quality images with few to no restrictions always working to improve platforms”
- “Collections without easy catalogue access are not much good for anything serious, but mostly it’s way too hard to find a) digitisation b) that actually works and isn’t broken all the
time c) with a decent catalogue entry and not just metadata. Also "public comments" are a waste of space. People should stop doing that.”

- “Google Books--- only because their search functions are outstanding.”
- “Google—simply for being so huge”
- “Twitter—although any social media will do, for people plugging interesting resources, or new ones, or trial periods, or whatever”

Only one respondent, an independent researcher working in musicology, mentioned IIIF compliance directly as an advantage. No respondents mentioned IIIF compliance indirectly (for example, by mentioning the ability to compare images from multiple institutions side-by-side).

5.1.4 Research functions of digitized special collections

Respondents were asked what they use digitized rare books and manuscripts for. “Situations where it is not possible to consult the original object” was the most popular option (see fig. 3), suggesting that digitized items still primarily function as straightforward surrogates for the originals. Meanwhile, comparatively few respondents (8) reported using digitized items for computational analysis, suggesting either that digital humanists tend not to work with rare items, or that the digital surrogates are inadequate for their purposes.

Figure 3: Reported research uses of digital surrogates
Respondents were also asked if there was anything they would like to be able to do with digitized materials that they are not currently able to. These responses can be roughly divided into the following categories of desirable functions:

- full-text search: 6
- downloading: 2
- higher-quality images (for example, to detect ruling): 3
- more metadata (for example, physical dimensions, or to improve discoverability): 4
- better viewing software: 1
- transcription and annotation functions: 1
- more complete digitization (for example, more non-English items): 4
- images of pages in context: 2
- better reuse options: 2

Two respondents mentioned dimensions, and one mentioned specifically the shortcomings of microfilm digitization. One mentioned that textural analysis would be desirable but impossible. Eight respondents, however, said that there was nothing they would like to be able to do that they can’t currently do, and 57 respondents did not answer this question at all, suggesting that the general level of satisfaction is reasonably high.

5.1.5 Metadata

Respondents were asked what terms they would like to able to search by in order to discover digitized special collections items.
Responses to this question were relatively evenly distributed (see fig. 4). Title and author were the most popular fields, but even the least popular (licensing/reuse information) was deemed desirable by 22 respondents. Responses also depended heavily on area of study. All but three of the 74 respondents who study printed books selected title as a desirable field, compared to only four of the 18 who study only manuscripts. All of those who study manuscripts, however, selected shelfmark as a desirable field.

5.1.6 Important features of digitized items

Respondents were also asked to select the most important features of digitized rare books and manuscripts. Most respondents selected more than one feature, but the most important features—selected by more than 50% of respondents—were zoomable images, high-quality image downloads, downloadable PDFs and searchable text. Searchable metadata was deemed important by only 40 respondents, suggesting either that respondents have other ways of discovering digitized objects than via the associated metadata, or that they did not understand the term “metadata”. One respondent chose “other” and wrote in “Catalogue and bibliography - the scholarly contextualisation, if you like,” suggesting that the latter possibility might be correct.

In this question, several features were listed as options even though they are relatively new and, to the researcher’s knowledge, little used. These were side-by-side comparison, annotation and image adjustment. As expected, these features were among the least popular, but they did garner
between 15 and 29 responses each, suggesting that some respondents either are using these features or simply consider them desirable.

**Figure 5: Advantages of digital surrogates**

- zooming in to a high level of detail
- high-quality image downloads
- downloadable PDFs
- searchable text
- portability and remote access
- searchable metadata
- the ability to adjust image qualities such as contrast and...
- side-by-side comparison of multiple items
- preservation
- the ability to make notes and annotations
- other

5.1.7 Downloading

Respondents were also asked whether they preferred to download digitized items or view them online. Unsurprisingly, given the responses to the previous question, downloading proved more popular; 64 respondents reported that they prefer to download, while 29 reported that they prefer to view online.

5.1.8 Research planning

In an effort to gauge the perceived importance of digitized items in research, respondents were asked whether they had ever built elements of their research around whether particular items had been digitized, or if they knew of others who did this. 57 of the 89 respondents to this question said yes. In order to gauge the effectiveness of respondents’ methods for finding digitized items, they were also asked whether they had ever found out only late in the research process, or after the research was over, that a relevant item had been digitized. Here, 55 of the 92 respondents said yes. While this is a large number, which may indicate that libraries need to do a much better job disseminating digitized resources, it is also possible that it simply reflects the pace of digitization, and that the relevant item had not been digitized at all until late in the respondent’s project.
5.1.9 Preferred method for finding new digitized items

When Digital.Bodleian was launched in 2015, several scholars on Twitter lamented the lack of an easy way of discovering newly digitized items. In order to learn more about options for this, respondents were asked if they would subscribe to or make use of a number of methods for announcing newly digitized items (see fig. 6).

![Figure 6: Preferred methods of announcing newly digitized items](image)

The most popular option, chosen by 51 of the 87 respondents, was an email newsletter. This was unexpected, given that research suggests that students at least are easily fatigued by email or SMS communications from libraries (Walsh 2012), but it may be that the audience of the questionnaire consists of more committed researchers who are more likely to appreciate news about digitization. Twitter, a library blog, and an option to browse new items in the digital library were also reasonably popular responses to this question, with each chosen by 40 respondents. “RSS feed” was chosen by only 15 respondents.

5.1.10 Citation

Citation and resource attribution are subjects of considerable concern for the managers of digital library collections, as demonstrating use of the collections and their impact on research is key for acquiring and justifying funding. In order to gauge the current state of play regarding citation of digitized special collections, respondents were asked whether, supposing they had used a digitized rare book or manuscript, they would cite the original, the digitized version, or both. 41 respondents—just under half the total—said they would cite the original only. Some of these said this was because it was unimportant whether they also cited the digitized version, or because they
weren’t sure how to cite the digitized version; but most (21) said that it was because it was customary in their field. Fourteen respondents said they would cite the digitized version, and 28 said they would cite both. Although there may be nuances to these answers that were lost due to the multiple-choice question format, they suggest both that the picture is not as bad as librarians might fear (just under half the respondents said they would cite the digitized version one way or another) and that the situation will improve with time, as citation practices slowly adapt to accommodate digitization.

5.1.11 Questionnaire conclusions

The responses to the questionnaire demonstrated that digitized special collections are important to many researchers; that respondents, by and large, had no trouble distinguishing between digitized special collections and other types of digital resource (only two respondents made reference to JSTOR and other secondary resources); and that researchers take a wide range of approaches to finding and using digitized special collections. The responses also indicate a high level of engagement with digitized collections: a willingness to subscribe to email newsletters, to find resources through various means, and to think critically of how the resources available to them could be improved. While it is likely that respondents self-selected as particularly enthusiastic users of these resources, it is nevertheless encouraging to librarians looking for proof that the resources they provide are well used. Meanwhile, the importance placed by independent researcher respondents on general public indexes serves as a reminder that library resources that adequately serve a local set of professional academics may not answer at all to the needs of a wider community of researchers.

Several questions were raised by the questionnaire responses and flagged for more in-depth investigation in the interviews. These include why researchers prefer to download images; whether standard citation practices have been changing to accommodate digitized materials; why colleagues are such a common source of resource discovery help, but not a preferred source; and, in detail, what metadata researchers would like to search for.

5.2 INTERVIEWS

5.2.1 Role of digitization in research

While the 13 interviewees were chosen because of their use of digitized special collections materials, the specifics of this use varied. Some were working on projects involving a handful of digitized objects, which in several cases had been digitized specifically for their research. Others
were working across an entire corpus, building databases or catalogues of, for example, Arabic manuscripts from a particular period. Consequently, the needs of the scholars varied. A scholar working on a specific manuscript witness might be interested mainly in image quality, while one gathering information about different manuscripts might be more interested in searchable metadata and an effective discovery system.

Regardless of their specific requirements, the interviewees held in common a dependence on digitized materials. A scholar working on a digital edition from a handful of digitized manuscript witnesses estimated that he consulted the digital surrogates more than 99% of the time, returning to the originals only to inspect the binding or view the illuminations in ultraviolet light. Other interviewees remarked on the amount of time and money digitization had saved them in eliminating visits to consult materials that did not end up being relevant to their research, and one said that digitization allowed him to maximize the value of his time with the actual manuscript, as he would first be able to conduct a “preliminary analysis” of the digital surrogate.

The interviewees who were least enthusiastic about the role of digitization in their research were those working in poorly catalogued areas. A scholar of book history, for example, reported that in order to find out about copy-specific details of a library’s collection (e.g., inscriptions and marginalia), she often had to visit the library in person and consult a folder of loose papers, as this information had not been electronically catalogued. A scholar of Arabic manuscripts, despite using digital surrogates and electronic catalogues extensively, was less enthusiastic than his counterparts in Western manuscripts, due to the amount of time he spends searching for poorly catalogued materials.

Interviewees also reported using digitized materials in teaching. This might be to liven up or illustrate a lecture, or to “give a sense of the material document”—something that, particularly in manuscript studies, undergraduate and taught postgraduate students have little experience of, as access to special collections is often restricted to research students.

5.2.1.1 Availability of digitized materials

Whether or not the desired materials are available in digitized form appears to depend considerably on field of study. Thanks to EEBO and Google Books, the interviewees working with printed books reported good availability of digitized resources and especially of British printed books, despite criticisms of the search tools, viewing interfaces and image quality. (EEBO’s images are scans of microfilm, and Google Books’ are black-and-white scans.) In the field of manuscript
studies, scholars appear to be aware of the historical tendency of library digitization programs to prioritize famous and beautiful objects that will generate broad publicity and engagement, rather than obscure items that will be useful to a handful of scholars. Consequently, a Hebraist art historian and a medieval musicologist reported that the manuscripts she studied had mostly been digitized; the musicologist noted that “the ones that haven’t been digitized are often the scruffy, not-very-nice-looking sources”. On the other end of the spectrum, a German medievalist working with letters held at monasteries reported that she had taken over 21,000 photographs herself, and that three of the four institutions that had digitized items she was interested in had done so on her request and with her funding. As well as paying for digitization themselves, several interviewees reported resorting to microfilm when digital images weren’t available, or writing directly to the library to ask for images, with variable success.

5.2.1.2 Inadequacies of digital surrogates

Some interviewees said that digital surrogates were completely adequate for their purposes, obviating the need to visit the originals in person. Most, however, had reservations. Some of these could be resolved by higher-quality digitization: doubts about color fidelity, for example, or problems with image resolution. Other concerns related to aspects of books and manuscripts that are seldom adequately photographed: bindings, the edges of text blocks (in order to see quiring and decoration), and page gutters. Several interviewees expressed a belief that there are aspects of physical books and manuscripts that simply cannot be captured digitally. One, who reported that his consultations with manuscripts are “without exaggeration, 99 percent digital”, nevertheless said that even a 3D virtual reconstruction of a manuscript would be inadequate because “there are some things sometimes that you just need to see or feel”. The quality of the parchment or paper, the depth of the paint or ink, and the feel of the binding were all mentioned by interviewees as belonging to this category, and while one might expect that high-quality digital images would make it easier to see textual details or erased text, one interviewee recounted several occasions where she noticed something when consulting the original that she had missed in the digital surrogate because she hadn’t expected to see it there. Nevertheless, every interviewee was convinced that digital surrogates, even if they didn’t always obviate the need to consult the original, sped up the consultation process by allowing them to focus only on the things they hadn’t been able to ascertain from looking at the surrogate.
5.2.2 Discovery methods

5.2.2.1 Where they search

Preferred resources for discovering digitized special collections vary across areas of research. The scholars working with Greek manuscripts, German manuscripts and printed books reported trying subject-specific centralized databases first: Pinakes\(^{15}\) in the case of Greek manuscripts, the Handschriftencensus\(^{16}\) in the case of German manuscripts, and, in the case of printed books, the subscription databases EEBO and Eighteenth Century Collections Online\(^{17}\) (ECCO). These sites may not contain rich metadata in every case, but they are relatively easy to search and navigate, and provide up-to-date links to digitized copies. One scholar of Greek manuscripts reported that “it’s easier...to find the permalink of Greek manuscripts from the Bibliothèque nationale de France through Pinakes rather than by searching the BnF’s website due to better stemming in Pinakes.

Aggregated finding aids like Pinakes and the Handschriftencensus do not exist in every field. Scholars of French, English or Oriental manuscripts reported using the Internet Archive\(^{18}\), DMMmaps\(^{19}\) and Arlima\(^{20}\) as centralized resources, but they most often search library sites directly for digitized materials, assuming they already know where the manuscript is held. This may be true even when they don’t yet know whether the manuscript has been digitized; an early-career researcher working for the most part with relatively obscure Arabic manuscripts reported that “whenever I am interested in a [French] manuscript I go on Gallica and check, but they are never there.” With familiar resources such as the BnF or the British Library, this might be an easy task (difficulties of shelfmark format aside), but several interviewees described difficulties in finding libraries’ digitized collections. For example, according to a late-career Greek scholar:

There are two ways of getting to the Bodleian website, and one of them has got access to collections, and if you go to special collections, you'll find something about digitized manuscripts. Somewhere one can find a heading saying digitized manuscripts but you have to hunt for it a bit. And then you get a series of options, and if you’re after the Greek manuscripts, it’s more helpful to go down further and find the Polonsky website, where you can find a bit more information. So in other words it's not very coordinated.... Initially I found myself going around in circles, and then not being able to refind my route. When I looked yesterday, that’s when I’d been practicing a bit more [laughs], there was something on one of the Bodley main sites, there was a menu item that said special collections, and then

\(^{15}\) http://pinakes.irht.cnrs.fr/
\(^{16}\) http://www.handschriftencensus.de/
\(^{17}\) http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx
\(^{18}\) https://archive.org/
\(^{19}\) http://digitizedmedievalmanuscripts.org/
\(^{20}\) http://www.arlima.net/
catalogues. I had a faint feeling that something was being done with the 19th-century
catalogues and they were being fed into a sort of an easier access system and I thought I’d
found that at one point, but I’ve not been able to find it since.

Most interviewees reported using Google for different stages of their searches. Several
reported Googling shelfmarks or names of manuscripts as a preliminary discovery method,
although one reported doing this “only out of frustration” after he had failed to find the desired
manuscript via an institution website. Several interviewees expressed doubts about Google’s
effectiveness in indexing digital library collections. One admitted to using Google occasionally
herself, but said that she wouldn’t recommend it to her students, because it tends to turn up blog
posts or news items rather than links to digital surrogates. Another reported using Google to find
the collection itself, but said that “even if [digitized manuscripts] are in relatively helpful
repositories, there’s no immediate kind of straight-in-from-Google route into looking at them. My
instinct is that that’s a mixture of...that not being a priority for your library, and also Google not
seeing that as a crucial thing that they should [index]”. Another manuscript scholar said similarly
that “Google I think doesn’t search immediately to the deeper of libraries’ collections”.

Several interviewees, recognizing this deficiency, use Google to cast a wide net. A printed
book scholar has set up a Google alert for the names of printers and owners she is interested in,
with “lots of different kinds of searches and spellings.” An Arabic scholar has used Google to find
travel blogs or amateur scholarship that have in some cases alerted him to manuscripts relevant to
his research, although this information is often inaccurate. One of the Google skeptics quoted above
also reported using the “slightly lateral” strategy of Googling to find announcements of digitized
manuscripts: “So for example, I might remember to, instead of searching for the shelfmark...search
for ‘really exciting copy of X digitized by this library’. And I’m glad that libraries do that, because
Google is much better at searching news posts than searching databases of digitized images.” This
strategy was echoed by another manuscript scholar, specifically in reference to the British Library,
which frequently publishes blog posts about recently digitized manuscripts. A third scholar,
however, said that she avoids simply Googling shelfmarks for the very reason that these searches
usually turn up blog posts rather than links to digital surrogates, and that, if the blog posts do link
to digital surrogates, the links are often out of date.

As expected, given the results of the questionnaire, most interviewees reported receiving
help or direction from colleagues either at Oxford or at other institutions in finding digitized
materials. One called colleagues “very very useful,” and a late-career scholar reported that her
research collaborator is “very up on all things digital...he’s my first port of call”. Several reported
hearing initially about digital resources either at conferences or lectures or by word of mouth. When it came to specific digitized resources, interviewees mostly reported turning to colleagues—usually by email—after trying and failing to find resources on their own. As an Armenologist put it,

I think in these niche things where many aspects are still in written print form, you rely on people telling you where to look. And that might be a librarian or it might be a colleague, and the main exchange of information probably occurs by email.

Some interviewees were more reluctant to ask colleagues for help: “I would generally assume probably incorrectly that we all have the same awareness of what is available.” Interviewees were largely even more reluctant to request help and information via social media, although several reported seeing other colleagues make requests via Facebook or Twitter. Only one interviewee, a Hebraist, mentioned a “specialized interest group” on Facebook, specific to Hebrew paleography and codicology, which she finds useful for exchanging resources and information. Several other interviewees, however, reported that they have found social media groups inefficient and too diluted to be useful for research purposes, and one said that asking for help on social media amounted to saying, “I’ve failed in my scholarly rigor, can someone please help me out”. This suggests that at least in this interviewee’s field—Anglo-Norman studies—there is an expectation that individual research skills should be sufficient to discover the desired materials. When this does not turn out to be the case, scholars might be more likely to turn privately to colleagues with whom they are already friendly, rather than posting public requests on social media.

Interviewees reported approaching library staff with similar reluctance. One explained that while librarians are “almost always responsive and helpful,” he rarely contacts them for help finding resources because “often I can establish reasonably confidently that it’s just not there or it’s definitely there”. Several reported contacting librarians only if they know them personally or if they are attached to a small (and therefore, generally speaking, more helpful) library. One interviewee recalled a librarian at a Cambridge college who took pictures of an undigitized manuscript and emailed them to him. He added that “[if] you send something like this to the [BnF]...you know they’ll be too busy deleting it to actually get round to despising you for asking the question in the first place”.

5.2.2.2 What they search for

The results of the questionnaire indicated that, depending on field of study, shelfmark and title were the most commonly searched-for metadata items, suggesting that by and large scholars already know exactly what they are looking for. This was generally borne out in the interview portion of my study. Shelfmarks and titles were not the only popular search terms (one interviewee
described searching for "title or shelfmark or person or date or content"), but they were the most popular. Interviewees reported finding shelfmarks in references in secondary literature or in print catalogues, and then going to Google or a library repository website to find a digital surrogate.

One reason for searching by shelfmark is that manuscript and even printed book scholars encounter many problems with transliteration systems and variant spellings. It is rare for a repository or finding aid to reproduce Greek, Arabic or Hebrew catalogue information in the original alphabet, and there is little consensus between institutions as to what transliteration system to use. Inconsistencies can crop up even within a single institution; at the Bodleian Libraries, for example, the online catalogue of printed books, SOLO, uses a different Arabic transliteration system from the new TEI catalogue of Arabic manuscripts. A Greek scholar reported "hideous spelling problems...you get used to just bashing away at it: you know, is it Latinate or Greek, is it a C or a K, and what have they decided to call it". An Arabic scholar reported searching for titles in Arabic script via Google, because "with Arabic there's only one way of writing the title of the book, but if you write in English there's ten ways of writing it"; the result is that "I spend a lot of time Googling stuff". They regarded this process of trial and error as tedious but inevitable.

Shelfmark searches have their own problems, however. Shelfmarks may change over time, so that the shelfmark referenced in a piece of scholarship from the 1990s might not still apply. Moreover, shelfmark format is not standardized, and library repository search algorithms may not accommodate variations:

So for example when you order manuscripts digitally at the [British Library], you have to format them as major element of shelfmark, then MS for manuscript, then number, so for example Royal MS E 9. Most scholars working in English at least would conventionally write MS Royal E 9. So you just have to remember that as far as the BL is concerned, that's Royal MS E 9, in their online ordering thing. I think their online digitized image thing might be more forgiving actually. But not every library—I mean the BL’s a big outfit with more resources, and not every library is going to be putting things online and thinking about have we accounted for the five different ways someone might format a search...and that's a big problem for searchability, is just trying to work out how the shelfmarks will be thought of in the library that owns the book.

Another manuscript scholar reported a similar experience, adding that "as far as I know these rules [for formatting shelfmarks] aren't written down". Another said that "finding stuff by shelfmark is really difficult" and that "getting a standardized way of talking about a physical manuscript is I think really important, but everyone has to agree on it and they just never will". Interviewees praised resources such as Gallica and Pinakes, which have more tolerant search algorithms, capable of accommodating variations in punctuation and abbreviation of shelfmarks. To
further address this difficulty, Pinakes has recently begun incorporating Diktyon numbers\textsuperscript{21}—the equivalent of a unique manuscript ID—into its item records, making individual manuscripts easier to find and cite. One of the Greek scholars interviewed expressed appreciation of this feature.

Scholars of printed books, meanwhile, are more likely to search for title or author or other even less standardized fields. One printed book scholar described a painstaking search process:

If I knew what I was looking for...I would go into EEBO here on my desktop, and I would search for [the author's name], because I know that the title search on EEBO is...actually really difficult to do somehow, because the spellings are kind of weird.... So that's what I would do. And then I would get about 80 hits I guess, and I would just kind of scan through them until I found what I wanted. So I could do a much more specific search I can see, but I very rarely would do. My experience—it's probably my own mismanagement but my experience is that quite often it says there are no results, so it actually seems easier to do a broader search and sift through by eye rather than getting it to do that.

The other printed book scholar interviewed reported a similar experience. He is often able to find items in EEBO by their short title catalogue numbers, but as he is often not looking for a particular book, he searches by keyword and is vigilant about false positives due to variant spellings: "It's knowing the limits of what's available that produces better results, and also distrusting the results in the first place". Nevertheless, he drew a contrast between digitized manuscripts and digitized books: "What libraries tend to do is catalogue printed books very efficiently, because of author/title/date, and their cataloguing of manuscripts may be very good but it's hard to find what you're looking for and then to get from the catalogue to a digital form of the manuscript, and that I think is something that needs more work."

The only interviewee who reported no problems finding what she needed in a given repository was the German manuscript scholar. Thanks to the Handschriftenbuch and the enormous cataloguing and digitization efforts of German libraries, she can simply go to the website, find the institution, choose the shelfmark from a list, and see a comprehensive catalogue entry with (where applicable) a link to a digitized copy. While she admits that the strategy of browsing to institution and then shelfmark is not always intuitive to new users of the website, who prefer to do keyword searches or browse an aggregated shelfmark list, we have already seen that manuscript

\textsuperscript{21}http://www.diktyon.org/en/identifians-du-reseau/manuscrits/
scholars mostly already know the institution and shelfmark before they begin searching, making this an effective system.

### 5.2.2.3 Attitudes toward the search process

For the researchers interviewed in this study, locating digitized items online is a constant, iterative process, involving a range of strategies and skills. Describing his typical search process, one said:

> First of all I’d have a think about, is it part of an obvious collection. So I think all the medieval manuscripts at John Rylands are digitized, so if it has a JR shelfmark I’ll just say, well that must be digitized, [and] I’ll go to the JR website, try to remember where their online thing is. But of course that’s normally not the case. My next step would be to ask…my colleagues on the project, because they know the…corpus better than I do and they might know it. Then I Google, usually for the library—if I remember the name of their digital site, so University of Pennsylvania has a site called Penn in Hand, I’d Google for that, and then I would search that site for it. I might in parallel try Googling the shelfmark of the manuscript, because if you’re very very lucky, occasionally that will actually turn up a digitized version.

Asked about how long this process usually takes, he said:

> I mean, it’s not something I’ve tried timing, but I feel like it’s the sort of task where I might spend five or ten minutes on it, and at the end of that I’d either have found it or I’d have not found it and been annoyed by that, and I’d come back to the problem later, I think. Because in the next few hours I’d probably think of a couple of other ways into—things I should search for or things I should check or even people who aren’t librarians, just friends who have worked on manuscripts. It’s one of those tasks where if it isn’t going well, it’s not as though you spend four hours concentratedly trying to determine—I mean, after fifteen minutes you kind of know whether or not you can find it easily—but it can kind of linger as a problem on your to-do list that sits there and sucks up time over a few days as you think of things and try them and don’t succeed.

Competence in this area is viewed as a professional necessity; interviewees described it in terms of “scholarly rigor” (see above) and being “a proper researcher”. Unsurprisingly, many appeared to take pride in the knowledge base they had built up, remarking on how they had figured out many things by themselves, or citing instances where they had been able to pass their knowledge on to others. Two, when asked about times when they had turned to others for help, expressed that they believed they knew as much as anyone else did about available digital resources.

This pride in and value of resource discovery skills was in many cases paradoxically accompanied by a stated lack of confidence. A researcher might know how to find something, but be doubtful that their method was the best way of finding it. These disclaimers about methodology occurred frequently in the interviews:
• “I think only out of frustration...would I go to Google to see whether I am just too stupid to find it via internal means”
• “Again, this may be just that I’m inefficient and there may be a better way to do this”
• “…my use of and promulgation of digital books...feels a bit random to me”
• “…it’s probably my own mismanagement but my experience is that quite often it says there are no results”
• “I’ve been able to—amusingly, as I am technologically incompetent—to show it to people who are much more technologically competent than I am”
• “…I don’t know whether this is right or not…”
• “I suppose I was just being dim”
• “…maybe I’m not particularly good at this”

It’s possible that these disclaimers were merely conversational tics and not indicative of actual insecurities, but the pattern suggests that researchers are liable to blame themselves for an unintuitive or ineffective resource discovery system. This may result in a higher degree of tolerance for suboptimal systems (for example, one late-career researcher said that she doesn’t mind struggling with difficult interfaces because she assumes they are intuitive for someone who has “internalized” how digital interfaces usually work), which could conceivably be a blessing for libraries struggling to meet the high usability expectations of Internet-savvy researchers accustomed to Google’s hypertolerant search algorithms. Given that researchers view the discovery and use of digital resources as an essential scholarly skill, however, this would seem to be unnecessarily damaging to their self-confidence.

It is worth noting that one interviewee expressed not resignation or even pride, but actual fondness for the process of discovery, saying, “I quite enjoy looking for things. I quite enjoy the difficulty in finding things. It’s quite nice to, you know, find [a digitized manuscript] in obscure places.” This response echoes the questionnaire respondent who said they were “happy having a happy constellation” of digital resources, and crucially, both the questionnaire respondent and the interviewee identified themselves as librarians. Considering that many digital library collections are designed by—or at least with input from—librarians rather than academics, this finding may serve as a warning: we must not expect that our users share our enjoyment of the search process.
5.2.2.4 **Guidance for students**

As discussed above, many of the interviewed scholars reported encouraging or requiring their students to make use of digitized rare books and manuscripts. Depending on the level of study, their instructions on how to find these resources may be quite granular. One reported gathering links together in a blog post and directing students there, while another reported giving his undergraduates direct permalinks or specific instructions, but giving his postgraduate students only the URL of the digital repository, "because really they will very quickly have to be finding manuscripts themselves, so they may as well learn how". Along a similar line, several set their students short discovery assignments to introduce them to this area of research.

In a pattern across the interviews, scholars reported setting their students modest discovery tasks and finding that the more “enterprising” or “diligent” students would expand on these to find other digitized items—in some cases, items of which the scholars themselves had not been aware. At least one interviewee called this “embarrassing”, because prior to her students’ discoveries she had thought she knew about all the available resources. This phenomenon may be linked to the general perception among interviewees that their students are naturally good at online resource discovery: “It’s probably an erroneous assumption, but I tend to think that [students] are going to be a lot better than I am, generationally, at finding stuff online.” The interviewee who had compiled blog posts of resources for her students said that she had originally thought “I’ll just tell them to find [the resources], you know, they’re all better at this than me”, but that she had deemed the blog posts more efficient in the end. Even one of the youngest interviewed scholars attributed his students’ success in this area to being “five or six years younger than me...[and] usually even more comfortable with websites than I am”. None of the interviewees said that their students frequently ask them for help in this area.

5.2.3 **Citation**

Reported citation practices varied widely across interviewees. In some cases, the interviewee’s standard practice depended on research context. One said that if he consulted a manuscript digitally, he would cite the digitized version specifically “only where I think it makes a difference”, meaning if he were discussing material aspects of the manuscript, not simply the content, or if the digital images were particularly low-quality. Another said that he would cite the digitized version if he were writing an in-depth study of a particular manuscript, but not if he were simply mentioning it in passing. Others indicated that their citation practices depended on the end product; one said that he would follow whatever style guide he was using, citing the digitized
version only if the style guide specified that he should, and another said that he would cite the digitized version in an electronic publication, but that it would be “too cumbersome” to cite URLs in a printed book.

Several interviewees were more dogmatic about citing digitized versions. One Greek manuscript scholar reported that he always cites both, saying that “you have to declare and say clearly, yes, I've seen the digital image plus the original, or just the digital image.” Similarly, a printed book scholar said that “I think you have to, when you’re referring to something, let the reader know that it is available online in one form or another”. One interviewee, when asked about citation, said immediately that this was something she had recently been “discussing extensively”, and that she and her colleagues had settled on giving the shelfmark for each manuscript along with the domain name of the digital library, “because that's pretty sure to stay the same and should normally be enough to find” the digitized manuscript, and because top-level links are generally shorter and therefore easier to typeset and read.

The interviewees who did not report regularly citing digitized versions did not seem to believe that using digital surrogates was stigmatized. Only one interviewee said tentatively that using a digital surrogate without ever having consulted the original might be “viewed negatively, in a way”. One book scholar acknowledged that she had felt this way before, but no longer did:

I think in the past I would have sort of elided the digital one into the original copy, partly out of a kind of feeling that I wasn't quite sure about how to cite a digital copy, and partly...maybe because of a feeling that it wasn’t quite so good to have consulted a digital copy rather than the original. I don't feel that second thing anymore.

Furthermore, even the interviewees who indicated that they didn't always cite digitized versions nevertheless appeared to perceive the benefits of citation. The scholar who found printed URLs cumbersome acknowledged that he regularly follows URLs to digitized resources in the literature he reads. Another, who said that he didn't “feel that the precise mode of consultation is an essential thing”, acknowledged that he might cite digitization projects he’d worked on “as a self-promotional thing”, and that in any circumstance “it is quite nice if you can point somebody towards the resource if they wish to...check the claim themselves”. Another expressed doubt that anyone would ever follow links in a printed book rather than simply Googling, for example, “University of Stuttgart First Folio”, but reported that she did include URLs for the sake of those who might want to. While it was less common for interviewees to acknowledge the benefits to libraries of citing their digitized resources, a few did so:
I think in a way lots of colleagues who don't have [Early English Books Online]...find it really hard to get their institutions to subscribe...because if you look at EEBO in terms of citations, it looks very low. So it looks as if nobody uses it, but of course absolutely everybody uses it, even if you've got an amazing rare book library across the road as I have...so it seemed to start to be kind of an ethical point to say this is a resource that we're all using.

Similarly, a manuscript scholar, while acknowledging that citing digitized versions isn’t “a common practice”, said that “it’s good to indicate that these things are available” and that there is an “ethical or philosophical” element to the question, as it’s “worth acknowledging the effort...to create” digitized resources. He mentioned a digitized collection of 19th-century records:

They’re very keen for people to cite them rather than the text that they took the information from, so they have stripped out page numbering from the information they present, so you can’t know where in the physical book you are, so you’re forced to [cite the digitized version].... There’s a little page on their website where they explain it—not at all aggressively—and they have a reasonable case.

The principal disincentive for citation of digitized versions appears to be inconvenience. Scholars lack guidance (one reported that the most recent MLA Style Guide “was almost silent on how to cite a digital copy of a book”) and may not want to clutter up their footnotes and bibliographies with long URLs that “look messy” (five interviewees mentioned the hassle of transcribing and typesetting URLs). They may also be discouraged by past experiences of link rot or unaccommodating typesetting; one interviewee went to significant effort to include a list of links to digitized manuscripts referenced in the text of a recently published book, only to find that even in the electronic version of the book the links were flat.

Given that, for almost every scholar interviewed, citation of digitized versions appeared to be a matter of case-by-case consideration (only two or three described their citation practices as mandated by style guides or as “just how it’s done”), and given the appreciation the interviewees showed for the work that has gone into digitizing library holdings, it is important for libraries to make it easy to cite digitized versions and to make clear the advantages of doing so.

5.2.4 Desires

5.2.4.1 Discovery systems

The interviewees were asked to describe their ideal discovery process for digitized rare books and manuscripts. As in the questionnaire responses, efficiency was much prized; one interviewee said the most important thing was to “find everything as quickly as possible”. Beyond this, however, the responses centered around two desires.
A majority of interviewees expressed a desire for a searchable online catalogue with links to digitized items. In some cases this already exists; the two Greek scholars praised the British Library’s new digital collection of Greek manuscripts, which smoothly integrates catalogue entries and digital surrogates. Further, the existence of Pinakes, which includes brief metadata for each item across institutions, and the Handschriftencensus, which includes full catalogue entries, mean that Greek and German manuscript scholars are in relatively good shape when it comes to online catalogues. Fittingly, these interviewees were mainly concerned with improving these existing resources by planning for future funding and making it easy to submit corrections. Researchers in other areas are in a worse position. The Bodleian Library, as several interviewees pointed out, does not currently have up-to-date digital catalogues of any of its manuscript collections. In fact, the Arabic scholar reported that very few of the Bodleian’s Arabic manuscripts are catalogued at all, even in handwritten or print form. Scholars in these areas wished for online catalogues with links to digitized items.

A centralized catalogue that incorporates resources from a range of institutions was also perceived as potentially useful by several interviewees, particularly for the early stages of discovery. It was not, however, perceived as crucial, and interviewees tended to be skeptical of the feasibility and effectiveness of such a catalogue. One interviewee expressed concern that the “fragmented” nature of digital manuscript resources made a union catalogue a tall order, but that in the future there might be “a possibility of coming together”. Another said that a union catalogue “would be awesome…if it worked”, but that the union catalogues he had tried hadn’t helped him to find things. Another acknowledged that a “more global search engine” would be helpful in cases where he didn’t know exactly what manuscript he was looking for, “but I doubt that would be the primary means for me to find something, unless I’m really desperate”; if he knew where the manuscript was held, he would go to that specific library’s website. A printed book scholar, meanwhile, said that linking from an individual library catalogue to COPAC or WorldCat would be “great, although it would be a massive job,” but then added that as her research mainly involves copy specifics that are rarely catalogued online at all, it would have limited use for her: “I don’t think I...have particularly sophisticated search requirements, I’d just like to be able to find things”.

While desires for browsing interfaces and metadata facets varied from one interviewee to another (see 5.2.4.2 below), almost every manuscript scholar interviewed mentioned that it would
be useful to have a single list of digitized shelfmarks, either aggregated or institution-specific. For researchers who already know the shelfmark they’re looking for, a resource like this can be much faster and more reliable than wrestling with different institutions’ search syntax requirements. A Greek scholar, working with Bodleian materials digitized for the Polonsky Foundation Digitization Project, praised the Polonsky Foundation Digitization Project website, which is not searchable and contains no metadata, for its inclusion of a list of shelfmarks with links to their digital surrogates. She uses this page rather than searching for the manuscripts in the Bodleian’s new digital library platform, Digital.Bodleian. For some interviewees, the list of shelfmarks would not even necessarily function as a way in, but simply as a time-saving guide to what is out there, as well as a template for how to format shelfmarks. An English paleographer described his desired discovery tool as

a…page where the library would just say, here’s a full list of every medieval manuscript we’ve digitized, and you could ctrl-F, or there’s a search box at the top, but you can ctrl-F for them. Ideally I guess it would be nice if those shelfmarks were then hyperlinks to a stable URL for the manuscript images, but I’d be perfectly happy to [copy the shelfmark] and bang that into a search interface. Just knowing it was there would help.

One interviewee, in fact, had compiled such a list in the front matter for a book she had recently worked on. The list contained every manuscript referenced in the book, along with URLs, where possible, for digital surrogates.

5.2.4.2 Metadata

5.2.4.2.1 Manuscripts

The most important metadata field for manuscript scholars is the shelfmark. In most cases, scholars are coming to digital collections from manuscript inventories or handlists or mentions in secondary literature, so they already know the shelfmark they’re looking for. Scholars struggle, however, with shelfmark formatting conventions (see 5.2.2.2 above), suggesting that in order to make discovery via shelfmark truly effective, libraries need to standardize their notation, publish their rules, and employ search algorithms that accommodate variations in notation.

In cases where scholars do not know the shelfmark of the resource they’re looking for, the list of desired metadata for discovery expands. An Arabic scholar building a database wanted to search by date, place, author and more, in order to find manuscripts he might not already know about. A Hebrew scholar said that while she usually used shelfmark, she might also look for a manuscript either by popular name (for example, the Kennicott Bible) or by genre (for example, any mahzor). An English medievalist allowed that while he always has “an independent handlist of copies of” whatever text he is looking for, a scholar of a later period “might not be in that lucky
position” and would want to search by, e.g., author or date range, because there might be a larger number of potentially relevant manuscripts. (Regrettably, no post-medieval manuscript scholars were interviewed, so it’s impossible to confirm whether this is true.)

Despite having such modest requirements for searchable metadata, most interviewees expressed interest in seeing a broad range of additional metadata alongside the digitized object. Two mentioned quiring information as a desideratum, and those interested in art history or musicology mentioned indexes or annotations of illuminated pages or musical notation. Two mentioned the names of calligraphers and scribes. One mentioned origin and provenance as the most important, but added, laughing, that “I guess in some ways you just want as many things to filter by as possible”. Others echoed this sentiment. The only interviewee who couldn’t think of any additional metadata she would like to see included was the German scholar, who had access to a comprehensive digital catalogue in the Handschriftencensus.

Metadata may matter less, however, if it is perceived to be unreliable. One interviewee expressed indifference to accompanying metadata, saying that "the things you might want put into the metadata—provenance, dating, physical size, quiring—all of these you’re probably going to want to check yourself one way or another”. Another said that it was essential to see where the metadata had come from:

I'd like datings, but dating of manuscripts is a complicated issue, so I would want to have the datings and I'd want to know who's dating them to those dates, like which person and when, and in what publication and on what grounds.

She acknowledged that this desire was "pretty hardcore", but said that she had run into problems before with libraries that “just put the dating that they had in the old catalogue” into their digital metadata, which may have been made obsolete by new scholarship. To combat this problem, she suggested a system that allows input from the scholarly community:

I think it would be useful to have a better mechanism whereby people could feed back stuff. Clearly it can’t be totally crowdsourced because it’s very specialist information, but you know, if you could register and they would know for example that I’m registering and I’m not just any old body...maybe my registration would be subject to a process of minor review, to just say who you are, and then when I send in stuff then it might get put up. But it would be ascribed to me, because then if one of my colleagues says no I don’t believe that dating at all, and sends in a rival bid for the date, then that could also go up with their name and their grounds for thinking that too. So I think it has to be slightly more of a discussion around manuscripts, because they’re complicated objects.

The idea of an ongoing discussion surrounding metadata is particularly important. It is easy for librarians and developers crafting a digital library to forget that attributing a particular date to a manuscript—or even formatting the date in a particular way—often involves a statement of
opinion rather than fact. A system that could display multiple versions of each metadata field, with multiple degrees of certainty and explicitly stated sources, might more closely mimic the true state of knowledge in manuscript studies. This is not something that most of the interviewed manuscript scholars seemed to expect, however. No other interviewee suggested such an involved solution to the problem of incorrect metadata as the one quoted above, but two others said that it was essential to see what catalogue the metadata had been taken from, so that they could know whether or not it was correct and up-to-date.

5.2.4.2.2 Printed books

The interviewed scholars of printed books, having access to EEBO, ECCO and the Incunable Short Title Catalogue (ISTC) as well as reasonably detailed MARC catalogue entries provided by their libraries, had fewer requirements for metadata that were not already filled. One scholar said that he encountered the most trouble when looking for manuscripts related to his bibliographic research, manuscripts being on the whole less well-catalogued than printed books. The other printed book scholar, however, who is most interested in provenance and other aspects of books' material history, reported that most libraries don't catalogue this kind of copy-specific information online. Instead, they might keep it in a physical folder, which she would have to visit the library to consult. While digitized copies make it easier to check if a specific book has annotations or marginal notes, they do not make it much easier to search among the books that do. For this purpose, the interviewee wished for a brief note indicating the presence of copy-specific notes or features in a given book.

5.2.4.2.3 Links

Permalinks—persistent links to digital resources—were identified by several interviewees as important for research. Interviewees mentioned permalinks firstly within the context of streamlining their own research. Several noted the inconvenience of digital libraries that do not provide stable URLs for digitized items; one mentioned a particular library that generates new session URLs for each visit, making bookmarking impossible. Another reported discovering BnF manuscripts using Pinakes rather than Gallica, simply because Pinakes makes it easier to find the permalink. Most, however, mentioned permalinks within the context of citation. One interviewee, describing her practice of including links to digitized images in citations, said that "we just have to hope the stable URL is really stable". Another said that he would always cite the digitized version, but that he would only include the URL if it “really is a permalink”. He went on to express frustration with the British Library, which is in the process of updating its digitized Greek
manuscripts portal and had not made clear to him whether or not the old manuscript URLs would still be valid.

In addition to being stable, scholars also want the links to be short. Three interviewees mentioned the inconvenience of including long URLs in citations, both because they are ugly and because they are difficult for readers to then type up. One said explicitly that he would be more likely to include URLs in his citations if there were a mechanism for obtaining a short, stable URL.

5.2.4.3 Images and viewers

Compared to metadata and discovery systems, considerable effort has been spent researching and creating reliable and usable images and viewing interfaces. Unsurprisingly, interviewees generally expressed positive opinions about the current state of the art in this area. Many of those interviewed still work with black-and-white microfilm or scans of microfilm, on which high-resolution full-color digital facsimiles are a transformative improvement. This is subject to geographic variation, however. One interviewee reported having some trouble with low-resolution images from the National Library of Algiers, and another, describing the first time he had ordered a scan from the Bodleian, said he had been “astounded that it was a 150MB image and I could see every single hair on the parchment, as opposed to the 3MB we got from somewhere in the Caucasus, where I could see every single pixel.” Another recalled that some of the libraries he has worked with in Greece digitize intentionally poor-quality images in order to protect their copyright and prevent unauthorized reuse. This scholar expressed a particular wish that libraries would “agree on a standard high-quality version that is to be uploaded”, so that money and time would not be wasted on images that are unsuitable for research.

In addition to resolution, scholars are concerned with color fidelity and size indicators. Many institutions use colorcheckers, photographed alongside the book or manuscript, to provide context for the colors of the images. Several interviewees noted that this was useful, although one of these said it was enough to photograph the colorchecker once and then to leave the researcher to analyze the images to find out if the exposure and color matching were the same across an entire digitized object. Color fidelity is especially important for art historians, of course, but even for paleographers and codicologists it can have scholarly implications, as researchers may build their arguments on the color of ink or parchment. Generally, however, researchers seem to be unlikely to trust digital surrogates on these points. One paleographer described size and color as “things I’m just not happy to rely on digital images for”; even when a ruler has been photographed alongside a manuscript, inconsistent depth or warped pages might give an inaccurate idea of the manuscript’s
size. Another said that "seeing something on a computer screen can be a bit treacherous," as it's difficult to gauge size, or to observe subtle qualities in the support or ink without looking at the manuscript from multiple angles. Similarly, an art historian noted the drawbacks of thinking of manuscripts as two-dimensional digital objects, as it is necessary for her to look at illuminations from several angles to determine, for example, how the paint was applied.

Regarding the experience of viewing the image, interviewees had few major requirements. Two expressed dissatisfaction with the Luna image viewer, used by several major institutions, which allows for very detailed and granular metadata but makes it difficult to page through images in a sequence. When prompted, several said that the ability to rotate and adjust contrast or brightness would be helpful, but that they manage this themselves by downloading the images and editing them in software such as Photoshop. Page labels are key; several expressed dissatisfaction with the BnF’s unpaginated microfilm scans, one said that it would be helpful if signature numbers were provided as well as folio numbers in printed books, and one criticized a Princeton library for making it possible to navigate by quire at the expense of navigation by folio. (He maintained, however, that including quiring information or diagrams would be helpful if it were done in a different way.) In addition to folio numbers, adequate zoom and faithful representation of a book or manuscript as a single object, rather than a set of disjointed page images, appear to be the most important considerations.

5.2.4.4 Downloading

With the high proportion of questionnaire respondents who indicated a preference for downloading images in mind, interviewees were asked for their opinions on downloading. Nine said that they did like to be able to download images, either to work offline (at “a beach or a café where there isn’t reasonable internet”), to edit or manipulate the images, to include them in PowerPoint presentations or lectures, or to remix and juxtapose them to assist with analysis. As a musicologist, who described downloading as "crucial", put it:

I often just need to have a reference picture that is easily accessible to me, and particularly be able to put several different versions of a song side by side on a table. A physical tabletop. And you can’t really do that on a screen. I mean, I could have four different screens and blah, but actually having a physical printoff of a file is really valuable.

Another said:

If you are allowed to download...the image, you can do whatever you want. And this is what you have to do if you are a proper researcher. I mean, if you are researching and you are
trying to find out what's happened... sometimes [downloaded] images are really helpful, and
they give you a better angle than the manuscript itself. Imagine that you have a very tight
manuscript and something written has been faded or something like that, and then you have
the good quality image: you can zoom in, you can play with software, and you [can] find a
better image. Like using X-ray.

Two of the interviewees who rarely if ever download images said that the downloadable
quality is too poor and that high-quality downloads are too unwieldy. Another said that whether he
would like to be able to download images depends on the repository; he has no desire to download
an image that is already easily accessible online, but “with older repositories...it’s tempting to do it”
because the online images load so slowly. This reflects the view of the questionnaire respondent
who said that downloading images was greatly preferable on a slow network, where the download
could be started and left to run overnight, rather than waiting several minutes for each image to
load in the browser.

It is worth noting that none of the interviewees who reported downloading images in order
to more easily manipulate and compare them were aware of the tools that are already available to
do these things online. The Mirador viewer for IIIF-compliant resources, which has been available
for more than two years, makes it possible to view different digitized objects side-by-side in the
same viewer, and has recently added features to allow annotation and manipulation of image color,
contrast and brightness. One early-career scholar, discussing his ideal viewing interface, seemed to
be describing Mirador exactly:

In an ideal world maybe there would be a single viewer, right, across institutions, which
would allow you to compare manuscripts and go straight to particular points in a manuscript
or multiple manuscripts, that would be wonderful.

When asked, however, he said that he had not heard of it. More effective dissemination of these
tools is clearly necessary.

6 CONCLUSION

This research was undertaken with the aim of answering a few preliminary questions about
how a specific scholarly community discovers and makes use of a specific set of digital resources. It
was expected that participants in the study would report frustration with the available resources, a
standard set of discovery tactics, and an unwillingness to go beyond these tactics to locate hard-to-
find resources. Instead, participants were largely very appreciative of the resources available to
them, and patient with their limitations; moreover, they appeared to view the discovery of digitized resources as a core research skill, for which they called on a large body of knowledge gained primarily through personal trial and error.

While it is not astonishing that a group of professional researchers should be diligent in their research, the particular diligence of rare book and manuscript scholars is most likely linked to the uniqueness of the materials. While the digital age has brought a surfeit of information to many researchers, a scholar seeking a particular digitized manuscript either finds it or doesn’t; there is no option to settle for something more convenient. The relative uniqueness of special collections makes them “one of the last reasons that a scholar needs to go to a library” (Madsen 2010, p. 42)—or to a particular library’s digital collection—in the age of e-books and interlibrary loan, but it also means that libraries need to tailor their digital offerings to those who cannot satisfice, as well as those who can. There are two main points here:

1. Librarians concerned with discoverability should not be content with evidence that scholars have found their digital collections, as this by no means indicates that they are easy to find.

2. The discovery behavior of most library and web users, with its tendencies toward satisficing, is not representative of the behavior of specialized scholars, and discovery interfaces must be designed to accommodate both.

Point 2 is epitomized by the concept of the shelfmark index, which would be useless to non-specialists (and which has no analogue in the commercial web, where keyword searching and facets take primacy over long unfiltered lists), but which was an obvious good to the manuscript scholars interviewed. There are exceptions, of course; for example, a scholar seeking a reproduction of a 12th-century illumination for a lecture might not have a particular manuscript in mind, and would be grateful for a more commercial-style interface with search facets. Moreover, the need for specialized and old-fashioned-seeming discovery tools, like the shelfmark index, may fade as libraries develop better metadata and more tolerant search algorithms, allowing commercial-style searching strategies to work better for specialized scholars. The crucial point is that resources designed for specialists as well as non-specialists must accommodate a variety of information behavior types. The recommendations that follow are intended to make life easier now, for specialist scholars seeking specific resources; as technology changes, better options will most likely arise.
6.1 RECOMMENDATIONS

6.1.1 Creating discoverable resources

As one might expect, and as this study has confirmed, the most useful means of discovering digitized special collections resources would be a complete and up-to-date electronic catalogue—either global or institution-specific—with direct links to digitized objects. This is a tall order for libraries that are struggling to fund digitization and cataloguing projects at all. There are smaller measures that can be taken, however, to ensure that a digitized collection facilitates discovery as much as possible.

6.1.1.1 Standardized notation

Variations in shelfmark formatting pose a significant hurdle to manuscript scholars, as they must try searching for many variations and may, simply because they haven’t thought of a particular variation, fail to find a digitized manuscript. For scholars of Arabic, Greek, Hebrew and other languages with non-Latin alphabets, or with early printed texts, this problem is multiplied, as each search term may be transliterated in a number of ways. While it is clear that scholars using special collections are willing in most cases to put in the time to discover things through trial and error, measures must be taken to minimize the need for this, both to lower the barrier to entry for earlier-career scholars and less experienced digital searchers, and to ensure that scholars do not miss out on digitized resources through an infelicitous series of attempts. It is recommended that libraries:

- ensure that their own shelfmark notation is standardized across resources and across platforms.
- make available documentation of their shelfmark notation system, so that scholars can refer to it in searching and citing.
- adopt and accommodate unique reference systems such as Diktyon IDs as these develop.
- employ more sophisticated and tolerant search algorithms that accommodate variations in shelfmark notation.
- make explicit which transliteration systems they use.
- where possible, create dual-language records and allow searching in non-Latin alphabets or characters.
- employ and contribute to VIAF and other name authority files.
6.1.1.2公共索引

为了进一步防止无效搜索，建议图书馆实施公共数字目录。作者、主题等索引可能同样值得。这些索引可以被聚合为一个全球查找辅助工具，虽然比集中化的目录要不那么雄心勃勃，但仍然可能非常有用。简单列表也有优势，它们不像电子目录，可以被搜索引擎爬取。虽然创建可爬取目录的问题需要在长期中得到适当解决，图书馆需要做更多来优化其数字图书馆网站的搜索引擎发现，简单列出目录可能是一个简单的方式来缓解学者试图通过Google搜索的材料时遇到的问题。

6.1.1.3永久链接和引用指南

正如这项研究已表明，学者们忽视了对数字副本的引用，因为他们不知道如何或为什么这样做，因此，记录引用的必要性和正确程序至关重要。Kaislaniemi（2014）和Froehlich（2015）就EEBO-TCP的引用问题提供了优秀的经验，图书馆在其数字藏品页面上发布的引用建议可以借鉴Kaislaniemi和Froehlich的建议，不过最好要咨询跨机构的数字人文学者，并尝试形成关于最佳引用实践的共识。已发布的建议还应包括对引用数字版本为什么重要的解释。由于长期和不稳定的URL也是一个问题，图书馆应确保（理想情况下）永久链接在数字图书馆界面中很容易找到。DOI可能成为特藏数字化领域广泛使用的方式，此时应显眼地列出。目前，这并非如此。

6.1.2创建可用、可靠的元数据

接受项目采访的学者从完全不需要描述性元数据，只要他们能找到书架号，到那些可能会根据数字副本的描述来选择是否亲自访查手稿或书籍。值得注意的是，接受项目采访的学者都对他们在寻找的材料有相当的经验，因此，更随意或缺乏经验的研究人员可能会需要更多的元数据。为了适应这些用户，图书馆应尽可能地显示描述性元数据，甚至包括所有可能在印刷目录中找到的内容。
important than the amount of metadata, however, is its accuracy. Libraries should of course make
every effort to ensure the accuracy of digitized metadata, but they must also ensure that a
mechanism is in place to update the metadata in response to scholarly feedback or new discoveries.
In order to avoid errors perpetuating themselves, changes must be reflected across all digital
instances of the metadata in question. Linked data may be a natural choice here, as changes can be
made in one place and then simply linked to across catalogues and interfaces. Libraries should also
indicate the source of the metadata, so that scholars can evaluate its accuracy for themselves (as,
for example, a date taken from a 19th-century catalogue may have been contradicted by recent
scholarship). A more collaborative and iterative cataloguing system might also be worth
investment, one that would accommodate multiple values for each field, display degrees of
certainty for each, and allow users to contribute their own conclusions to the publicly displayed
catalogue. Whether or not such a system would be widely practical, it would at least generate a
useful picture of the volume and character of the ongoing scholarly dialogue about metadata.

6.1.3 Creating images that facilitate scholarship

Perhaps unsurprisingly, given the resources currently being devoted to photography in
digitization projects, participants in this study were generally satisfied with the quality of recent
digitized images. The basic requirements are deep zoom, color fidelity (ideally demonstrable
through a photographed colorchecker), and presentation as a paged sequence. Bindings should be
photographed where possible. Especially with difficult objects such as palimpsests and damaged or
decorated books or manuscripts, hyperspectral imaging is a useful addition, as it creates a digital
surrogate that not only mimics but enhances the experience of viewing the original. Three-
dimensional imaging and recently developed interfaces that allow users to switch between
differently lit or angled images will most likely mitigate the essential and deceptive flatness of
digitized objects. It is recommended that libraries work with scholars to choose pilot projects for
these technologies, as they will not be equally useful in every circumstance.

Given the popularity of downloadable images among participants, libraries should make it a
priority to allow downloading of reasonably high-resolution copies of digitized images. While
libraries may argue that the images are free to access online and that high-resolution downloads
would be cumbersomely large, it is clear that scholars want high-resolution downloads in many
cases and are willing to put up with large file sizes. It would be ideal, however, to offer multiple size
options for downloads, to accommodate scholars with different research needs and hardware
capabilities.
In addition to providing downloadable images (or instead of doing so, in the case of licensing restrictions), libraries should work to ensure that scholars can undertake basic research tasks using the images provided in the online viewing platform. Many participants reported downloading images in order to more easily juxtapose and edit them, so libraries should use viewing software that allows side-by-side comparison of images and manipulation of color, contrast, brightness and other image qualities. Given that software already exists to do all of these things (Mirador), this amounts to a recommendation that libraries make their holdings IIIF-compliant and educate their users about what they can do with IIIF-compliant materials. In any case, juxtaposition and editing are two of the main advantages that a digital surrogate has over an original, and libraries should be delivering those advantages to their users.

6.1.4 Dissemination and instruction

With few exceptions, the scholars interviewed in this study had received no training or instruction in locating and using digitized special collections materials. Their substantial arsenals of discovery techniques were largely the result of personal experience and experimentation. While the interviewed scholars reported passing their skills on to their students, there is a lack of library-administered training. Libraries and academic departments may offer digital information literacy training (see for example the popular training programs at Lisbon University (Sanches 2016)), but these do not cover the skills required for digitized special collections discovery in particular. Even the exemplary online paleography courses recently given by Stanford University, which engaged directly with the advantages and disadvantages of consulting digitized manuscripts, devoted no time to the specific skill of resource discovery. Given that several participants viewed their expertise in this area as a core research skill, we must ask ourselves why this skill is not being taught. An effort should be made to integrate training in the use of digitized special collections into taught postgraduate programs, as at any later point researchers are likely to dismiss training because they believe they should already know how to find things. Training sessions are likely to be useful for library staff as well, as students examining digital resources with fresh eyes will flag up pain points.

Training at this early point in scholars’ careers is not sufficient, however, as technology changes. The problem of educating researchers and later-career academics is significant and by no means specific to this area, but it is vital that researchers be made aware of tools like Mirador that could vastly simplify and enhance their work. Given scholars’ tendency to learn things from

http://online.stanford.edu/course/digging-deeper-form-and-function-manuscripts
colleagues, a good strategy may be to educate a few keen scholars about these tools and trust that word will spread. Where possible, searching and browsing recommendations and guides to the structure of digital collections should be readily accessible from the digital collection pages. Given the willingness of a majority of questionnaire respondents to receive digital library updates by email, an email newsletter may also be a good way to propagate advice for using digital resources. Blog posts devoted to discovery tips (see for example British Library 2015) are also likely to be useful to professional researchers, who will be able to bookmark helpful posts for later reference. Perhaps the most important thing, however, is to integrate scholarly input into digitization projects. Scholars must be given the opportunity to weigh in on interface design and metadata, among other things, to ensure both that the projects are well-conceived and that people in the research community are aware of them.

6.2 LIMITATIONS OF THIS RESEARCH

The limited number and range of participants in this research, as discussed above, is one principal limitation that should be taken into account in forming conclusions. The method of recruiting participants to both the questionnaire and interview portions of the study favored researchers in close contact with librarians and other scholars, either in person, by email or by social media. Furthermore, the participants in the interview portion worked predominantly with manuscripts, and no scholars from outside the University of Oxford were interviewed in depth. Neither were scholars working with American, East Asian or sub-Saharan African materials. If resource discovery techniques or expectations for digitized materials vary from region to region, this variance is not reflected in the study. Moreover, problems such as poor image quality, slow or network-taxing websites, and foreign-language interfaces are most likely underrepresented in this study, as Oxford scholars have access to better resources than many. For the same reason, this study may underestimate the importance of digitized materials in research, as Oxford scholars tend to have access to excellent collections at the Bodleian Libraries, and therefore do not need to rely on digital surrogates in as many cases as, for example, an American scholar might. Libraries in other parts of the world or with a high proportion of scholars in an underrepresented area would be advised to conduct their own studies. In general, working closely with collection curators and local researchers is most likely the best way for a digital team to ensure that any digitization project is tailored to the library’s collection and community.

Another significant limitation is the fact that all accounts of resource discovery methods and behaviors relied on individual testimony. Scholars were asked to describe their own views and
habits, but they were not observed conducting any research tasks. This was partly because, in the researcher's view, scholars' perceptions are worthy of study, and partly because even an in-person observation session, where the participant carried out discovery tasks as instructed, would still be far removed from the actual experience of looking for a digitized object in the course of normal research. This belief was borne out by the substance of the interviews, in which scholars reported a wide range of discovery techniques, each specific to a particular situation. Those further investigating this subject more narrowly, however, or planning to implement these recommendations, would do best to include a component of practical observation.

6.3 Future directions

Given the shortage of existing research into the subject of resource discovery among digitized special collections, this study was intended to provide a set of preliminary recommendations and crucial areas for development given the deficiencies pointed up by current research practice. Beyond the most basic elements, however (e.g., shelfmark searching among manuscript scholars), the needs of, and resources available to, researchers vary widely across career stages, research stages and areas of scholarship. Additional variation most likely occurs from one institution or country to another. In order to create collections that are accessible and useful to as many people as possible, more research is needed into the behavior and requirements of more specific user groups. Once more research has been done, libraries will be able to—and will need to—build resources that accommodate multiple modes of discovery and use, rather than simply catering to the highest-profile or most easily-satisfied demographic.
7 APPENDICES

7.1 APPENDIX A: QUESTIONNAIRE
Calling all humanists: how do you find and use digitized rare books and manuscripts?

This short survey, by Emma Stanford of Bodleian Digital Library Systems and Services, is intended to help digital librarians understand how researchers discover and access digitized rare books and manuscripts. Understanding how this process works will help us to create useful resources and to place them where researchers can find them. The survey is part of Emma's research for her master's dissertation in library science at City University London; it is being undertaken with the support of the Bodleian Libraries, but NOT as a Bodleian Libraries project. You can learn more about participating in this survey at goo.gl/ftQx8W.

1. How would you describe yourself?
   *Mark only one oval.*
   - Undergraduate student
   - Postgraduate or graduate student
   - Lecturer
   - Professor
   - Alt-ac or independent researcher
   - Other: .................................................................

2. What is your field of research?
   *Check all that apply.*
   - Languages
   - Literature
   - Paleography
   - Codicology
   - History
   - Sociology or anthropology
   - Other:

3. What type of materials do you work with?
   *Check all that apply.*
   - Printed books
   - Western manuscripts
   - Eastern manuscripts
   - Maps
   - Archival materials
   - Other: .................................................................
4. Approximately how often do you consult digitized rare books or manuscripts in the course of your research?

Mark only one oval.

☐ At least once a week
☐ At least once a month
☐ Several times a year
☐ Every year or so
☐ Less than once every five years

5. How important are digitized rare books and manuscripts to your research?

Mark only one oval.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unimportant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. How did you learn about the digitized rare books or manuscripts you have used in your research?

Check all that apply.

☐ Colleague or advisor
☐ Librarian
☐ Library website
☐ Department website
☐ Published research
☐ Mailing list or web forum
☐ Online list or directory
☐ Social media
☐ Other:

7. How would you prefer to find out about digitized rare books and manuscripts?

Mark only one oval.

☐ Colleague or advisor
☐ Librarian
☐ Library website
☐ Department website
☐ Monograph or article
☐ Mailing list or web forum
☐ Online list or directory
☐ Social media
☐ Other:............................................................................................................................
8. Why is that?

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

9. Name up to four institutions or websites you think are particularly useful for finding and accessing digitized rare books or manuscripts, and explain in brief what you prefer about them.

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................
10. Do you use any of the following for access to digitized rare books or manuscripts?

Check all that apply.

- Bayerische Staatsbibliothek
- Biblioteca Apostolica Vaticana
- Biblioteca Real de la Academia
- Bodleian Library
- British Library
- Cambridge University
- Digital Scriptorium
- DMMmaps
- earlymedievalmonasticism.org
- e-codices
- EEBO-TCP
- Europeana
- Gallica or the Bibliothèque nationale de France
- Google Books
- Harvard University
- Hathi Trust
- Incunable Short Title Catalogue
- National Library of Australia
- National Library of Israel
- National Library of Scotland
- National Library of Wales
- New York Public Library
- Pinakes
- Rijksmuseum
- Staatsbibliothek zu Berlin
- Stanford University
- Trinity College Dublin
- Universitätsbibliothek Heidelberg
- University of Leuven
- Walters Art Museum
- Wellcome Library
- Yale University
- Other: ............................................................

https://docs.google.com/forms/d/1Ungj5FcQV3s42IsawJqWuvfA0pP4562bUDobF39-A/edit
11. What do you use digitized rare books and manuscripts for?
Check all that apply.

☐ Browsing
☐ Early-stage research
☐ In-depth analysis
☐ Comparing multiple items
☐ Transcription
☐ Situations where it's not possible to consult the original object
☐ Computational analysis
☐ Image analysis
☐ Analysis of the book or manuscript as a physical object
☐ Other:

12. Is there anything that you would like to be able to use digitized rare books and manuscripts for, but can't? Why not?

........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................

13. If you have used a digitized rare book or manuscript in your research, do you generally cite the digitized version or the original?
Mark only one oval.

☐ The original, because that's what is customary in my field
☐ The original, because I'm not sure how to cite the digitized version
☐ The original, because it's unimportant whether I also used a digitized version
☐ The digitized version
☐ Both
☐ Other: ..................................................................................................................................
14. When searching for digitized rare books or manuscripts, what terms would you like to be able to search for?
   Check all that apply.
   - Institution
   - Shelfmark
   - Title
   - Author
   - Catalogue number
   - Subject
   - Date
   - Language
   - Place of origin
   - Provenance
   - Marginalia or annotations
   - Decoration
   - Licensing or reusability information
   - Download options
   - Type of image (e.g., color or black-and-white)
   - Other: .................................................................

15. What features of digitized rare books and manuscripts are most important to you?
   Check all that apply.
   - Searchable text
   - Searchable metadata
   - Zooming in to a high level of detail
   - Portability and remote access
   - The ability to make notes and annotations
   - Side-by-side comparison of multiple items
   - The ability to adjust image qualities such as contrast and brightness
   - Preservation
   - High-quality image downloads
   - Downloadable PDFs
   - Other: .................................................................

16. Do you prefer to download digitized images or view them online?
   Mark only one oval.
   - Download
   - View online
17. Have you ever built elements of your research around whether or not certain items have been digitized, or do you know of others who do this? 
Mark only one oval.

☐ Yes
☐ No

18. Have you ever found out only late in your research, or after your research has been finished, that a relevant rare book or manuscript has been digitized? 
Mark only one oval.

☐ Yes
☐ No

19. Would you subscribe to or make use of any of the following ways of announcing newly digitized items? 
Check all that apply.

☐ Email newsletter
☐ RSS feed
☐ Library blog
☐ Twitter
☐ Facebook
☐ Option to browse new items specifically

20. Please use this space for any other comments about your experience and preferences regarding digitized rare books and manuscripts.

...........................................................................................................
...........................................................................................................
...........................................................................................................
...........................................................................................................

Powered by

Google Forms
7.2 APPENDIX B: QUESTIONNAIRE RESPONSES
<table>
<thead>
<tr>
<th>Timezone</th>
<th>How would you describe yourself?</th>
<th>What type of research do you work with?</th>
<th>How important is digitized scholarly research in your research?</th>
<th>How long do you spend looking at digitized rare books and manuscripts?</th>
<th>Why so? (Please explain)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13/05/16 12:59 PM</td>
<td>PhD student</td>
<td>Palaeography, Bibliography, Medieval manuscripts</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>13/05/16 1:05 PM</td>
<td>Professor</td>
<td>History, Library Science</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Graduate student</td>
<td>Literature</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Graduate student</td>
<td>Language, Literature</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Undergraduate student</td>
<td>Literature</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Professor</td>
<td>Language, Literature, Palaeography</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Independent researcher</td>
<td>Oral history, Library science</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Independent researcher</td>
<td>History, Religion</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Graduate student</td>
<td>Languages, History</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
<tr>
<td>15/05/16 3:30 PM</td>
<td>Graduate student</td>
<td>Languages, Palaeography</td>
<td>At least once a week</td>
<td>In order to access the digitized version for research and comparison</td>
<td>Because I don't have access to the original, but I would like to be able to get access to the digitized version to compare and contrast with the original.</td>
</tr>
</tbody>
</table>

**Summary:**

- Digitized scientific research is important in various fields, with a focus on accessing original material for comparison.
- The duration of access varies from at least once a week to more frequent use depending on research needs.
- The reasons for using digitized content range from comparing with original material to the need for specific images or versions.

---

**Digitized Rare Books and Manuscripts:**

- **Digitized Versions:** EEBO-TCP, Google Books, HathiTrust, Archive.org, GoogleBooks.
- **Digitized Projects:** Digital Scriptorium, DMMmaps, e-Scriptorium.
<table>
<thead>
<tr>
<th>Researcher Type</th>
<th>Institution</th>
<th>Materials Requested</th>
<th>Frequency of Use</th>
<th>How Important are Digitized Materials Used in Your Research?</th>
<th>What Type of Materials Do You Prefer to Use (Digitized, Original, Both)?</th>
<th>What are Your Reasons for Using Digitized Materials?</th>
<th>How Important Are the Digitized Rare Books or Manuscripts Used in Your Research?</th>
<th>Do You Prefer to View the Digitized Rare Books or Manuscripts, or the Original?</th>
<th>Is the Digitized Version or the Manuscript in Your Field the Preferred?</th>
<th>What Are Your Main Areas of Research?</th>
<th>Additional Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student in Early-stage research</td>
<td>Library</td>
<td>Manuscripts, Archives, Printed books</td>
<td>At least once a week</td>
<td>5</td>
<td>Both</td>
<td>Jeremy</td>
<td>Both</td>
<td>The original, because it's customary in my field</td>
<td>Both</td>
<td>Both</td>
<td>5</td>
</tr>
<tr>
<td>Institution, Title, Author</td>
<td>Name up to four institutions or websites you would like to be able to consult</td>
<td>Possible to consult the original object?</td>
<td>Open digital access?</td>
<td>Are the digitized rare books or manuscripts most useful?</td>
<td>Please describe any situations where it's not possible to consult the original object.</td>
<td>Digital Scriptorium</td>
<td>Incunable Short Title Catalogue, TCP, Google Books, Harvard College Dublin, Wellcome</td>
<td>Download options</td>
<td>Download</td>
<td>Image analysis, Maps to original object</td>
<td>Possible to consult the original object?</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>---------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>-------------------------------------------</td>
<td>----------------</td>
<td>--------</td>
<td>-----------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Harry Ransom Center (UT Austin)</td>
<td>Digital Scriptorium, British Library, Library of Congress, Bodleian Library, British Library</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
| Topic | How often do you use digitized manuscripts or rare books? | Libraries, Language, Philology, Manuscripts, Digital resources (e.g. microfilm, CD, online resources, etc.) | Where do you use digitized manuscripts or rare books? | Do you prefer to download images, PDFs, or to access the original object? | What features of digitized manuscripts or rare books are important to you? | Would you be more likely to use digitized manuscripts or rare books in your research if they were more easily accessible? | Have you ever built your own digital repository of digitized manuscripts or rare books? | Do you use any of the following institutions or websites you listed? | What features of digitized manuscripts or rare books would you like to be able to search for? | Would you make use of any of the above options if digitized manuscripts or rare books were more easily accessible? | Would you be more likely to use digitized manuscripts or rare books in your research if they were more easily accessible? | Name up to four institutions or websites you think you might use or make use of any of the above options if digitized manuscripts or rare books were more easily accessible? | How often do you use digitized rare books or manuscripts because that's what is customary in your field of research? | Do you think the digitized version is always a useful substitute for the original object? | Do you think the digitized version is a useful substitute for the original object? | What would make your research about 10x easier? | How do you find out about digitized rare books or manuscripts because that's what is customary in your field of research? | Library website | \[
\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|} \\
\hline
| Topic | How often do you use digitized manuscripts or rare books? | Libraries, Language, Philology, Manuscripts, Digital resources (e.g. microfilm, CD, online resources, etc.) | Where do you use digitized manuscripts or rare books? | Do you prefer to download images, PDFs, or to access the original object? | What features of digitized manuscripts or rare books are important to you? | Would you be more likely to use digitized manuscripts or rare books in your research if they were more easily accessible? | Have you ever built your own digital repository of digitized manuscripts or rare books? | Do you use any of the following institutions or websites you listed? | What features of digitized manuscripts or rare books would you like to be able to search for? | Would you make use of any of the above options if digitized manuscripts or rare books were more easily accessible? | Would you be more likely to use digitized manuscripts or rare books in your research if they were more easily accessible? | Name up to four institutions or websites you think you might use or make use of any of the above options if digitized manuscripts or rare books were more easily accessible? | How often do you use digitized rare books or manuscripts because that's what is customary in your field of research? | Do you think the digitized version is always a useful substitute for the original object? | Do you think the digitized version is a useful substitute for the original object? | What would make your research about 10x easier? | How do you find out about digitized rare books or manuscripts because that's what is customary in your field of research? | Library website |
\end{array}
\]
**Question 1:** How would you describe your research?

**Question 2:** What type of materials do you generally work with?

**Question 3:** Do you use any of the following online services for your research?

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library website, Mailing list or directory</td>
</tr>
</tbody>
</table>

**Question 4:** Have you ever found out about digitized or rare materials that lists all digitized MSS and provides easy searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 5:** Have you ever built a searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 6:** What type of materials do you usually work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 7:** What would you like to have in your online searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-quality image downloads</td>
</tr>
</tbody>
</table>

**Question 8:** Do you prefer to consult the original or the digitized version or the digitized rare book or manuscript as a physical object?

**Question 9:** When digitizing, what is your primary concern?

<table>
<thead>
<tr>
<th>Concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservation</td>
</tr>
</tbody>
</table>

**Question 10:** What would you like to have in your Online list or directory, Social Librarian, Library website, Published research?

<table>
<thead>
<tr>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotations</td>
</tr>
</tbody>
</table>

**Question 11:** When digitizing, what is your primary concern?

<table>
<thead>
<tr>
<th>Concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservation</td>
</tr>
</tbody>
</table>

**Question 12:** What type of materials do you typically work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 13:** What type of materials do you usually work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 14:** Have you ever built a searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 15:** Have you ever found out about digitized or rare materials that lists all digitized MSS and provides easy searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 16:** What type of materials do you usually work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 17:** What type of materials do you typically work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 18:** Have you ever built a searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 19:** Have you ever found out about digitized or rare materials that lists all digitized MSS and provides easy searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 20:** What type of materials do you usually work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 21:** What type of materials do you typically work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 22:** Have you ever built a searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 23:** Have you ever found out about digitized or rare materials that lists all digitized MSS and provides easy searchability, or after your research has been published? Please check any that apply:

<table>
<thead>
<tr>
<th>Online Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online list or directory</td>
</tr>
</tbody>
</table>

**Question 24:** What type of materials do you usually work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
</tbody>
</table>

**Question 25:** What type of materials do you typically work with?

<table>
<thead>
<tr>
<th>Material Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscripts</td>
</tr>
<tr>
<td>Timestamp</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>9/12/2016 16:53:50</td>
</tr>
<tr>
<td>9/13/2016 11:26:51</td>
</tr>
<tr>
<td>9/16/2016 13:06:39</td>
</tr>
<tr>
<td>10/6/2016 22:08:12</td>
</tr>
</tbody>
</table>

**Undergraduate student researcher:**
- Social media
- Twitter, Facebook
- Library blog, Twitter, Facebook
- Library website

**Alt-ac or independent researcher:**
- Social media
- Twitter, Facebook
- Library website, Published media

**Postgraduate or graduate student:**
- Social media
- Twitter, Facebook
- Library website

The original, because I missed something I think is important.

**Why is this important to you?**
- To search for?
- Marginalia or annotations
- Date, Place of origin, Title, Author, Subject, Language
- Catalogue number, Shelfmark, Institution, Provenance
- Downloads, Licensing or reusability information, Downloadable PDFs
- Portability and remote access, Side-by-side comparison of multiple items, Searchable metadata, Searchable text
- Zooming in to a high level of detail, High-quality scans

**If you have access to a digitized version of your research, what is the main benefit of using a digitized version compared to the original?**
- It gives the opportunity to learn casually about things
- It allows you to see what you might have missed
- It provides more information than the original

**What are the most important features of a good online directory?**
- High-quality scans
- Searchable text
- Searchable metadata
- side-by-side comparison of multiple items

**Would you recommend this to other researchers?**
- Yes

**Do you use any specific online directories?**
- Google Books
- National Library of Wales
- Gallica
- Biblioteca Apostolica Vaticana
- BNF
- Bodleian
- British Library
- BNF
- Staatsbibliothek zu Berlin
- Cambrige University, e-codices, Gallica or the Biblioteca Real de la Academia

**How do you consult digitized manuscripts?**
- At least once a week
- At least once a month
- Every year or so

**Approximately how often do you consult digitized rare books and manuscripts you have found?**
- At least once a week
- At least once a month
- Every year or so

**When did you last access the digitized version?**
- 2016-09-12 16:53:50
- 2016-09-13 11:26:51
- 2016-09-16 13:06:39
- 2016-10-06 22:08:12
- 2016-10-07 07:54:00
- 2016-10-07 07:16:18

**Timestamps:**
- 9/12/2016 16:53:50
- 9/13/2016 11:26:51
- 9/16/2016 13:06:39
- 10/6/2016 22:08:12
- 10/7/2016 07:54:00
- 10/7/2016 07:16:18
| Timestamp | How would you describe your field? | What type of materials do you work with? | How important is digitized rare books or manuscripts to your research? | What do you use for digitized rare books and manuscripts? | Why do you use it? | How important is it for you to be able to access digitized rare books or manuscripts? | If you have ever used digitized rare books or manuscripts, is there anything that you couldn't do with the digitized version or the original object? | What do you use digitized rare books and manuscripts for, but can't do if they're not digitized? | Why is that? | Is there anything that you can do with digitized versions of rare books or manuscripts that you can't do with the original object? | Name up to four institutions or websites you would subscribe to for access to digitized rare books or manuscripts | Do you subscribe to any of the above institutions or websites? | Would you subscribe to more? How many? | Why or why not? |
|-----------|----------------------------------|----------------------------------------|--------------------------------------------------|--------------------------------------------------|---------------|-----------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| 10/6/2016 21:36:16 | Postgraduate in graduate studies | History | Approximately how often do you consult digitized rare books and manuscripts? | What do you use digitized rare books and manuscripts for? | Why do you use it? | How important is it for you to be able to access digitized rare books or manuscripts? | If you have ever used digitized rare books or manuscripts, is there anything that you couldn't do with the digitized version or the original object? | What do you use digitized rare books and manuscripts for, but can't do if they're not digitized? | Why is that? | Is there anything that you can do with digitized versions of rare books or manuscripts that you can't do with the original object? | Name up to four institutions or websites you would subscribe to for access to digitized rare books or manuscripts | Do you subscribe to any of the above institutions or websites? | Would you subscribe to more? How many? | Why or why not? |
7.3 **Appendix C: Interview Transcripts**

Transcript portions used in analysis have been divided roughly into categories below:

### 7.3.1 Role of digital surrogates in research

“Absolutely essential” role played by digitized manuscripts or microfilm copies or “A digitized form of microfilm they created way back when. It’s pretty rare that they’ll go through and do a nice new digitization to be honest, but occasionally I’ll ask them to pop it up online.”

“I’ve rarely gone into the library to call it up and consult it physically—usually when I need to check measurements or put it under ultraviolet light because I think there’s something which is not being picked up well, or as well as it might be, on the digital image.”

“Without exaggeration about 99% digital…. When I was preparing the bid for the…project I had it out then and it was one of five MSS I was looking at for about 1 or 2 days in the library. I’ve had it out again now and that’s the first time I’ve physically pulled it for inspection, and I’ve spent about three hours physically looking at it, rechecking the binding, chasing things up with the UV light, which didn’t work… and in the meantime I’ve transcribed the entire thing, looked at it with digital copies of comparable manuscripts up on screen” “99 to 1 ratio is probably actually understanding it, it really is that much of a bias”

“Knowing stuff is there would encourage me to look at it because you know it’s going to make life a lot easier. Conversely there are some things I’ll be moving onto after I’ve finished this article...are problematized by the fact that I don’t have decent access to images and I know I’m going to have to go to the libraries to get something that would be a firm basis for producing research of which I’m confident. Basically I’ve got microfilms but these are all black and white and I’m really not happy with them in terms of saying anything with confidence about the manuscripts”.

“If it was a question of checking a reference that I couldn’t do online, I would take a note and then at some point go to the OI and check the physical copy. I wouldn’t say it holds up—you can always postpone your referencing. And I think the project as a whole has already got the basic resources it needs to carry on. Just some specific resources might not be available, but they can be found elsewhere. Certainly—if it’s about a certain poet, there may be six or eight editions of his work, and some poems may not be in one, they may be in others, so if you can’t find the one where this poem’s cited online then you have to go and check the physical copy. So I wouldn’t say that would hold you up, but just postpone it.”

“I use the manuscripts obviously for source material. I usually work from the reproductions of them, which has been largely microfilm, I haven’t moved onto CDs because I collected my microfilm a long time ago, and one of the projects I’m involved with has 2 MSS which have got superb illustrations in which have been very difficult to access, and now that they’ve been digitized—one’s in the Vatican, one’s in the Bn in Paris—then I consult those constantly, and that has been a huge boon”

“All my projects have got reasonably advanced, so—I’m spending actually a lot of time with digitized versions of MSS, which is—that’s not quite true, it’s photocopied from a microfilm, which is rather vile, but when I’m looking at the illustrated manuscripts I’m going to the digital site constantly for that. And otherwise I’m using digitized catalogues, old editions, 19th-century printed texts, that is a huge boon actually, that one can consult things online and very often download them.”

“With printed books the problem is flipping through and sticking your finger in one page and going to another page. With a manuscript you’re normally working much more slowly and precisely, and you’re not allowed to flip pages if you have the thing in front of you, so. And of course a huge advantage with digitized images is that you can zoom and really get into the detail of the hand and things, which magnifying glasses don’t always do.”

Don’t base plans around digitization “because in the past, if one had to travel, you would always, if you were going to work on something extensively, get a microfilm because you could never take sufficient notes on the
spot to be quite certain, so you’d always want a record of it, and that meant you didn’t actually have to be over in a strange library for a long period of time. It would certainly—now it just makes life easier if one can access DMSS, but I don’t know that it would have affected me in my choice of subjects.

“hard to put a number” on digitized vs. physical; his PhD thesis was “deliberately built around seeing manuscripts in person”—marked digital but not physically viewed books with an asterisk; admits that it “maybe raises some questions” that he didn’t mark the ones he viewed digitally and then physically (thesis was on ruling etc.; “even if there is a ruler included in the image you feel uncomfortable” using it; also weighing books in medieval bindings); “less than 5% of the manuscripts that featured in any significant way” were consulted digitally; for current project, though, “maybe half at the moment, and possibly more than half by time consulting MSS”—using “some sort of digitized surrogate” where he’s photographed them himself; “I really only need to transcribe the text, and then go back and check the manuscript in person to make sure I’ve got it and there’s no oddities or anything, and do some codicological work, but I don’t need to do all my transcription in the Bodleian. It’s more efficient for me to take 20 photographs and then sit here with my photographs and work from those. So probably more of my time than not, maybe 60 or 70%, even with Bodleian manuscripts.”

“how I understand if a MS is interesting to me. Sometimes the catalogues are very incomplete, especially when it comes to paleographic data, so just by reading the catalogue I cannot be 100% sure that this MS is in [ ] script and was produced in Andalus or Morocco in this particular period. I have to have colophon in front of me so I can read it and I can make sure that the data contained in the catalogue is correct. Usually most of these catalogues were written in the 19th century so they tend not to be too accurate. But sometimes certain MSS yes were copied in this time frame, so they tend not to be too accurate. But sometimes certain MSS yes were copied in this time frame, but somewhere else.”

“I have a lot of images of digitized MSS in my database that I normally use when I want to make sure of certain elements.”

“Of course if all these MSS were digitized probably someone would already have written my thesis. So it’s good to discover new things. So you ask for images, you are sent images, and you realize, my God, the catalogue entry says one thing, but this MS is actually completely different. This is an interesting discovery.”

looks at “the incipit of the text, the colophon...and then images of folios from the body of the text that give me an idea about the layout, the scripts, the treatment of the headings as opposed to the treatment of the main body. Sometimes MSS are copied by different hands, and in that case you have to ask for a lot of images because you want to understand how many hands were involved. But normally all I need is five or six good images from different parts of the MS, and that will do for a preliminary step, and then of course when I examine the MS in person then I look at the entire thing.”

“I have the stuff on my laptop in files, but I also tend to print it off. I have a color printer at home and I just spend a lot of time looking at these images and noticing certain things about them [how text and music patterns vary together or separately].”

never has to go see the original. “Thankfully not.”

role of DMSS: “Without them I’d have to spend a lot more time in Paris I suppose. Most of the witnesses are in Paris...and then there’s one in Oxford and then in various French regional libraries. So I guess I rely on digitizations or at least scans of microfilms for transcriptions but also to do a preliminary analysis—codicological analysis shall we say, and then I will hopefully get a chance to visit all of the witnesses in situ as well. So I guess that’s what I’m using digitizations for primarily.”

“Before designing the project I’d obviously look to see what was available. I knew that the bulk of the MSS were already available on Gallica, and I assumed, I guess, that I would be able to get my hands on—I knew I would have the budget to purchase microfilms or digitizations for the rest of them. So I guess I did do that
preliminary work in that I checked on Gallica what was available, but the majority of them were, so it wasn’t going to break the bank purchasing the rest of them.”

“Especially if I haven’t got a huge amount of time in situ--I’m thinking at the BnF, where they’re likely to limit the amount of time--it’s, yeah, I’m going to have to do at least some work or as much as I can from digitizations. Also to make the time I do have with the document as useful as possible.”

“I often use an image from a MS not just to give a sense of the material document, which I think students are often intrigued by, but also as a way of structuring a lecture”

do students work with digitized materials? “Not the students that I’ve taught I don’t think. In that I will often point them towards resources that they might want to follow up on, but in terms of setting work that requires them to consult digitized material, I don’t think I’ve ever done that. It’s always been ‘have a look and be fascinated by this’ rather than ‘to complete this essay you will’”

consults originals “almost always. There are a very few occasions where say, I haven’t been given access by a library, and if I am doing an in-depth study of a MS I will always attempt to see the original, or I will make it--yeah, that’s where I will say ‘I have not been able to consult these in person’”

“Oh definitely, yes. Depending on what exactly you’re looking for. If you’re trying to find ...and you know that the provenance of certain MSS is similar, you need to see the entire MS. Even binding can lead you to the provenance if it’s the original...So all these digital media are really important.”

“For stuff where it’s not copy-specific, I mean, they play a huge role now. So EEBO, some of the Folger resources, digitized early texts...that’s huge, and I use that kind of material probably every day, but there’s a definite part of my work which still or perhaps even more needs travel to look at specific copies.”

EEBO: “I’m generally working on a particular book that I want to look at and have in front of me on the screen. That’s one way. So I will call up that book and look through it, and see what it has to offer me.”

“I’m not just looking for texts, I’m looking for what MSS actually look like as material objects. So I often can’t do it with published texts, even when they are critical editions. I need to see the manuscript. The next best thing after seeing the MS is seeing the digitized version, which is often cheaper than traveling.”; “For my students access to digital material is essential because they have almost no other way of getting to know what MSS look like”; in the mid-90s was using exclusively print facsimiles; more difficult to access “in the provinces” than in London, as they are “almost impossible to obtain via interlibrary loans. The rule is no. Too valuable.”

7.3.2 Resource discovery

“Either I’m looking for a very particular fragment, which usually doesn’t occur on a library website, or rather I’m not aware that it has been digitized, or when I’m looking for larger manuscripts then I usually start on Google and it appears as a link from a university website or library website”

“I think it depends on how deeply they are engrossed in the whole question of working with manuscripts. If it’s someone who actually does a critical edition of something, then they probably will start with a print catalogue that they have, and work their way out from that. I think in these niche things where many aspects are still in written print form, you rely on people telling you where to look. And that might be a librarian or it might be a colleague, and the main exchange of information probably occurs by email.”
"The languages I work with tend to either have manuscripts in weird locations or in let’s say very secure locations, like the Vatican Library, where people are reluctant to let you look at things you might want to look at. So digitization would be a help, but again, it hasn’t come to that yet.”

Google: “If I’m looking for something particular, a manuscript or a fragment—then it’s a pain. First of all because I don’t know whether it has been digitized, so you start with looking for the actual manuscript, the shelfmark or whatever it might be, and then for a more general term, and I would say in 60% of cases or something like that, in the majority of cases, I actually get lucky and find something close to what I wanted. And in those cases it’s usually been quite easy. In other cases you try your best and see whether it’s on Archive or on a university website or whether some Russian or Eastern European colleagues have more or less legal copies of things. So far I’m quite happy with how Googling has worked out. Then again, with the Armenian manuscripts, they’re all here, so I know where to go.”

“If I know what university or library [a manuscript] is held at, I would probably start with their website, because it’s going to be faster and less cumbersome to do it that way, and then yes, whether it’s an integrated system where the regular library holdings and the manuscripts holdings are listed together or separated I think doesn’t really matter as long as you can toggle between the two. But yes, searching would be the beginning, and searching for the title or shelfmark or person or date or content, then going through the search results, I think that would be very useful. If they are then also discoverable via a more global search engine I think that would be helpful too, but I doubt that would be the primary means for me to find something, unless I’m really desperate. If I’d read about it somewhere then I think I’d go to the place where it was listed at. If I was more generally searching for ‘Oh, are there any manuscripts on, I don’t know, medieval Armenian stones,’ then you wouldn’t go to a random university website unless this was the main place for Armenian holdings, because they might not have everything but they would have a little bit of most things. For discovering things a global search engine would be more immediately relevant.”

“If I knew what specific object I was looking for and where it was held, I would look there, on the institution website, and then use the resources provided there to go about finding it either by actually finding it or by contacting them. And I think only out of frustration after that would I go to Google to see whether I am just too stupid to find it via internal means”

“Google, pretty much. That’s turned up a load of things. And it’s amazing what you can find through unexpected means.... I was particularly interested in checking out the dimensions of a manuscript from the Vitez collection at Liege University, and I could not--I looked and looked and couldn’t find any indication that this had been digitized or anything, and then somehow by just browsing around and through some complete back door I then chanced upon Liege has actually digitized the whole lot of it, perfect one, downloadable in full quality, but it was so badly publicized that it was amazing when I found it” “I didn’t find it by searching for the name of the manuscript or going through the repository of the Liege library—I can’t remember how I got there in the end, but so that was by pretty much random browsing, I think in the end I was looking for something else and then somehow got crosslinked onto it. But usually you try the actual host library, so you go to the Bibliotheque nationale, they’ve digitized a lot of things, or British Library, because the apocalypse manuscripts are pretty there’s a good chance that they’re going to be there, and gradually as I’ve been working more and more of those have come online in a properly done modern color form. A lot of the Paris nationale scans are just microfilm”—which is fine for text but not for images

would Google the shelfmark, “check the repository itself, though that’s not always the best way to go, then other than that it comes down to emailing colleagues and sort of saying ‘Are you aware of anything?’”

“So this is the [Brussels] Royal Library 2.282 I think...I couldn’t find that anywhere.... Even though it’s not part of the same corpus of MSS as the one that I’m looking at at the moment it became apparent there was some affiliation which I wanted to check out, and just by chance reading an article on something unrelated, I then saw a reference that it had a list of contents”...a colleague published a description that mentioned the list of contents... “so then I had to contact him, to say, you know, I see you’ve done this work, do you have a
microfilm I can borrow? and he said no, but I did take some pictures, and just by chance one of the pictures he’d taken was of the apocalypse content list, so he sent that to me, and I could see the resource was of real interest, then I had to think how on earth can I get a copy of this. I looked at the means of acquiring the microfilm from the Brussels library, and the route that they prescribed was so tortuous—you know, you have to print out things and then complete it by hand and send there—I could see I would be tied up for a longer duration just trying to get the microfilm than I could possibly manage in terms of keeping the deadline for the article. So I then had just received a book which had been sent to me by the publisher of somebody who’d done a book on—an art historical book on apocalypses and I knew she’d written about this manuscript, so I wrote to the author directly and said, You’ve done a really really good book, how did you get hold of a copy of this microfilm? And she said well I’m not working in the area anymore, so do you want to borrow mine?

doesn’t like to publicly confess ignorance/incapability; uses social media for personal but not professional stuff usually; “I’ve failed in my scholarly rigor, can someone please help me out” is an admission of failure; expectation that he should be able to find things by himself; certain colleagues he wouldn’t hesitate to ask, though—but not as a first port of call; do the leg work first

for a 19thc European edition, would go to Meradoc; for an Arabic edition would try Archive.org, "but Archive.org has crowdsourced items and the transliteration system used for Arabic is sometimes incomprehensible, so what I would do, I would search for the book in Arabic and perhaps go through a discussion group which talks about this work and gives the link to the Archive.org copy. That’s simpler than trying to work out what transliteration system they might have used. There are a couple of scholarly discussion sites, and if you type in an Arabic book of logic by [] and you see these sites discuss it, if it’s available they normally would give the link.”

looking for MSS: “You wouldn’t generally find what you were looking for online. If it was something really specific like the project we’re working on we had to apply for the scans just from the imaging services at the libraries rather than find them online. We used one Bodleian MS as well, and that was a scan from microfilm, which is perfectly good for textual editing, because with the equipment they have now you can do quite good scans from microfilm. If it’s a really famous MS then yeah, you can usually find it, if it’s in Gallica at the BNF or other American or German libraries, but normally you wouldn’t really find what you were looking for.”

Google? "Well yeah, for the names of Arabic books, and then usually directed either to Arrchive.org or---so I google or search the Arabic title of the book and then put Archive.org beside it in English and it usually comes up."

colleagues: "Yeah, usually by email." Doesn’t use social media

no training received: "in fact I’m in charge of the online resources page for our project website. I add links to digitized editions of works and resources that are useful. I mean everybody shares what they have but I put most stuff up.” [just for project members]

“I quite enjoy looking for things. I quite enjoy the difficulty in finding things. It’s quite nice to you know find it in obscure places.” [laughs]

“I’d find the digital portal. These day there seems to be a digital portal of the digitized manuscripts, and as far as I can see, the thing to do now is you put in something--you’re looking for a chronicle and you’re not sure what its shelfmark is and you just type in the chronicle name and see if it crops up, and if it then says that it’s been digitized… One doesn’t always know the shelfmark…. I might go to a library and want to know, do you by any chance have any manuscripts of Herodotus here, and then, are they digitized? And start off with the name. Or I would actually go to Pinakes, which has increasingly detailed information.”

“With Bodley? Quite hard. There are two ways of getting to the Bodleian website, and one of them has got access to collections, and if you go to special collections, you’ll find something about digitized MSS. Somewhere one can find a heading saying DMSS but you have to hunt for it a bit. And then you get a series of
options, and if you're after the Greek MSS, it's more helpful to go down further and find the Polonsky website, where you can find a bit more information. So in otherwise it's not very coordinated.... By reading, I think--Initially I found myself going around in circles, and then not being able to refind my route. When I looked yesterday, that's when I'd been practicing a bit more [laughs] there was something on one of the Bodley main sites, there was a menu item that said special collections, and then catalogues. I had a faint feeling that something was being done with the 19th-century catalogues and they were being fed into a sort of a easier access system and I thought I'd found that at one point, but I've not been able to find it since."

“The Cox 1883 Quarto Catalogue of Greek MSS, which is the one that I’m constantly going to, I've discovered is actually slightly searchable, but it's very hit and miss because it's 19th century so broken type and therefore it doesn't always get recognized.”

“I'd search for the shelfmark, to actually find it. I'd search on the author or the title of the work. I would also be interested in possession, I mean previous owners, previous locations, I would then depending on how things were developing I'd be interested in paper or parchment, are there illuminations, then things like lines per page, in Greek MSS the rulings, all of which helps to group things together. And all of which in a full on catalogue would be there, but that's sort of secondary things. The first stuff one wants to hone in on is the author and work. Scribe. And place of possession, and the other stuff could be sort of further back in the system.”

spelling variations: “Oh yes.” [trial and error] “There are hideous spelling problems, and two major tools--the TLG and Pinakes--have their own idiocratic spelling systems, so you get used to just bashing away at it, you know, is it Latinate or Greek, is it a c or a k, and what have they decided to call it...in the TLG it's called Bellum troianum, and nobody has ever called the text that"

Greek characters: "no, because it’s not usually recognized by the browsers .... I would do more if it were more widely used."

Google: "No, except I was told by someone last week that there's a splendid way of getting to MSS in the BN by going to Google, which I slightly tried but not worked so I must get back to him--so no, I haven't. I use Google quite a lot as a shortcut, and it’s surprising what it will throw up, but I haven’t encountered it as a way of getting to DMSS."

Colleagues: “A lot. There’s a young man who’s doing a research project with me who is very up on all things digital, and he’s always producing endless material and he's my first port of call [laughs]”. email and in person.

“First of all I’d have a think about, is it part of an obvious collection? So, I think all the MMSS at John Rylands are digitized? So if it has a JR shelfmark I'll just say, well that must be digitized, so I'll go to the JR website, try to remember where their online thing is. But of course that's normally not the case. My next step would be to ask..my colleagues on the project, because they know the WB corpus better than I do and they might know it. Then I Google, usually for the library--if I remember the name of their digital site, so U of Pennsylvania has a site called Penn in Hand, I'd Google for that, and then I would search that site for it. I might in parallel try googling the shelfmark of the MS, because if you’re very very lucky, occasionally that will actually turn up a digitized version [laughs] My experience is that when things have been digitized, even if they're in relatively helpful repositories, there’s no immediate kind of straight-in from Google route into looking at them. I don’t know whether that’s--my instinct is that that’s a mixture of libraries not--that not being a priority for your library, and also Google not seeing that as a crucial thing that they should--and equally, to be fair, if it's available online for free, and the images load fairly quickly, and the interface isn’t a pig, I don’t feel like I’m hard done by if I had to search the library catalogue for five minutes to save myself from being on the train for three hours. I suppose the thing is that if it just popped up on Google immediately, that would make finding out whether they had it in the first place much easier. That would be intellectually satisfying.”
“at most once or twice I’ve had to email the library and say look, is this available in any form online or do I have to come myself...often a library will tell you that they have a repository of digitized MSS, but it’s --not everyone will give you a list of the MSS they’ve digitized. And I guess if you’re in the middle of digitizing things, that could be changing every week anyway, and it could be as simple as like a completely unformatted HTML file that you can ctrl-f file, I mean that would work, from my point of view, that would let me know--the Parker Library actually lets you browse by Mss shelfmark, and it’ just gives you a list of numbers of every MS they have digitized, which is of course every MS in that case, but even though--because that’s now in the history of DMSS quite an old repository, and they’re building a new one, or at least they have a thing on their website about building a sequel--but it’s interesting that they thought of that quite a long time ago”

“Often for me, when I’m coming to a library site, I know the shelfmarks of what I want to see, and I’m not going ‘do they have copies of x or y’ because I have an independent handlist of copies of--and I suppose also if you worked on 18thc letters or something, you might not be in that lucky position. You might be saying--you might need those search functions to say do you have any letters from someone named Walton between the year...and that’s not the kind of question I normally find myself asking because usually MMSS aren’t numerous enough that there’s just a mass of stuff that you have to search through, or if there are unknown MMSS they’re so unknown that there’s no hope”

“suppose I couldn’t find the MS, have I tried searching for--so often when an MS goes online--if a library hasn’t digitized that many MSS, and they digitize an MS, they often make a news like a blog post about it or something, and sometimes you can find it by searching for the announcement that it’s been put online, which is slightly lateral, but--so for example I might remember to, instead of searching for the shelfmark, should I search for “really exciting copy of X digitized by this library. And I’m glad that libraries do that, because Google is much better at searching news posts than searching databases of digitized images. And also, there’s an obvious--it’s good or libraries to be telling the public about what they do put online. But that’s one example of the kind of thing I might think of, or sometimes shelfmarks change over time, so I might realize the next day that I’ve been searching for the shelfmark it had before 1992, like an idiot. Or that the library might have catalogued it online under a different shelfmark to the one I use, or sometimes search functions within repositories are finicky about how you format. So for example when you order MSS digitally at the BL, you have to format them as major element of shelfmark, then MS for manuscript, then number, so for example Royal MS E 9. Most scholars working in English at least would conventionally write MS Royal E 9. So you just have to remember that as far as the BL is concerned ,that’s Royal MS E 9, in their online ordering thing. I think their online digitized image thing might be more forgiving actually. But not every library--I mean the BL’s a big outfit with more resources and not every library is going to be putting things online and thinking about have we accounted for the 5 different ways someone might format a search”

“And that’s a big problem for searchability, is just trying to work out how the shelfmarks will be thought of in the library that owns the book”

“For the kind of work I do, it would be really nice if there was just that page I mentioned, the kind of imaginary page where the lib would just say here’s a full list of every MMSS we’ve digitized, and you could ctrl-f, or there’s a search box at the top but you can ctrl-f for them. Ideally I guess it would be nice if those shelfmarks were then hyperlinks to a stable URL for the MS images, but I’d be perfectly happy to and bang that into a search interface. Just knowing that it was there would help my--but yeah, it would be nice if they linked immediately. And you could if you wanted also--you could consider if you were the library listing the MS material you’ve digitized by text, like if you had a really sophisticatedly well organized collection of digital images, where you could say this folio range in this manuscript and also this later bit are both scrappy copy of a particular work of John of Wales or something, and we can make it so that someone can click on John of Wales and get a list of all the different MS witnesses to any of J of W’s works, including duplicates in our repository”--“except that I’m not actually sure that would be good from my point of view. I mean it would be convenient for me, but I’m used to thinking of things in term of MS which contains texts in unique forms, and particularly for undergrads and beginning postgrads I think it’s quite useful for them to be made to remember
that it is a unified physical object which will not behave in the ways that you might want it to, that really
Troilus and Criseyde is a set of witnesses that disagree with each other in different objects in different kinds
of hands, and if you want to consult that evidence you can’t--the gold standard or looking at that evidence is
to have it in front of you physically, and you can't full-text search it on the Bodleian’s table in special
collections, you have to manually read it, which means learning to read the handwriting and so on, and there’s
no way to avoid that. And even if you look at it online, my understanding is that we’re nowhere near having
full transcriptions, so even if you look at it online it’s completely different to having a searchable concordance
text online. So yeah, I think there are advantages to presenting MSS just as MSS digitally, even though I have
hesitations of the way that things become so slavishly imitative of the original object”

“First of all I would visit the website to see if there are some DMSS available online... Sometimes just by
Googling the shelfmark of a manuscript I may come across interesting images, perhaps not legal images but
images that have been taken by researchers and put online without any right to do so”

he always checks: “For instance with the National Library in Paris, whenever I am interested in a MS I go on
Gallica and check. But they are never there [laughs]”

“It can be, like, MSS mentioned in blogs, travel blogs, like an American tourist that for one reason or another
has the chance to visit the library of this particular mosque in Morocco, which of course I would never be able
to go to because it’s beyond my means, and then this guy talks about seeing this MS in the library--but then if
it's realistic, you investigate further. If it’s unrealistic, you just discard it. Or there’s proto-scholarly literature
on Arabic MSS written by amateurs, sometimes seems out there, I don’t know, I haven't seen. Sometimes the
information is accurate, sometimes it’s not. Powerpoint presentations, all sorts of material that you can find
online. Sometimes you find stuff on Facebook.”

searching in Arabic is easier because you don’t have to deal with transliterations; “with Arabic there’s only
one way of writing the title of the book, but if you write in English there’s ten ways of writing it. So you may
actually be luckier if you look for a title in the original language.” SOLO’s transliteration is “peculiar”: has to
remember how it’s done every time; Fihrist’s is different (more scientifically accurate) than SOLO’s

“I spend a lot of time googling stuff”

“So at the moment I just bump into MSS when I read the footnotes of certain articles relating to aspects of
society or culture, and then they mention a MS, and then most of the time I already know the MS, sometimes
the MS comes from a godforsaken library somewhere that has no catalogue, so I couldn’t consult it, and then
in that case I have to contact someone who works in the library who can facilitate”

does contact Arabists at other institutions over email
catalogues for Arabic MSS “are very limited and very outdated” at many institutions. “The only institution that
I can think of that have a very good online catalogue for Arabic MSS in the NL in Paris. But all the other
libraries, most of them, they have PDFs of the printed editions of the catalogues, so what you do is go online
and download a PDF and then you leaf through the PDF as if it was a real book. You can’t really make a search,
you just have to read the whole thing or go to the index.”

“I probably--so I’m using a published book that’s got a list of all the contents of my MS. I would almost
immediately look at the MS page of my source just to check that what’s in the published source is exactly what
there is. And then I will make a table using the published book with the relevant folio numbers, and then I’ll
immediately go and hunt down the images, and basically download the individual images from Gallica and
stick them in a file”

when looking for specific literary texts, “I tend to go through Arlima or one of the other sites that will
aggregate MS listing for particular works of literature. And I’ve also done a few blog posts on my website
aggregating...sources for specific French poems”
finding things on Gallica: “I would go there if I knew what the call sign of the MS I wanted to look at was, and that’s often because I’ll come across it in something I’m reading. And then I’ll search in their general search. I actually filled in a questionnaire they sent me about how I use their site only the other day, and they said how often do you use the advanced search, and I was like, there’s an advanced search? Because I just use the simple search bar but because I put things in quotes, and I want manuscript francais 12784, the pages that then come up and then you can constrain the materials, if you select manuscript there’s usually only one result, and there’s your manuscript, that’s just two clicks. So I don’t need fancy searches.”

“I will always know the shelfmark. And the reason for that is because basically all these MSS have been inventoried, and all of the stuff I’m looking at pretty much has been at least described if not actually edited.”

“I can’t imagine a circumstance where I’d think, ooh, there must be a MS of such and such, I have to try and find it. It just wouldn’t happen.”

“I search for stuff in Google all the time.”

“It would be useful if there were a standardized way of referencing shelfmarks. You know, if you abbreviate them or not...finding stuff by shelfmark is really difficult. And musicology has—there’s RISM sigla for libraries and collections, but even musicology journals don’t always use it, which really annoys me. So getting a standardized way of talking about a physical manuscript is I think really important. But everyone has to agree on it and they just never will...[Arlima] just write them out in full, which is very long-winded and you’d never get away with it in an article, so then when you write an article you have to choose what citation style you’re using”—also there are multiple 19thC sigla for each MS; you have to start by knowing all these; “they’ve got different sigla depending on whether you’re talking about them in one context or another...so once you get into sigla you can’t tell what you’re talking about unless you have quite a lot of knowledge already. So in terms of discovering them you just have to know what you’re looking for.”

“Proper inventories of MSS—well, the closest you get is Arlima, but typically if you try and find out what’s in a MS that’s got some music in and some other stuff, RISM will tell you what music it’s got in it but won’t tell you anything about the other stuff, and that’s as if the other stuff has got nothing to do with the music, which I don’t believe at all. One of the big things I want to do with my MS is show how they’re collected, but the inventories are often very—they’ve got huge disciplinary hangups, so you either get inventories of music or you get inventories of the other stuff”

search process on Gallica: “I guess because I work on medieval French texts, Gallica is my go-to resource, so I would just go to the Gallica homepage and type in the classmark. And I guess I know that you might have to do that in a particular format, because I know from using other sites for example that it can be very difficult to search for a particular MS. And I think that’s partially true of Gallica as well but I know that I have to do Francoais plus the number, or NAF if it’s [], I know--I’m familiar with--as far as I know these rules aren’t written down [laughs] but--you know, so I’d go straight to the site, and I’d learned how to input a number and see what comes up. I suppose I might well—for nonGallica MSS I guess I would go to the llnstitute de Recherche e Destrutte de Texte, so they have lists of all their MSS which are microfilmed, so I know I can source material from them if it isn’t digitized. And then I guess for libraries I’m not so familiar with I might go to the--is it the DMMap site? That for me is a quick way to get to, if I don’t know what a library or an institution offers in the way of digitized MSS, that’s the best port of call I think. Or another one for medieval French is, I guess it’s more useful for printed bibliography but Arlima. It’s very incomplete as a resource but they will often list MS witnesses of a particular text, and they’ll have links not to the digitizations of that particular MS normally, but at least to the site that will offer you a digitization.”

“Hmm. I would probably stop there in most cases, or contact the specific library, or say if it was the BL I might think, ooh, I know their manuscript blog so I might think ooh I’ll search that and see if anything comes up as a recently--because often they publicize what’s just been made available, don’t they? So I might check that I suppose. I guess any recent scholarship on a text I might check to see if any information about digitizations is
put in that, but then that seems to be very incomplete. Well, I feel that people aren’t necessarily in the habit of always giving information for the digitizations that they’ve used in their scholarly matter.... If those sources aren’t complete yeah, I’d ask a librarian at the institution. That’s where I’d go.”

Colleagues? “I would generally assume probably incorrectly that we all have the same awareness of what is available. But especially if something hasn’t been digitized but I know someone has worked on a particular text or sort of domain, I might ask whether they know of any digitizations or if they have one. So yes, I do, but not as immediately as I would approach a librarian perhaps.” would do this by email, “or--yeah, I have occasionally tweeted, not recently, but I have tweeted for if anyone--but that's more to do with if anyone knows of any scholarly material or secondary criticism on a particular MS rather than inquiring about digitization specifically.”

“I think so, yeah, unless--I’m just thinking back to previous projects--yes, normally yeah, I’m looking for a specific MS. Maybe I’m searching for particular images, I don’t know. I’m just thinking of coming up with teaching material, perhaps, in which case I might be say looking for a particular type of text or a particular date or MSS within a particular period. So there are filters that I would use I think. But not so much with the current project.”

“The IHT (?) has the Jonas database, which is comprehensive for certain genres of medieval French literature, particularly pious material...so that’s a very very useful resource for me. And then that’s--yeah, I mean, I’m aware that that’s not comprehensive, but that’s often a first port of call, and then Arlima, I guess. So those are the sort of subject-specific databases I use very regularly.”

“There are two ways as far as I know for the Greek MSS. The first is Pinakes, which is a very fine database, a very important tool. And it’s not only the cataloguing of all the MSS that you can find there, but it is also the version and the hyperlink that are there. And there is also something from Princeton but I am not sure how updated it is.... Otherwise I search in the particular libraries I am aware of. Sometimes it’s not very easy. With Vienna, you cannot find particular permalinks for each MS. Or it used to be--I haven’t used it in six months or so--but it generates every time you are searching. But then you need to search within the library, within the catalogue, you have to eliminate it by the MSS, and then you find it. And sometimes--maybe they have changed it, I’m not sure--but it’s easier for me to find the permalink of Greek MSS from the BnF through Pinakes rather than by searching in the BnF’s website. Again you need to put grec or something like that.”

Ease of finding DMSS link: “Depends on the library itself. Sometimes yes sometimes no. But I think it’s kind of a growing business, so hopefully it’s getting better.”

Googling shelfmarks? “Yes.” Accurate? “Not always, because Google I think doesn’t search immediately to the deeper of its libraries’ collections. Sometimes it might be an OPAC or something like that, so it’s a good risk. I would rather go straight to the library or drop an email. And of course Pinakes--I think they give one shelfmark, maybe I’m wrong. And sometimes, the problem with this OLIM case [where shelfmarks have changed as MS changes institutions]” You find out “by yourself, in your searching. Or perhaps you find the most recent publication and hopefully the researcher knew this detail, otherwise you may regard it as a different manuscript. Sometimes this happens with Pinakes. They listed 2 different MSS as if they were different, but they are exactly the same, but they have different IDs.” contacted Pinakes to correct this and they fixed it. “They were good.”

would contact scholars to find materials by email. others post requests on Facebook, but he wouldn’t do it himself. Contacting directly is “more efficient, I would say.”

“Well I do, but not in a very--I don’t feel there are very good ways of doing that. So for example the way copy-specific material has been folded into catalogues is very different. There isn’t really a standard. I think libraries are developing standard protocols for how they record that, but they’re most of them not retroconverting their catalogues, so it’s not very easy to find--I mean you can find the material that’s been--it’s a bit random the way one might try and search for material I think. Because lots of libraries that have copy
specific material about their collections don’t have that as part of the electronic file you can look at from a distance. So they might have a folder of kind of print material or something in the library which says these are our books which all belonged to X person, you know, or whatever. But it’s not very easy to find that. I’ve talked to lots of libraries where that tended to be stuff where in their first sort of rush to try and digitize catalogues, they didn’t do that material because it wasn’t their priority. So yeah, provenance and ownership and those kind of things that I’m interested in and that lots of book people are interested in, that’s really difficult to do online."

“The Folger catalogue Hamnet does have copy specific information for their books, but that in itself is a very--I don’t know if you know that catalogue but it’s a very sort of--it was probably the start of the art kind of twenty years ago. SOLO has some. It depends a bit on the libraries, the contributing libraries.”

“Well, I don’t really know how I find out. [laughs] So, I mean I think one of the things quite a lot of people are interested in doing is trying to reconstruct--so for example there’s someone called Frances...who was an important provincial book collector in the middle of the 17th century, and she writes in her books, she’s got Shakespeare books and various things, and there’s been a sort of long term project to try to sort of reconstruct that library, because you know for example because the Shakespeare stuff, they’re books that people are always going to be looking at, so they see oh that’s got this woman’s name on it, they’re much more familiar books, but she must have had a ton of other you know religious books, sermons, that kind of stuff that it’s much less likely that people are going to go and look at, so we never find the fact that she also owned those. So that’s been quite a long project, just trying to find, going to libraries, so lots of people are trying to reconstruct ownership patterns just by going around to different libraries and saying do you have anything do you know anything about this signature, do you know anything about this person. So it’s still--I mean it is better, I think digital resources do make it better in that you can--but I mean what I do for that, which is something a lot of people do, is I just google, do lots of different kinds of searches and spellings, I keep doing that, and have an alert for these kind of names, and see if something comes up, because there doesn’t seem to be a kind of meta way of doing it.” results are usually “most often what comes up is book sales, because that’s the kind of stuff that changes most.”

“So if I knew what I was looking for, so say I was looking for what does the title page of the first printed Duchess of Malfi look like, I would go into EEBO here on my desktop, and I would search for Webster, because I know that the title search on EEBO is really, is actually really difficult to do somehow, because the spellings are kind of weird and it’s really hard to get a title search. So that’s what I would do. And then I would get about 80 hits I guess, and I would just kind of scan through them until I found what I wanted. So I could do a much more specific search I can see, but I very rarely would do. My experience--it’s probably my own mismanagement but my experience is that quite often it says there are no results, so it actually seems easier to do a broader search and sift through by eye rather than getting it to do that.”

“I would look at--actually interestingly for Shakespeare books there’s quite a lot of--well there’s a number of places that I would know to go, but for other early modern books I don’t--I think it’s quite hard to know where digitized collections are. So I might look at the Furnace collection in Pennsylvania, or I might look at the Folger digital collection. The Folger digital collection is all page openings, so they’re not in any way sort of collated as books, as digital objects. There’s no sense that you would turn the pages, it’s that here is one high-res jpeg that’s the first opening, and then here’s another one the second opening. So I would look at that. But interestingly I was just looking the other day for a copy because EEBO texts are not really great--well obviously they’re not great to print out, they’re certainly not great to print out and then photocopy, because the quality is not very good, and because one of the things I was trying to encourage my undergraduates to do was to think it would be fun to look at these texts, they don’t want some kind of blotchy thing they can’t even read. So I was looking for a Shepherd’s Calendar, a really well delicately printed book, and it was really hard to find a modern digital copy.”
"I think it’s also that I’m not aware of a project which has brought those kind of holdings together in a way that makes it—so there are digital editions, some of them quite good, but as a— I suppose I’m less interested in them actually, what I really want is a digital facsimile, a digital object that’s just the book."

old booklists and auction catalogues; “Quite a lot of 18th-century booksellers and auctioneers’ catalogues, which I don’t buy because they’re too expensive, are available on ECCO. I’ve found a way of isolating them, which is quite useful. By clicking boxes you can just turn up the book catalogues, which I’m quite excited about, and I’ve been able to—amusingly, as I am technologically incompetent—to show it to people who are much more technologically competent than I am” (unclick other boxes, click reference, use word catalogue, divide it into ten-year blocks)

"I heard about EEBO from a friend who was compiling a 3-volume catalogue of Elizabethan printed verse, and he came to Oxford on the basis that he would be able to look at the original editions in the library, and then he discovered that he could get it in his room in college. While he was staying as a visiting fellow he told me about it. So quite a lot of things I hear from other people. I’m not a great goer to conferences, and I’m technically pretty primitive, but there are things—most of the things I hear about I hear about from other people. Occasionally I read a review which mentions a website that’s helpful, but not often. And the thing that I was looking at this morning for example, this website [lead books]. I was talking to the woman who runs the website last night...so you meet someone who is working on these lead book...so the answer to the question is word of mouth. I don’t think there’s anywhere that tells me with my range of interests where I should be looking specifically."

finding on EEBO: “It incorporates the short title catalogue. I also use ESTC, which is wonderful. So I would know about the author or the title or have read those somewhere, so I’d call the book up either by title or by short title catalogue number. I have three sets of the short title catalogue [at different houses] so I’m very lucky to have those so I can work very quickly. But otherwise I search EEBO by word. I’m particular interested in for example the history of terms for printing in English, so I ask EEBO to give me early instances of a technical word like skeleton in printing, and it brings them all up and then I work through them decade by decade trying to find instances of where they’re using this word in a technical sense. So I do a lot of word searching, and when I’m editing someone’s edition of a Shakespeare play...the editor was saying Shakespeare made this word up according to the OED, and I was searching EEBO for that word and discovering that it had been used in lots of publications before Shakespeare, so I could say to the editor no you’re wrong, the OED is unrevised at this point, look on EEBO"

"I’ve done so much searching of EEBO and resources like that, I’ve worked out ways of making it more efficient and giving better results. And also quite a lot of work I’m interested in is attribution stuff, and there’s quite a lot of work in 16th and 17th century drama particularly on how to use EEBO”--searching TEI: "there are ways you can use the transcribed text in EEBO more efficiently, and books on attribution talk about this quite a bit. But on other occasions I’ve worked it out for myself."

strategies: “That you have to remember variant spellings for example, the variant spelling box can give false positives as it were. It says there’s nothing there, but if you try another spelling of the same word in another variant form it will bring up examples of it. So it’s knowing the limits of what’s available that produces better results, and also distrusting the results in the first place, and saying this word [ ] doesn’t occur before Shakespeare, well if you spell it slightly differently you can find it on EEBO. So that’s largely how I use it. But I’m not a—EEBO has TEI conferences at the Bodleian, I’ve never been to one of them and discussed it with anyone. It’s all home-baked [laughs]"

"So maybe I’m not particularly good at this, but I will often try and identify the library that holds it and then look for its digital library, digital collections...via the institution website. Sometimes I’m lazy and I just type in Bavarian State Library digital collections into Google rather than going to the state library first. And I’m afraid I’ve done that with the Bodleian as well, so I will type into Google Bodleian digital collections instead of going to the Bodleian website...it’s not always absolutely obvious. The problem comes then, you know, how they set
up--there are myriad different ways in which each library has set up their search mechanism. So with some you have to know the exact shelfmark, but it has to be typed in a particular correct exact format, so if there is no dot after manuscript, then you’re stuck, you know, and sometimes you’ve got to play around with it a bit. So it’s not very error friendly. And even if you’ve sort of got it right--but you know in the printed literature MSS are cited in a variety of ways. Which one is THE correct one is not always easy to figure out.”

would already know the shelfmark. “it’s a little harder if I don’t know the shelfmark, if I’m looking for a genre, so say, a different type of Bible commentary, or an author--with Hebrew authors, you know, there’s usually more than one way of transliterating them. And it also depends on the language, so for example the sound sh in English it will be an sh, but if you had a German digital collection it’s sch. If you’re in a Polish one it’ll be an sz maybe, and so on.... There is a fair amount of trial and error.”

how long she’d spend: “It depends. If I happen to know that the MS I’m looking for is digitized, I will probably be spending as much time as is necessary to track it down. But you know there comes a point where I might just--if something isn’t going well, so for example I managed to track down an MS in a Polish digital collection by using Latin. But then I had trouble with actually accessing the digitized MS, and in the end I had to write to them and say I don’t know what’s going on but I get as far as this and no further.... They did write back and said they would pass it on to their IT department but I haven’t had a resolution. And, you know, resources, I understand.”

type shelfmark into Google: “I have done that, yes. I would discourage my students from doing that...because I’ve found that it usually doesn’t work. It usually takes you somewhere else which isn’t what you want. You know, it’ll take you to somebody’s blog or somebody’s website that mentions the manuscript, and it might even mention that there is a digitized version with a link and the link is broken, so it doesn’t tend to work. I think there’s some kind of a wrapper around the digital collection that you have to actually go into--again there’s some exceptions, if you know how to operate it you can get to some of the Munich digitized Hebrew MSS that way, but it's usually not worth it.”

colleagues: “very rarely. But that's because you know I've been at it for a long time, so I think of myself as experienced. It's probably wrong but that's just me, you know, I tend to work on my own most of the time.... There is however one exception, which is a group on Facebook called Hebrew Paleography and Codicology, which has people like me writing in to say does anybody know how to--so occasionally I'll do that, and people will help each other. So that's quite a specialized interest group.”

searching in Hebrew alphabet: “I tend to use Western computers and I don’t always have access to a Hebrew keyboard, so that only works where there’s a little onscreen keyboard. I’ve not seen that on offer. It could be a very good thing because you could bypass the transliteration problem, but it would have to be accompanied by an onscreen keyboard, which I gather is not that difficult. So that would help, and I imagine it applies to a lot of other languages as well.”

7.3.3 Contacting library staff

“I would see if there was a contact form on the website I was expecting it to be on, of the library, and failing that, email/call them to see if it has, indeed been digitized and I am too stupid to find it, or if it hasn’t been digitized”

“I have been doing this for four years now, and I think maybe twice. Again because there just isn’t all that much digitized material on what I work with, or it’s all here, and if it’s here I usually just went and knocked on someone’s door. There’s been one case in Berlin where basically the image was meant to be there, there was a link, but it didn’t appear ,and I just wrote them an email querying, and never got an answer but the image was back a few days later. The other contact was with Lerevan, where we asked for images to be digitized specifically, and I think it only took a year to be realized, and once they’d done it we asked them to do it again because the quality was very unique.”
Help from librarians: "If I know them personally then yeah... because I'd worked with the manuscript previously I know the staff there, and I wrote to somebody directly to ask if it was possible to buy an image of this and request permission to use it, and they replied in the same day saying well here's an image for free, just include an acknowledgement"--chasing up permission at the Bodleian was "a bit of a pain in the arse" "generally speaking, if it's a helpful institution"... librarian at Cambridge college went off and took pictures of the manuscript itself to send along, "he was fantastic. That was just great"--writing is good "especially if it's a smaller library. You send something like this to the B n in Paris, and you know they'll be too busy deleting it to actually get round to despising you for asking the question in the first place"

library staff: "I don't think I have. Just to order scans."

"Actually no, I don't. Here, if I had an unwieldy problem I'd go to Martin Kauffmann by email, he's difficult to find in person, but otherwise no. I wouldn't go to the library staff, because I'm not sure quite what the library staff around in the various regions, what their expertise is."

library digitizing a Wycliffite Bible "let us know that they were going to do it, and then very kindly said just come along and, you know, plug the hard drive into the computer"

"Once or twice to ask specifically about, is this digitized, and I think once to say 'I think your digitized site isn't working, is it broken, are you doing maintenance or something'--but it's not often."

"Yeah. Actual librarians are almost always responsive, and helpful, though sometimes all they have to say is we haven't digitized that yet, or you know, we plan to but we haven't...it's not something I often find myself doing because often I can establish reasonably confidently that it's just not there or it's definitely there. It just takes more time than you'd like."

"Often hugely helpful. I mean, cases where, you know, they might reply with a PDF of the digitization for free, you know, within an hour. That doesn't happen every time of course [laughs] but generally yes, yeah. They're only too happy to point you at least in the right direction."

"and then what I would do is either go to the library or write to someone there and say could you send me an image of whatever it is I'm interested in"--libraries are helpful with this, and ability to photograph in libraries is huge, "makes an enormous difference to research...and then there are wonderful resources which I use like State Papers Online, which is fantastic, and I sit there printing out Elizabethan, Jacobean documents, and I'm happy"

librarians: "You know it sometimes takes quite a long time to figure out who the correct person is to write to. So I've not done that very often... I tend to only do that when I need to order photographs, you know, for publication. But also I think I probably have a perception that they're very busy people, and they're probably sitting there thinking, 'Why can't she find this? What's wrong with her?' which is probably not fair, but I sort of feel that I should only bother them if I'm absolutely desperate, not as a routine measure."

7.3.4 Advice to students

There are a couple of useful dictionaries that I mainly told [students] where to find and how to use the interfaces"--mainly linguistic teaching so not so many manuscripts although "I've occasionally shown them images of what the language they're studying looks like in real life as opposed to print, but those are images I had already as opposed to going to Digital Bodleian and finding something"

"It's probably an erroneous assumption, but I tend to think that [students] are going to be a lot better than I am generationally at finding stuff online"--has only given guidance to using Digital.Bodleian

"At the MA level I use it for paleography, so it's absolutely essential there. I ideally have students consulting stuff online--occasionally I've actually circulated the images to them--so that of course is fundamental. Anything involving images...in my lectures I'm constantly using digitized images as a pretty adjunct to the textual component of slides and things"
gives undergrad students “quite detailed help”—stable URLs for particular leaves or MSS, or instructions “go here and find this box, enter this particular phrase into it, and the MS should be this result”

undergrads don’t do much with medieval MSS in the most part; but “I might want them to look at the images” in a particular illuminated MS; get “a taste of how different MSS are”—or compare different copies of Troilus and Criseyde so students have an idea of how it appeared to readers in the 15th century

“For postgrad teaching, unless it was a very annoying repository, I’d be inclined to say, it’s online, on this site, here’s the URL for the site in general, let me know if it’s not loading properly or you can’t find it—because really they will very quickly have to be finding MSS themselves so they may as well learn how”—fairly quick at picking it up; attributes this to being “5 or 6 years younger than me” and “usually even more comfortable with websites than I am”—never hears about them having trouble normally

“The aggregated lists of MSS on my blog I did as teaching resources. I mean I partly did them because I was doing them for my own benefit and then I thought I’ll put these up online and then require my students to use them. So all of my courses I think without exception have looking at digital manuscripts as part of the deal—looking at mise-en-page is really important....I mean they like looking at them. They’re often very pretty manuscripts, with pictures of dogs chasing rabbits and things like that. So it’s a good thing to do in general.”

“I tell them to look on my blog under the teaching resources”... “The more enterprising students occasionally come up with stuff that isn’t there, or I’ll get an email saying oh you say this source is not available online but it is now. So I’ll update my page. But that tends to be my graduate students more than the undergrads”

“When you need to show something, to see all the scribal habits, and also to make people transcribe.”

“I actually started thinking well I’ll just tell them to find it, you know they’re all better at this than me”—but then decided it would be more efficient overall for her to just do it

“Over the vacation they had a kind of EEBO exercise to try and get more familiar with that and work out how many editions of something there had been and some element of how the book had changed over these multiple editions. So they used a combination of ESTC online and EEBO to do that.”

“I give students individual mini research projects, and say to them well you won’t be allowed to go see the original as an undergraduate, but you can actually get an impression of the entire manuscript in its entirety by going to this website. That has been very helpful and it’s yielded some brilliant student work.” lets students find things themselves “but they’re often not quite ready for that. The really brilliant ones often do come back with things that oops I didn’t know about. Often there’s time pressure, you know, they’ve got two weeks to produce the work, so often I will suggest to them why don’t you work on the Barcelona Haggadah, and this is where you find it on the British Library website. And then if they’re adventurous they will play around and find other examples.”

7.3.5 Opinions of existing interfaces

“The Luna viewer...is quite useful if a little fidgety. I haven’t yet had the opportunity to work with Digital Bod that much. Octopus is a pain. The University of Salamanca, that has been doing this vestment project, digital vesti, that was well presented for the user when you first come to it because they work with too many levels of layering, if you like—so from manuscripts to tractates to God knows what might be in any individual binding, and how they’ve taken it apart—so it took a little while to see it through their systematization—but otherwise it loaded reasonably quickly, you could have only a single page if I recall but that was usually enough and you could just flick through it if you liked...in Vienna the emphasis was more on the actual text, so everything was organized by fragment number... you would have the transcribed and then transliterated and translated text underneath, and a reasonably high-res image of the manuscript page. But again, very focused on one particular language tradition and manuscript holding, without concerns of interoperability”
“I’ve always been surprised that with the Bodleian being such a big research library, and the amount of people here at all levels working with manuscripts, we’re still relying on print catalogues and—when I started—green slips, rather than it being the norm that you have an online system or email someone.”

ten seconds of silence when asked about digital resources that are unusually good, followed by “Trinity College Cambridge. I find stuff relatively quickly there. I’d love to say Digital Bodleian but it ain’t true. Because the search mechanism there turns up some very strange results when you’re looking for stuff, and the collections feature doesn’t work very well, at least in my application of it, so I can’t say that... Cambridge university Library I found pretty quickly. British Library no. You find stuff there via Google. Their own internal search mechanisms are diabolical. Paris again...search, even when you tweak it, turns up some very strange results... So yeah, no, not really is the answer.”

“The things which are coming up online now which one can access--the quality is on the whole superb, and I looked at the Laurentiana in Florence, the Bn is a bit dodgy in a few things it has, the BL ones are splendid, Bodley ones where one can find them are pretty good quality images, but they’re not always terribly easy to manipulate. I don’t quite know what’s going on. Sometimes you get a manuscript which you can sort of turn the pages of, you know, sort of glide from one to another, other times you have to sort of go back and click on each image separately”

“The one that I’ve found most useful has been the BL, and I was impressed by the way you could go from a manuscript back to the catalogue entry and then into the portions of the MS, and very easily move through the MS. That’s the easiest one I think I’ve found so far.”

“My personal experience is that things have just got better. The accessibility and ease of use has been improving.... The speed with which things have changed is fantastic.”

new systems frustration: “Oh yes. Muttering to myself. But no, one simply bashes away and prods and eventually hopefully sort of gets something to work.”

“when I was thinking some very rude thoughts about the BL and also about Bodley, I then did discover that if you read things properly there was stuff on the site. One has to be patient, one has to not just sort of look quickly at the screen and then rush away--there is a reasonable amount of material and if one pays attention it is helpful. But I think sometimes it could be signposted rather more properly.”

“I had been very frustrated with the Bodley site, which has a digital portal, and I suppose I was being dim and not realizing that you could put anything you wanted in there to get up some sort of response, and found myself going round in circles and looking at the sort of headings of Western manuscripts and exhibitions and whatever, and thinking ‘I can’t get at the manuscripts!’ so yes.”

UPenn: illustrations and images separately tagged so they have two dropdowns, one for folios and one for images, although “I’d probably want to do one steady clickthrough of the MS to check that it was accurate, because what one person thinks of as decoration might not qualify in another person’s head” preferred and unpreferred portals: “BL’s new setup is pretty good. Slightly counterintuitive. I think they phrase the link to the actual MS images in a way that’s somewhat misleading or something like that. But it’s easy to use and fast. Penn in hand...that’s good. Princeton’s is slightly hard to get into and has that collation issue. But they offer you images pretty rapidly. Anywhere that uses Luna is a bit of a pain [JR and Bod]. Harvard I think--there’s only one MS at H that I’ve had to consult online so far, I think that was easy enough. The Huntington Library hasn’t digitized very much stuff, but they have this weird thing where they have five photographs of each medieval MS, usually the front and back cover and first and last pages and a page from the middle or something, and it’s a bit weird... And it’s not very easy to find, and yeah, it’s just a bit weird. But it exists, I guess [laughs]”
smaller libraries “have simpler and more antiquated interfaces, which is annoying but I don't think less of them for it, because I can imagine that if you have three or four staff it's not easy to just say, ‘We need the snazziest, most up to date, most comprehensive system’”

Gallica is easy to use because it doesn’t require you to punctuate the shelfmark exactly; “it’s very user-friendly. You can immediately realize if a manuscript is there or not. You don't have to pay too much attention to describe exactly the shelfmark.”

“I have to say the inventories on Gallica are terrible”-- “down-and-dirty uploads of 19thc scholars saying what's in the MSS—it's basically serviceable but it's not very good for online searching because they don’t standardize the orthography, which is always a problem for medieval sources...so that's why I go to Arlima, and I just hope that Arlima's got all the stuff aggregated. Which it doesn’t always have, I mean sometimes I send them updates, but I think there's enough people who are sharing info and updating these things that it's fairly useful--I mean I'm updating Wikipedia...as I go, just because why not”

“I do tend to tweet Gallica from time to time and they either ignore me or retweet me depending. They mainly ignore me. And I sent them several emails about the dating of [MS] and I think eventually it got changed. But they never replied to me... I send updates to Arlima too, and I'm very happy to send updates, but just often the updates are not done. And I know with DIAMM, people would send updates and it was difficult for us to do the updates because of the way that data was managed, it was managed through King's and they didn’t give us the ability to change stuff via our site ourselves, and so we were then dependent on them and it got more and more delayed and that's one of the reasons we’re moving”

“I guess I’m often predominantly really using Gallica and then there are a handful of other sites I use time and time again, and I probably use them rather unthinkingly but have learned how to as it were.”

usability of Gallica over the years? “Well it’s pretty terrible. It’s improved, but to my mind there are still lots of problems. In terms of searching for your documents, like I said it's really useful to know how to enter--like with the BL, you have t’ know how to enter the information to search for it. Navigability is not great even with the latest version of Gallica. I want to be able to go straight to a specific folio, which you can't do for scans of microfilms in any easy fashion at all. Ideally I’d want--I guess when you’re navigating I’d want to be able to see the progression through a MS that makes sense, so I’m not just on a specific folio, I want to see whereabouts I am in the MS. Which you do get on some--the Bodleian I think might--so I find that useful. I mean, ideally you’d want--and Gallica is so far away from this it's ridiculous--you'd want some sort of--you'd be able to jump to the start of a particular text say, or even sections within texts, but, you know, that's--Gallica as I say is a long way off that. And you don't have bibliographical resources either”

“I find it quite frustrating. I guess one issue I find with it is--so yes you can zoom in, but you can't necessarily maintain the zoom or it's inconsistent when you go to another folio, so having that control would be useful, which I think say e-codices--I think you can do that, I think there's a checkbox you can tick that maintains, if you go to another folio, you maintain the same resolution.”

But I think [the British Library is]--great, but you are not allowed to download any image. Of course if you know how computers work, you can find tricks and ways, which is legal, it's not illegal. They are very high-analysis. I haven’t checked the new one, but presumably it hasn’t changed anything. It’s quite similar to that one, and--is there any logo? Perhaps not. Or watermark or something like that. Because for the Vatican, there is. Which OK, better than nothing, but--and you see some libraries have for the entire thing, the microfilm version.”

“If you are talking about digitizing manuscripts, you are talking about images, the quality of the images. And how easy it is to browse, how easily you can extract the folio, when you’re a researcher, you’re looking for a particular folio perhaps, or a particular section. And if you don’t give this opportunity to search, then it’s no use. OK, it's good to have it there, but then you cannot apply it. It would have been much better for the BL if they could allow you to download sections, and the same also here. There is a selection in some German
libraries where again in the form of PDF, as in BnF, black-and-white images of perhaps I think they come from microfilm that has been digitized. And this is something also happening I heard with the Sinai collection, the Jerusalem collection at the Library of Congress. They have the entire old collection in microfilm and they have digitized all of them and they will upload it.”

“The problem with Pinakes, except for France, is the Latin. But if you’re dealing with this kind of MS and you like paleography, then you have to know Latin. Otherwise you can’t deal with it at all. But OK, thanks to Google, even if you don’t know what...means, you just search and you find it.”

“In the worst scenario that Pinakes is not updated anymore or ceases to exist, what will happen? Given that it’s a free database, and OK, semiprivate let’s say, what happens with that? That’s the kind of problem that we need to address. Maybe we need to have something to make it more official, to make the presence of Pinakes more stable and more official.”

“I don’t have anything against Pinakes so far. Some people say it’s not user-friendly. I would say it is. That’s a common problem with all kind sof databases, some people find it easy, others not, and so on...some people are getting used to one form, and when something in the way of appearance changes, they get lost. This seemed to happen in Pinakes. I think it was three years ago, they made a great change. To be honest I don’t exactly remember how it used to be, but now it’s much better. And what I like with Pinakes now--which is not very clear that you can do it, but you definitely can--is that you can find the entire collection of hits in a library. So if you go to Bodleian, and then you get all the collections, and then you get listed each collection separately, and then you get all the shelfmarks. That’s fantastic” but not very obvious. “Some people use the search on a particular MS to find it easily, but no, you cannot find so easily a particular MS. You need to go through the library rather than the shelfmark itself. If I remember correctly.”

“It’s not that easy to use [Folger], to call up one picture after another after another, you know, that feels---I suppose that feels quite an early stage of digitization, that you take photographs of openings and then you put them up on a website, which is you know, great in all kinds of ways, but there isn’t any more kind of curated sense of let’s put this book back together as a digital object. I mean that hasn’t been done. But it hasn’t been done a lot, apart from with Shakespeare”--Bodleian’s First Folio for example--“whereas most of the other--there are other versions of that book online, like the...Library in Leeds, where you can turn the pages but you can’t extract any of the images from it, but I don’t think apart from Shakespeare I’m not very conscious of other books that have been done like that.”

“Handschriftencensus: “the biggest interface for German MSS”; can search by library, by work; “it started I think 10 years ago just with all MSS with any German in it from the 8th to 12th century...and now it’s all MSS that contain any German from anywhere around the world”; “threatened with closure because it had been more or less a private initiative...it just got permanent, more or less permanent funding from the German Academy...and from November it will be revived and then also with an English interface. At the moment you can’t send in corrections, which until March was possible, so there is a bit of a lacuna of updates since March [2015], but that will be updated in November”

ODNB “is one of the best”--“allows you to, especially with the advanced search and the full-text searches of various kinds--it’s the advanced search that I really like--it allows you to seek people not by name [but by role etc.] and I’ve used that a lot. So if I want to find what people were living in the area of in the 17th century, so could they have known someone interesting in that area, that allows you to do it. And it’s quite a good system. That’s one of the most useful searching tools that there are if you’re working on a specific problem. Otherwise I just go to Google and look things up and eventually come across”

“Some of the digitization programs that libraries use for non-printed book material, I find it hard to negotiate and navigate rather to find what I want, in particular Folger’s digitized MSS, which I’m very interested in...it’s very hard to get hold of it and to make sure you’re seeing what you think you’re seeing or what you want to see. And the same is to a certain extent true with Digital Bodleian, which I think is still pretty much under
development, and I don’t blame the Bodleian for that. But especially discovering MSS online is rather I find a rather random business because I’m not looking at the right place, and what libraries tend to do is catalogue printed books very efficiently, because of author title date, and their cataloguing of MSS may be very good but it’s very hard to find what you’re looking for and then to get from the catalogue to a digital form of the MS, and that I think is something that needs more work. I mean I think the Folger--I love the Folger and I go there quite a lot--presents you with all sorts of wonderful images--digital collection on Luna, this thing--it’s not terribly helpful. Browse all. And it’s got thousands and thousands and thousands and your chance of finding what you want is pretty low...I think I would find a way of going to the catalogue and looking for things there, but it’s not terribly well done I think.”

ECCO “lists all the illustrations, not terribly well, but it gives you a sense of what’s there, so if you’re looking for a particular image, it’s quite easier to find. Easier than on EEBO.”

interfaces “usually changing for the better” and not too difficult to orient to new ones. “It’s the old ones that use software that’s obsolete that can be more of a problem. Like the infamous Deja Vu software, which first of all you’ve got to download it and it’s free but if you’re working in a university environment you can’t download anything.”

7.3.6 Metadata

“I wouldn’t rely on the metadata, partly because the errors tend to perpetuate themselves. So a good example is this...manuscript...I was looking at this because I’d read some dimensions of it in a work and I thought this doesn’t sound right, I’d like to check whether it really is as small as they say...and what you could then see as you followed the crumbtrail back is that--and I don’t know where it came from--somebody had come up with some fallacious measurements at some point, and these had then been copied by someone else, and you could guarantee that these would then be taken as the basis for the metadata, so if there’s anything referring to the physical dimensions--and then when you look at the manuscript and see that it’s supposed to be 16cm high, it’s actually got 2 columns of 27 lines, and you know something is wrong there”

“the things you might put into the metadata--provenance, dating, physical size, quiring--all of these you’re probably going to want to check yourself one way or another”--cites difficulty of “accommodating to the standard model” of metadata when digitizing things

“Perhaps, yeah. Perhaps the roman numeral date was transcribed wrongly. There were some where I knew the title but somebody had given a different title. The proper title was An Introduction to Metaphysics but the title they’d given was something like Studies in Reading Religious Texts or something like that.”

metadata in DB is “very brief, but it gives the essentials, and I think one can track things down”; scribe would be the next thing she’d look for

important metadata “depends on what you’re looking at the MS for”; date “although I might find myself differing from it, but having it there is nice”; provenance info, which can “save you a bit of a wild goose chase if you’re coming across names in the MS in the margins”; “what texts the library thinks are in it”; “most people who work with MSS just don’t search by title” but understands it might be useful for other types of MSS; would care about reuse info but not necessarily right now; “the category of info I would want to have available but probably wouldn’t read”--but “there are points in your career where that becomes very significant”

“most library catalogues are written by people who are experts on lots of things, and they’re extremely good, but they might not be the world expert on the particular manuscript that you’re working on, and it’s not common to disagree with you know the dating in the Summary Catalogue. So it’s not so much that I might not trust it as that I might come to a different conclusion”

“I would distrust someone doing a search and then saying there are X many manuscripts in this 50 year date range in such and such a repository because that’s a really dubious way to handle a bunch of numbers which
are actually not—in most cases they’re determined paleographically and paleographic dating is very inexact in my view”

“I wouldn’t say I sit there being hyper suspicious of every bit of metadata. I’m aware that errors will creep into things, but you can turn that into a counsel of despair where you just don’t do any research, because some probably relatively small proportion of the things you’re dealing with—paleographic dating is a special category because the date ranges you’re putting on things are so wide”

all his MSS have at least one title, “or they can be named in several different ways, so you have to spend some time Googling all the different possibilities”

“you’d be surprised by how much stuff—of course you have to be able to distinguish between false information and what sounds weird, what sounds realistic. Of course if you find info about an Arabic MS of a certain type written in this particular period in Toledo, and of course that’s impossible because Toledo was already conquered by Christians in that particular period, so you rule that out, you say OK this is nonsense. Or for instance if you find info about a paper MS copied at the end of the 10th century, that can be ruled out because paper was not used yet in that period, so it’s probably a later manuscript that has a colophon that gives a date, but it’s not the date of the copying of the MS, it’s the date of the original MS that was copied, 100 or 200 years earlier”

“I’d like datings, but dating of MSS is a complicated issue, so I would want to have the datings and I’d want to know who’s dating them to those dates, like which person and when, and in what publication and on what grounds. [laughs] That’s pretty hardcore…. I mean they give crazy dates to things. I’ve moaned to them about it before.” [in one instance] “they just put the dating that they had in the old catalogue” which may be decades or more out of date, made obsolete by new scholarship. “I think it would be useful to have a better mechanism whereby people could feed back stuff. Clearly it can’t be totally crowdsourced because it’s very specialist information, but you know, if you could register and they would know for example that I’m registering and I’m not just any old body [professor in blank at blank], maybe my registration would be subject to a process of minor review, to just say who you are, and then when I send in stuff then it might get put up. But it would be ascribed to me, because then if one of my colleagues says no I don’t believe that dating at all, and sends in a rival bid for the date, then that could also go up with their name and their grounds for thinking that too. So I think it has to be slightly more of a discussion around MSS, because they’re complicated objects. Sometimes they’re put together later anyway”

“It would be useful to know which pages have illuminations on them and which don’t. So to have some kind of annotated—though really I find sort of big whole pages of thumbnails of the MS not useful because you can’t really see, and not big enough to be useful, but what would be useful is if you just have a list of 1r 1v, 2r 2v, but the list has a little plus sign that you can expand and it will tell you what’s on there and/or whether there’s an illumination of that bit. Or music notation.”

“I guess in some ways you just want as many things to filter by as possible [laughs]. So I worked on the medieval francophone project [encountered difficulties with metadata]...we would want things like origin and provenance as well as date. Maybe names of artists and scribes, and you know--I would want all of those things, but--yeah. It’s difficult in a way because you’re often working on such a narrow project, right, it’s difficult--well, what you want to filter by is going to be so specific to the project I think. But for me, yeah, I guess origin and provenance would be right up there, and language of the text.”

“If we’re talking about descriptions of MSS that accompany a digitization, which are a really useful resource I find, I like to know where the information has come from I guess, which you don’t always have, but--I’m thinking of e-codices, say, where that is basically just scanned-in material from print sources, but it’s--I like to know that that’s where it’s come from. It might be out of date, but it’s still good if there’s at least a record of where it has come from. So having some sort of author there is useful.... Gallica I don’t think is particular comprehensive in providing even a brief snippet of information about the document you are looking at.”
"There is one notorious case, Sinai Greek collection. It has at least three different shelfmarks with no logic. So you need to refer to the particular cataloguer...some people catalogue based in the contents, so all the biblical manuscripts have say shelfmark 1 or a, but that's not the case with some other catalogues, where they follow a kind of sequential, the way they've been once catalogued. Sometimes this happens also with Athens MSS, but not always. But the most notorious to my knowledge is Sinai."

What additional metadata would be useful? “Maybe yes, but I’d need to see that particular record to know if I need more or not. A good thing to include is bibliography. This is where Pinakes is good, but they do only the new bibliography. They presuppose that you know the old bibliography. So they made a selection--I’m not sure how this works, I mean their mentality. You might have a very famous MS where you need 10 pages of bibliography...then they may have to do a selection, and they choose the most important. As far as I remember, Pinakes bibliography cannot exceed one page of the normal screen of a laptop or whatever. So yes, looking through bibliography and finding some articles you were not aware of would be a good thing. But sometimes you might have just a single reference to the articles. But OK, you cannot expect to have this detail in a database like Pinakes. I think the library has to list all the references to a manuscript without discrimination.”

“In general you find what you are looking for. But ok, if I have access to the images, I don’t need the further description. If I can check, I can find my way there.”

problem of misbound MSS: “what would you do? especially if you digitize the entire MS, would you put them as they are, or as they used to be? Or sometimes you have a page upside down, and then you have the conservators who say ok, don't touch it, this is a historical object, you have to represent it as it is now, and then it's a kind of gray zone.” His preference “yes and no, depends on the case. Because sometimes you may have a kind of hint that ok, you have a MS of 13th century let's say, and then in the 15th century it was bound in a different way, and they didn't pay attention or they didn't understand that this happened, so you need to know that these guys were careless or there was something behind that. Perhaps this might hint you that they were foreigners. So you need this as historical evidence of how the MS or codex has been used. So you need both. It's not one way or other.”

“All additions are welcome. What I said is that OK, if I have all the images it’s fine for me. I do not speak for all the researchers. To be honest, let's say I'm looking for maths. And I have a mixture of Greek and Arabic text. I would like to have the division, of course, I need it. And some people are not familiar with poetry or hymns, they definitely need this kind of division. But if the library does not have the funding to support proper cataloguing, although cataloguing is less expensive than imaging, then I would welcome to have at least uploaded images, rather than waiting for the--”

search for keywords ever? “Very rarely now. It's really handy to have Pinakes and Diktion.”

metadata correct? “Depends on the library. And sometimes yes, I may have worked on a particular MS and found more things than the particular cataloguer did, especially the identification of authors.” we have more information now than in a 1990s catalogue. “e if the library just reproduces an old catalogue, you might have these kinds of things ignored...I wouldn't dare to say that this is quite common, but it happens.”

trust electronic metadata? “I suppose I sometimes have the impression that the metadata has been done by somebody who understands metadata but not the book. So you feel as if different kinds of catalogue descriptions or different kinds of ways of describing the book come from quite different kinds of knowledge. Because sometimes I see things that are inconsistent with--sometimes metadata is very careful about some aspects of the book and not very careful about others, so it might not necessarily preserve --it might give the title in the form that we would--an abbreviated form that we would usually use rather than the title that's there on the page or something...I think if you’re trying to get any kind of read across these different resources that might--any kind of sort of gathering of data from different sources, it would be a problem. You can already see that on something like SOLO, our library catalogue at Oxford, I was looking to see how many
copies of the Second Folio we have in Oxford, and there are about five different records for the same book, each with two or three different examples of holdings which together mean that we’ve got eleven copies of the book, but they—you know, they’re not—just because of the absolute specifics of how you title that book, they don’t come up under the same [record].”

“it’s a fairly tight-knit and quite digitally conscious community, the German medievalists, and there are a lot of people controlling it. So there was a huge outcry when they deactivated the field for sending in [corrections]. But people knew that it was the one-stop portal for anybody engaged in manuscripts in the community of German medieval studies. So whenever somebody had published something, they would immediately send in the details of any corrections, and they were very quick in updating it.”

“it’s peculiar and fortunate in that there has been an early digitization drive and cataloguing drive in particular—so the cataloguing drive preceded the digitization drive from the German research foundation, the DFG...so they did this extraordinary detailed cataloguing work, which is all open-access. So there were very detailed guidelines how to catalogue MSS and they have catalogued pretty much all major German collections...so the metadata are impeccable”...nothing else she’d like included in metadata. Catalogue includes info on damage, foliation, quiring, decoration, initials, sewn-in veils to protect initials, provenance history, etc.

“So the metadata question—yeah it’s—they put all these things up but they don’t always make it terribly easy to find them, is my experience”

“Some MSS are sort of known under a nickname, so here it would be the Laud Mahzor...the Ambrosian Bible...it doesn’t apply to many but there are some that have nicknames.”

“Date, date range is another possibility, and that’s what they do in Munich, and it works. But, you know, I’ve tried that somewhere else, in a Polish library, and I didn’t get very far because their dating is very different from my dating. It seems to all be a century out, like somebody made a mechanical mistake somewhere. So yeah, that only works where there’s checking and double-checking and it really is right.”

catalogue source attribution: “That’s essential. It must be added. It makes a big difference because you have 150 years of cataloguing, and standards have changed. Also sometimes there’s more than one catalogue of the same collection, so the same manuscript will have been catalogued more than once by different people at different times. You need to know which one you’re looking at.” Finds that websites do give this: “If they give any catalogue information, they specify which it’s from.”

7.3.7 Desires and priorities

“First of all that you can find everything as quickly as possible...shelfmark is important because if I come across a manuscript it will usually be named in one fashion or another, so that is the most immediately useful I think..if you’re searching on a particular topic then I suppose something like Library of Congress subject headers would be useful so if I look for person X I know that yes, he is mentioned on folio X of that particular manuscript...so I think the incorporation of metadata is important if you want to help researchers engage with manuscripts that they don’t know about yet...doesn’t have to be a great deal, but dates essentially and subjects. I could go haywire and do it for every single folio but that might be overkill.”

“In terms of interface, it depends entirely on what you want to do, you know. I like the idea of having the ability to portray multiple MS pages from varying MSS next to each other to compare, especially if you’re working with one MS that is likely to be in Vienna and the other is in St Petersburg, you’re never going to have them side by side, and in the interest of saving paper you don’t want to have floating copies around, and those might not be good enough anyway. So that I think is quite useful. Increasingly so, having a proper system like SOLO or basically a library cataloguing system that is digitized is useful for MSS, as it is becoming ever more internationalized and people I think are quicker to adapt to an online or digitized system than they are to
what form do I have to fill in now and how long is it going to take’--it’s probably going to speed up scholarship if they can order something before they come up, rather than having to sit around for a few days”

"Um...that’s a good question. In theory, it would be nice if the images were accessible in such a way that you could dynamically load them into a program, so if they had very easily accessible permalinks or something like that. I know that a colleague...is working on programs that aid in the transliteration of MSS...essentially you might be able to tell a program here’s a range of images from a MS or variety of MSS by URL, and the program can then load them and then do its thing. So for example look for all the lines in a manuscript and be able to then load it into its own interface and give you the opportunity to transcribe the MS line by line”

“In particular cases it would be useful if you could have a variety of images associated with a particular folio, so for example if you’ve got a particular kind of illumination where the angle of view completely changes the colors, or where we’re talking about gold or silver ornaments, and the luster is quite different, and you can only see from a different angle how raised for example the gold is or how it has been applied, that would be relevant--or palimpsests, which we have occasionally worked with, similarly the integration of hyperspectral images in as far as you ever do them just on a whim--but I think you know having the ability to see ah, we have another image of this manuscript either hyperspectral or spectral form, that would be--or the ability to request something like that”

“Metadata is mainly relevant if you want to discover things. If you don’t read catalogues front to cover, which I think only very few people want to do--that’s the only way of finding something that people haven’t talked about.”

“If you had a construction of the digital artifact in 3D, then you’d get to a point where you wouldn’t have to be messing around opening it like this, checking the binding and things, if you could actually get it in some digital form where you could go round it and do all this interaction without physically handling it that would be kind of useful”--“and there are some things sometimes that you just need to see or feel, and I’m not sure how practical it would be, but that would be something”

“It would be good if you could do the logical route, which is you have your reference, the shelfmark, and then you actually go via that--so it says London British Library, you go to the British Library website, and then you have a reliable easy way to find their manuscript catalogue and then what you type in actually turns up the thing you’re looking for. That would be great. It doesn’t sound like a particularly demanding desideratum but it doesn’t seem to correspond to what actually happens”

A union catalogue “would be awesome, but--I’ve used something similar to that for France...and I ended up finding something via some completely unexpected route, which wasn’t using the union catalogue--but if it worked it would be brilliant”--“has used Europeana (but barely remembers the name) and e-codices (doesn’t remember the name) and found a few things there, “but it’s probably because the resources I want haven’t really been digitized, but I certainly haven’t had enough success that I’m reliant upon them”

“That it’s complete, no pages missing, no folds obscuring pages, which happens sometimes. That it’s legible, it’s not too huge a file size. Freely available.”

“I would like to be able to find a list somewhere, a catalogue, which--now how is that different from now--I mean, to be able to put in--now in fact both the Vatican and Bodley, there’s obviously overlap here because of Polonsky, have got lists of DMSS, and you can scroll down and see whether the one you’re after is there. That’s helpful. And as Pinakes gets more organized and has more info you could go to Pinakes and see whether it’s digitized. It’s already starting to be useful to do that. So that means that’s a consolidated set of information. That would be the ideal way of going about it, one set of consolidated info which has a brief entry and an indication of whether it’s digitized.” Through the catalogue as well would be good. Lists of shelfmarks are important.
“If you’re going to store these things and organize them by shelfmarks, you might as well organize their surrogates by shelfmarks. And to be fair most of the use people put the surrogates to is as surrogates for the physical object, and that’s not a bad thing in itself.”

“Some kind of quiring diagram would be really nice. I think it would be nice if it came with a disclaimer, saying this is what we understand the collation to be, it isn’t you know—and I would know that it wasn’t necessarily right, but when I was a master’s student I wouldn’t have necessarily known that—but some kind of way of indicating—if you could pull—have something running across the top of the screen that would show you were in the physical structure of the manuscript—at least where the catalogue thought you were, that would be really helpful”—Princeton “sort of does this but in a really unhelpful way, where they’ve kind of gone too far in the other direction, where to navigate in the manuscript, it actually gives you a list of quires and then a list of leaves within the quires, rather than letting you pick a folio number”—would like a middle ground “where you could access the information but not be forced to navigate using it”—although “it’s wonderful to see a library foregrounding” codicology ruler is always helpful “but at the same time I just don’t think it’s safe to do work on size from digital images, so in that sense it’s not really needed and it might encourage people into a dangerous belief in the reliability of engineering measurements from rulers in photographs”—it would never be flush with the page, and a medieval reader would have seen the page curved; also reluctant to base conclusions about rubrication etc. on color of digitized images; “certain things I’m just not happy to rely on digital images for”

“It would be really nice to be able to grab a stable URL for any given image or any given MS in a clean obvious way, and that’s not always the case”

Things you’d like to be able to do with digitized images? “Downloading is crucial, making them large enough to see features—I can pretty much do everything I need to do with them I think.” quality problems: “Gallica’s got quite a lot of stuff where as a stopgap measure they uploaded a lot of microfilm images in black and white, and the pain with those is that there’s no navigation. So normally with the proper stuff you just put in 120v and up it comes, and with those ones they are all NP which means no pagination. But what I can do—and I was going to do a blog post about this because no one else has noticed it—is if you look at the URL at the top it has actually got a folio number in it. So I just go up and edit that. But how hard would it be for someone to go through and edit the folio numbers?” URL numbers are just numbered from first image, so you have to account for front matter in calculating.

Hyperspectral; “Occasionally I look at them if people have got them up and about. I don’t tend to need to see that kind of thing. My MSS are sort of luxury productions so they’re often intact....actually there’s one....that’s got some erased bits in the index, and I suspected I could read them just by making them large, but I know that [someone] could basically photoshop them to maximize the traces of the old ink color and could probably tell me what’s under there. But in fact someone else has done it already, and I was right, what was under there was what I expected”

“Well, in an ideal world maybe there would be a single viewer, right, across institutions, which would allow you to compare MSS and go straight to particular points in a MS or multiple MSS, that would be wonderful. Maybe you could access these things both online and offline. That’s a bit pie-in-the-sky at the moment I guess.”

List of shelfmarks? “Yeah, that seems like an obvious point of reference. I guess at the moment that’s so fragmented, it will vary from one discipline to another. Perhaps what needs to happen is the various separate projects need to actually talk to each other so at least there’s some sort of compatibility. So maybe thinking about the future, there’s a possibility of coming together.”

“I’d really love to able to navigate them far more easily by jumping to particular—I’d like bookmarks or—maybe the first of that is being able to annotate more easily. But then it would be great if that was also publicly available as it were, so that—I mean perhaps we’re talking about divisions based on editions, say, or
even divisions based on the structure of the MS itself, so major initials, say, the way that the MS itself is structured. Perhaps both."

"I’d love to be able to--yeah, compare multiple MSS easily"

rotating possible in Bodleian viewer. "And I don't know why--actually this changed, at the Vatican, but before that, at least in the Vatican one, they used to give all the folios in thumbnails. Not anymore. That was really bad. And it was quite annoying for the old, let's say before September, Bodleian website, when you could just go through and you couldn't get any thumbnails, and you just go through and zoom in zoom in zoom in without anything else."

"if you are allowed to download at least the image, you can do whatever you want. And this is what you have to do if you are a proper researcher. I mean if you are researching and you are trying to find out what's happened, that’s why also sometimes these images area really important and helpful, and they give you a better angle than the MS itself. Imagine that you have a very tight MS and something written has been faded or something like that, and then you have the good quality image, you can zoom in, you can play with software, and you find a better image. Like using x-ray."

ideal digitized book? "I think it would be a series of images that you could look at--you would probably have a series of thumbnails, so you could go to different elements of the book, but I would think you would also have an ability to sort of as it were turn from one page to the next page and the next page. I think it would--one really useful thing it would have done would be it would tell you what the signature of each page was, so that when it’s not there at the bottom of the page...you then have to leaf back, don’t you, to find out what the number is in order to make a reference of it, and if it would tell you the reference that would be very helpful. And I think you would be able to download individual page images, and I think you would be able to enlarge it reasonably well but not necessarily massively."

incorporating copy specific details: “Including them in the digital catalogue would be a start. I think certainly in some ways I think for sort of annotations and marginal notes and stuff, we probably just need it to say ‘marginal notes in a query 17th century hand’, something like that. I don't think it’s realistic--if anybody’s interested in that, they’re going to want to go and look at it, either go and look at it or look at a digital copy that can show them more. They’re not going to get what’s in the notes from the catalogue record. You would never use that as your scholarly source. That would just tell you this is worth, I need to look at this myself. I think it would be good if we had some kind of--one of the things that Hamnet does is to have different categories of proper name that are associated with the catalogue record, so it might have sort of an author who’s named in the book in one field, and an author that's been attributed later might be another field, and then publisher's name might be another field, and then associated owners, so you can search all of that. But that’s because I’m particularly interested in the people. But I guess also if you were interested in later ownership patterns they would want--yeah, the details you would get from bookplates or other kinds of things. So I think if it were in the catalogue that would be good--I think that is coming in the new online ESTC, I gather that that is supposed to have copy specific details."

ideal finding process? "I guess it might be useful--for digitized early modern books, I’d quite like to be able to do it as part of the collectiveness of SOLO here in Oxford, which then if it doesn’t have it, says take your search to COPAC, take your search to Worldcat, it would be great although it would be a massive job--not necessarily, actually, if people made a digital book it could become a catalogue resource like any other one and you'd be able to find it--I’d quite like to be able to do it that way, because at the moment what I do is--and again, this may be that I’m just inefficient and there may be a better way to do this--so I would you know kind of follow up and bookmark when people--particularly book history people tweet a lot, book history Twitter is quite good, and they will quite often say, did you know that the University of-- has just digitized this herbal or something, and I’d keep or bookmark those posts, and it would be great if that were just something which would tie to the record of the book in some kind of capacity. I don’t know quite what the process would be. I feel as if my use of and promulgation of digital books is very--feels a bit random to me. It feels like what I pick
up is out there, or what someone else says, or what you know I go to a lecture and somebody says this image is from X. So if I’m looking for one now, like the Shepherd’s Calendar, I thought ok what are my go-to places, Folger, Furnace, things like that, and had a look. So I’d like it to be easier to do that in some kind of way. And although for my research I would like to be able to--I suppose the copy specific stuff is what’s most useful to me, it would just be --I don’t think I would have particularly sophisticated search requirements, I’d just like to be able to find things” and would be willing to experiment searching

“So there’s a site...where they have digitized a huge number of copies of Shakespeare’s quartos...and those are high-resolution digital color scans, and for all sorts of reasons I’m very interested in individual copies of those quartos of plays, so being able to see 3 or 4 different ones in high-quality scans has been very useful, and I’ve taken advantage of them in lecturing to scholarly societies. The thing I’m interested in is when leaves were missing from valuable printed books...people would do pen-and-ink facsimiles or they would do lithographic facsimiles or they would do typed facsimiles, and I’m interested in the people who made those facsimiles and in when and how they came to be put in the book. So you look at half a dozen copies of the first quarto of King Lear, and you hope that you’ll come across one which was imperfect...so looking at those high-quality resolution pictures makes it much easier to detect facsimile material than looking at black and white. So you know that the images on EEBO are not of the original books, they’re images taken from the microfilms, and those are not good enough, and they produce very unpleasant images when you’re talking in a lecture...if there is another image available then I will try not to use the one from EEBO, which seems a bit ungrateful, but they’re very flat and they’re very dark, and they don’t give people the sense of excitement that you get from looking at [higher-quality images]”

what happens with DB is it starts with the image and they supply the metadata, which is, you know, OK. What I’d much rather do is you start with the catalogue, the metadata--but this is better than you’ll ever get on a library's metadata--and it takes you from the catalogue to the thing. That’s what I really want to be able to do, is click on an entry to an author...I mean it would be fantastic and it would advance knowledge enormously, because you would be able to see lots of MSS in terms of what they contain, what is known about what they contain. I mean you can sort of do that a bit with State Papers online, because they included the metadata of the Calendar of State Papers with the images. The trouble is that the Calendar was done in the 19th century and it's inconsistent...ESTC does this a bit, where if you go to an obvious--I think the last line, if you go to that, takes you from the catalogue entry to the thing itself, and you can click on that and you've got the actual book coming up from the catalogue. So ESTC links you to EEBO and ECCO”

“color makes all the difference I think” but no other preferences for viewing

“I think in a perfect world I would want to be able to go to the library’s main web page and find a little button somewhere, top left or right, or somewhere where the menu is, basically, the main menu, to see the digital collections. Not hidden away, but more prominent. And then once you’re in there the search engine, if that’s the right word, needs to be a little bit more tolerant to--you know they sometimes ask things that even I can’t answer, or it’ll be set up in a fairly inflexible way, like author and title--well, when you’re looking for medieval works that’s not always the best way to go about it. Or there’s only one way to type in the shelfmark and everything else is wrong. Now in Paris they’ve got a way--they’ve got a pull-down menu, where you can look for the shelfmarks and you don’t have to type something in and be told that’s wrong, but you can scroll down and find the shelfmark you’re looking for. That’s not a bad thing, I guess.”

a link to catalogue entry “would be fantastic. Or some other information on the side, you know, ‘catalogue, see this’.”

“In the end even with--yes obviously it would be good if the image was legible.” Inadequate images “used to be common”

“I’m sometimes looking for things where I do need to see the original, because of erasures and things where you have to look at it in a particular light, but that’s a very specialized--most of the time it’s fine [to use the
digital surrogate only]...you can sometimes make [erasures] a bit more visible when you manipulate the photo. I've done other things where I flip things upside-down, so it's sometimes useful to be able to manipulate the photo, but obviously one has to make it clear to one's audience that they're not looking at the genuine article"

"as long as [color] is consistent from page one to the end, that's the main thing. As long as they're equally wrong. But obviously color fidelity does make a big difference to me because I am an art historian."

7.3.8 Deficiencies of digital surrogates

"Depending on where they're from, image quality. The first time I got one from here I was astounded that it was a 150 MB image and I could see every single hair on the parchment, as opposed to the 3MB we got from somewhere in the Caucasus, where I could see every single pixel--so that's one aspect. Same goes for colors, so a good trueness to color is important and similarly being able to judge the size, which with the scales that are provided nowadays you can basically do, but that isn't given in all digitized images I've seen so far. It's difficult to gauge materials at some point, especially when the metadata is left so you can't tell if the paper is glazed or not, if you don't have a different angle to look at it, or--in the case of vellum--you can't see which side of the skin it has been written on. If it's single images, especially when it comes to ornamented manuscripts, you can't always see the next page, to know whether something is on the back... Otherwise I think...seeing something on a computer screen can be a bit treacherous, in the sense that you don't always have a good feeling of how large something is. This probably is not that relevant when you're dealing with regular-sized MSS, in our C-E range, but if you're looking at enormous ones or tiny ones, or similarly with scripts where it's hard to appreciate the enormity or the minuteness of a script"

"We've decided against asking for images of the bindings, because you couldn’t see all that much of the end stitch, let's say, and it's I imagine quite difficult to get a good image of the gutter, for example...we are focusing on the inside"

"I'm always torn between whether I prefer digital or physical format. I don't like reading full books on a computer, because I find it harder to make quick notes...but they're so immensely useful if you are searching for something, and the same goes for MSS--it's harder to flick through and quickly look for something you can recognize by a glance ,but at the same time the portability or accessibility is really important."

difficulty of "trying to align multiple witnesses of either the same passage of text or image. And that becomes really time-consuming when you want to have, let’s say, I’m working with a corpus of about ten manuscripts at the moment, and you want to look at the...whore of Babylon...and see how it’s represented in different manuscripts and see it all at once. That’s something you couldn’t do unless you were in a library that happened to house all these MSS, so you can’t do that kind of consultation in any other way"--but needs to make it quicker than searching through all his MSS references

"You can’t say anything meaningful about the composition and transmission of the text without [searchable texts]." (But this creates projects for him to provide these.) "If you weren't prepared to put in that kind of work and you knew enough about philology that you would recognize what the problems are, then yeah, it would be [a problem]." Can’t answer the question of where a text was composed by talking about "one guy's edition, which is actually based on one single manuscript, so anything meaningful that can be said about it is based, if there’s any philological component to it at all, on the spellings which are contained in that one manuscript, and that’s not even representative of the textual tradition. The fact that you then don’t have transcriptions with which to compare to the points you’re making means that you’re always going to be on shaky ground”. “These arguments are being made and propagated and accepted as standard, and there's actually no basis for them whatsoever” if you do “some spade work” and go through “a number of manuscripts”. “On the basis of this very limited perspective you have people saying this text was composed in Normandy, in Metz, in Paris”
“Not many. Perhaps if a word is obscured, or in the gutter, but for textual editing it’s fine. For codicological purposes you would need to consult the original but for text I think it’s fine. But it’s more convenient, obviously you can work from home, you can manipulate the item’s zoom, magnify it to look at marginalia, rotate to look at marginalia written in a different orientation.”

“When you’re involved in a project that involves five or six translators, they may be using different editions of the same work, or referencing--that could be a problem, because a certain book could have three or four editions and two or more of those could be used by different people...it wouldn’t be a problem with translators working in the same office where they all have the same resources, but because each is in their home office or whatever with their own resources--I’m not saying it’s a big problem, because they’ll all appear in the bibliography of the finished work, but it’s just something to think about maybe.”

Frequency of problems with images/metadata: “Not often. I was puzzled earlier on this year that I wasn’t seeming to find stuff in the Bodleian and I’ve discovered since that there’d been a great hardware issue, which explained that particular thing--but no, it wasn’t urgent that I find the Bodley stuff, that was more a sort of a mild query and puzzlement”--found out about the failure when “the Polonsky blog said triumphantly on the second of September ‘it’s now back’--and I thought ‘oh, that explains a lot’” [laughs] “otherwise no. At the BnF, if there’s a problem I just try again another day, and usually it seems to sort itself out. And the Vatican likewise.” Timescale is relaxed “because I’m retired, and I’m finishing projects, and I’ve been working on them so long that one day or the next is not an urgent matter.”

not good enough image quality: “I haven’t encountered that. But then, I suspect my forays have not been hugely wide, because I’m not now interested in paleography itself, I’m not exploring scribal hands, I’m focusing on a small set of texts and I don’t have issues that rise out of those.”

quality is usually good enough, “I had problems with the National Library in Algiers. They sent me hundreds of pictures, they were very helpful, but the pictures are not very high quality so you really have to buckle down one centimeter from the screen to try and decipher them. Most of the stuff is still legible but I could have used better quality images but of course you can’t demand too much from the National Library in Algiers. It’s already a miracle that they answered my request.”

Images are generally sufficiently high quality but “What would be good is for the libraries to agree on a standard high-quality version that is to be uploaded, and not give any kind of a hint or--there are some libraries in Greece that upload their stuff, but it is very bad quality on purpose.” “How to persuade the libraries to follow a standard of high quality images?”

“But we also need to agree that when you are a scholar and you need to publish something, you have to get the permission. But permission doesn’t mean that you have to pay a hundred pounds for just a single image.”

“It’s just about thinking about a book not as a series of files that you open--you know, open one picture, minimize it, open another picture--and particularly if you’re using a laptop and you’ve not got a lot of screen space, somehow it feels a very awkward process, whereas just going from one page and clicking and it’s the next page--it doesn’t have to sort of make a rustle or something--I think that’s a bit sort of precious in a way--but it’s just thinking here’s one image, in a way, and if you click in the left-hand margin you can go back a page, and if you click in the right-hand margin you can go forward a page”

“I’m sometimes looking for things where I do need to see the original, because of erasures and things where you have to look at it in a particular light, but that’s a very specialized--most of the time it’s fine [to use the digital surrogate only]. you can sometimes make [erasures] a bit more visible when you manipulate the photo. I’ve done other things where I flip things upside-down, so it’s sometimes useful to be able to manipulate the photo, but obviously one has to make it clear to one’s audience that they’re not looking at the genuine article”
Availability of digital surrogates

Has never paid for digitization; everything has already been done that she needs. “I’m aware that it’s an issue for current students. I was talking to someone last year who’s having problems accessing materials and getting photographs, and that’s going to be amazingly expensive. Basically I think they don’t want to let him look at the manuscripts” [laughs]

“Generally speaking, if it’s an American library, I would not be sure, but maybe have a higher than 50%—like maybe I’d be happy to put a low amount of money on their having digitized—especially things like Wycliffite Bibles, because especially in America, the Bible is something that the public can be made interested in, whereas the Prick of Conscience, for example, the kind of extremely non-canonical English poems that I was working on for my DPhil, are completely unknown and therefore there’s no kind of market case for—your library Twitter account can’t get any hits by that name”

“Some American libraries in my experience have digitized everything, and some have digitized things they think people will be really interested in, or they’ve got a digitization program which began by picking kind of the big hitters, and often that will have included a Wycliffite Bible—not always. The thing about major American research libraries is they’re extremely wealthy and they don’t have many medieval manuscripts compared to Oxford, London or Cambridge”

“Sometimes I have to ask for fresh images to be taken because, you know, some of the MSS were never digitized. It depends on what you mean by digitized—sometimes I just ask for a snapshot of a page, if that can be considered digitizing a MS I don’t know. But sometimes the libraries in Morocco don’t have machinery where you can properly digitize page by page, sometimes they just go there with a crappy camera and they take a couple of pictures and send them to me.” [laughs] dozens of requests like this; only two instances of use of already digitized MSS (at Bnf and Rabat)

“Sometimes I have to pay, sometimes I don’t need to pay. And when I have to pay my department will reimburse me.” but people are generally able to give him the images

“Institutions tend to digitize the highlights of their collections, but not the down-to-earth MSS that I work with. I mean, I also have some illuminated MSS in my corpus.”

“The ones that haven’t been digitized are often the scruffy not very nice looking sources.”

“It’s a real mixture. I guess of the ones in Paris half are scanned microfilms and the other are digitizations in full technicolor,” thinks all digitizations were done before he started this project; had some microfilm from related projects as well; “I don’t think any special work has been done”

problems with digitized images? “As I mentioned, the Bodleian Library, that was the technical failure.” For undone Baroccis, ”still, they could have uploaded the microfilm. And perhaps with certain software we can get a kind of a full-color image, I assume.”

has taken 21K photographs herself of manuscripts, has them in a private database, and “as far as it’s possible with the licensing libraries I’ve put the things online [on Flickr]”; “the American libraries have all put them online on their own website, which is ideally what I want, so it’s not me holding but rather just linking through to the professionally digitized” (personal-professional ratio: 10 to 1)

“Those that have been digitized have only been digitized because--well, I paid for the Oxford ones, the two in Harvard were financed by the Houghton Library itself, and the one in Michigan I contributed a fraction of the cost. The ones in Goetting they digitized for me two manuscripts and did it at a very good price. But otherwise I take all the photographs myself.”

Good availability of digitized books on Google and EEBO, although Google Books “drives me mad. It’s unhelpful and inflexible and then when you find the book you can’t cut and paste from it, which is fantastically irritating, and they say this page is not available and you say why not--continental printed books
is a much more hit and miss thing. Sometimes you can find continental printed books reproduced well on library websites. Sometimes you can find them by googling the date and you'll find something. But they have not got nearly as good a sort of bibliographically controlled area as British books before 1800 have got. So it's hit and miss. And you can use things like the USTC and you can sometimes find images on that, and they send you off to a library and then the library says no you can't look at this, which is pretty annoying.

"It depends on the institution, but I'm finding that as I go along more and more MSS are being digitized. But then there are I would say notorious non-digitizers, like some libraries say in Italy for example. So maybe if I wait another twenty years, but I can't really do that. So it's a very uneven picture."

Hebrew versus other kinds of MSS: fewer Hebrew ones out there to begin with, but there have been Hebrew-specific digitization funding sources. Hard to compare and "in flux. Everything is changing so rapidly."

7.3.10 Citation of digital surrogates

"Only where I think it makes a difference. So if I'm referring to the content then probably not, because I figure that whatever I can read on a digital copy I can read in real life. Where it comes to material aspects I wouldn't necessarily say that this was viewed digitally or something like that, but I might add that as a side note that I have seen this online and I haven't touched it...I think you would probably mention it if it were a poor quality image, where there was a certain amount of uncertainty, or if it was black and white, so an ancient digital image, but otherwise I don't think it makes a difference... [this is standard in the field] inasmuch as there are digital images, yes, I think so. The reference to MSS has only been codified say ten years ago, so prior to that it was a little haphazard what abbreviations people would use for various locations"--there was a seminal article that listed the terms they should be using

"Original. That was always going to be the first port of call. Then if it's available online I will include the URL, but the--I don't tend to feel that the precise mode of consultation is an essential thing. [in the case of this specific project] I will definitely be referring to the website which will allow people to look at this and other manuscripts, but that's a self-promotional thing. [laughs] If it's not that component I don't see it as an absolutely essential thing. But it is quite nice if you can point somebody towards the resource if they wish to substantiate the claim or check the claim themselves."

"for this project the only online editions we're citing are I think the Encyclopedia of Islam. All other things we're citing the physical edition...primarily because it's going to be a printed book, the end result, and if I was to write an article and cite the electronic resource I would put the URL and date accessed, and there's no point doing that for a physical printed book really. If it was an online edition that would be fine, so you could click through to the primary source. And that would be far too cumbersome for a printed edition...it would take up a lot of space if you're putting the URL, which can sometimes be long, plus the date accessed in every instance."

"It's very useful to have a bibliography with links to the articles."

"I would assume that at some point I had seen the original, so I would give the reference to the original. I would also probably give the finding reference for the digitized version, because that is now so easy. One used to refer to facsimiles, if you wanted to bring home some precise point, but now increasingly you will--I will put a URL for a digitized copy of a manuscript page."

"I'm not referring to MSS by URL in anything I publish, although that would be an interesting idea, and it's something I've considered. One issue with doing it in print is that most URLs aren't very typing friendly. If there was a way to get a URL shortener URL for a stable URL for a MS that would be nice."

"I don't think it's a common practice. I adopted it from [']s PhD thesis, just because it seemed like a good idea, because it's good to indicate when you've used a digital resource, and it's good to indicate that these things are available. Of course it doesn't work to do that because a lot of things I did consult online but I also did consult them personally, and therefore you can't tell that they're online by reading the bibliography. There's a
there's a question, isn't there, around, you know, if you consult a MS online, even if you then consult it physically, should you still in some way cite the online version, because it is a new object, and therefore you know it's—it's worth acknowledging the effort to kind of create it. I know the people who run is it British History Online? It's a huge searchable digitized collection of a lot of primarily 19th-century records like calendared and gazetteered will compilations...they're very keen for people to cite them rather than the text that they took the information from, so they have stripped out page numbering from the information they present, so you can't know where in the physical book you are, so you're forced to--unless you're me, and you just go to Duke Humphreys and look at the physical book. There's a little page on their website where they explain it--not at all aggressively--and they have a reasonable case, certainly if I used them they'd probably be in the bibliography, but I don't know, and once you've consulted the physical book, that kind of supersedes the--because there are keying errors in British History Online, which is the other reason I consult the physical book”--also EEBO: “similar qualms about the extent to which they're actually--there's a vast amount of unacknowledged work by early modernists where really they've looked at the digitized version of a microfilm version of an early printed book, they haven't really handled--and sometimes the copy on EEBO isn't particularly typical of that printing, so there are issues around that”

manuscripts are easier because “there's only one copy existing in the world” so “you're not obscuring that step, if you cite a manuscript, the digitized version is only ever going to be a digitized version of that manuscript, it's not going to be a digitized version of five other--”

would cite the original if saw both, and mention that it was digitized; would cite the digitized version if saw only digitized; that's just how it's done; “in the footnote I will mention that I have received these images from this senior curator in this library, and I will include reference to the images”

"Well, I actually just edited a book...where what we did is we included links to the online version of everything rather than--so we'd put the folio number of the original in the main text and then in the footnote there'd be a link to the image using the stable URL and we just have to hope the stable URL is really stable. And we hoped that in the PDF version of the book those links would all be hotlinks, and then they turned out to be flat links, so they weren't hotlinks, so we talked to them about it and I think if you buy the enhanced version of the book it now has them as hotlinks, which was our original design, and we also at one point wanted to put up a list--because the book has got a list of all the MSS and the URLs of the online versions”

"I'd consult whatever style guide I'm using I guess. I'd always cite the original, and--it would depend on the type of article or output in that if I was referring to a lot of MSS in passing then I would probably just cite the original, currently at least. If it's, I don't know, an in-depth study of a particular MS, I would certainly refer to the digitization as well. But I guess I'd follow whatever the style guide is telling me to do.... I guess if I didn't have any guidance my default would be--I don't know whether this is right or not, but my default would be to use whatever in-house way of--so however the institution cites its own MSS, I'd follow their practice.”

"I think there are a lot of assumptions made, in that we assume that the original has been consulted, but a lot of extensive use has been made of the digitization as well, unless someone says to the contrary, maybe? I can think of articles where the author said that they had not been able to consult the original. But that’s sort of viewed negatively in a way, as a lack, whereas perhaps we should be--yeah, I suppose we should be becoming more specific about what we have actually consulted, making that clear. But I just wonder how, you know, if you're listing a long list of MSS, for whatever reason, in your article, how practical it really is to include links to every single one. And that's where having a list or a catalogue of MSS that have been digitized would be really useful, because then you could just point to there, I guess.”

"Both. And you have to declare and say clearly yes I've seen the digital image plus the original, or just the digital image.” URL: “Not always. Depends. Whether it's a permalink or not. Or if it is really a permalink. Because I was quite cross, I didn't like the idea of BL changing the way they've been presenting the MSS. It wasn't entirely clear whether they changed also the permalinks or not.”
common permalink problems? “It’s not very common to have this permalink in other research.” other people
don’t give the URL “not as far as I know. Maybe I’m wrong,” do they specify digitized version? “Usually yes.
That’s our common practice. It’s what I was taught and what I did. Yeah I remember when I prepared my
critical edition, I had to put a list of all the MSS, almost 40, give the date, and the way--in situ, CD-ROM, or
URL. Back then there was no list of my MSS online, so I had either CD-ROM or in situ or perhaps a
reproduction on paper of the microfilm section.”

“I now would cite the digitized one, although I think that’s--I think in the past I would have sort of elided the
digital one into the original copy, partly out of a kind of feeling that I wasn’t quite sure about how to cite a
digital copy, and partly because maybe because of a feeling that it wasn’t quite so good to have consulted a
digital copy rather than the original. I don’t feel that second thing anymore, I think it’s important that we--I
think in a way lots of colleagues who don’t have EEBO, so American colleagues who don’t have EEBO, find it
really hard to get their institutions to subscribe to EEBO because if you look at EEBO in terms of citations, it
looks very low. So it looks as if nobody uses it, but of course absolutely everybody uses it, even if you’ve got
an amazing rare book library across the road as I have, but if you haven’t--so it seemed to start to be kind of
an ethical point to say this is a resource that we’re all using.”

“Somebody told me that the MLA Guide 2016, or the most recent one, was almost silent on how to cite a
digital copy of a book rather than sort of a website. It could grasp that you might cite a website or a web
address, but not that a digital copy of a book might need to be cited. So I guess the practical side, we haven’t
completely sorted that out. You know, how do you do that without it looking really messy. Because those very
long URLs that are often--they do look messy, don’t they, especially if your program automatically underlines
them or breaks them in a funny--doesn’t want to break them, and screws up the justification. It just seems
sort of messy.”

“I’ve always thought that’s quite a weird thing in a print book, to have those long references... I would have
thought nobody would follow the links. What you would do is you would Google University of Stuttgart First
Folio.”

“Something we’ve been discussing extensively...my preference has been to give the old-style shelfmark plus
the domain name of the library where you can find the digitized object, because that’s pretty sure to stay the
same and should normally be enough to find it on that--if you have the shelfmark that will always be part of
the metadata, while if you give these permalinks and people have to type them in, it’s bound to go
wrong...what I’ve been doing also is if I’m doing the print-ready copy, then I would give really the domain
name, but would have as hyperlink behind it the permalink...I’ve done several open-access books and for the
printed version you really want to have a short memorable link, but if possible you also want to have
something clickable, and to bring you not to the top level but directly to the item. But you want as a safety net
also the shelfmark, so if the link is broken because the library has decided to move all of its digitization to
another site, you can still find it via the shelfmark”

“I think I would probably cite the website, or both. Because I talk about copies for this sort of thing--you can
see this in the Folger copy and you can see the Folger copy here”--finds this strategy common in others’ work.
“I think you have to, when you’re referring to something, let the reader know that it is available online in one
form or another, yeah, no, I think you need to do that”

7.3.11 Downloading

“I think it would be nice to be able to download them. Probably impracticable in the sense that if you want
high-resolution images you’re going to end up with a few GB per manuscripts, but if you’re someone who’s
working with ten manuscripts because you’re doing a critical edition, then it’s nice to be able to take them to a
beach or a cafe where there isn’t reasonable internet, and given how academics work these days, that’s
probably what some of them do.”
“something which is very useful is if you can then download the images in some way shape or form in a
decent resolution...so sometimes you’re going to be working offline, and sometimes you’re going to want to
have multiple instances open, and it’s a lot easier to do that if you actually have the files in your own
repository and then use the same method for consulting them--as opposed to what I have now, where you’ve
got a browser window here, a PDF window or multiple ones open, some kind of image viewer here--it gets a
bit of tiresome setting it all up”

“I like to have that option. I wouldn’t say I download everything, but if I think it’s going to be useful now or in
the future.”

Too large a file size: “I suppose that would be weighed up against how much you need or like the resource.
Sometimes I would download some perhaps huge--like for example from this History of Medicine website,
there’s a Latin translation of [], and if you’re interested in 2 books out of the 8 books you can just download it,
use a PDF manipulator to extract pages you want and discard the rest, if it’s really huge.”

“Yes, and one can very often. Certainly you can download from the BL, and that is just incredible, to have
decent images at once on your desktop.”

“there aren’t very many reasons why I would want to take something off the internet that’s already hosted
somewhere and then host it myself. If I was writing a blog post about a MS I could just link to it if it’s publicly
available. If it isn’t publicly available I wouldn’t want to risk hosting it.”

not downloading “with most modern repositories, because they serve the images up fast. With older repos
yeah, it’s tempting to do it. Parker on the Web is a good example, it takes a while to load, it also takes a while
to navigate, but they’ve also made it so it’s quite difficult to download the images

understands there’s a concern about people who ”suck the repository into a hard drive and host it
somewhere else”; “in one sense there are perhaps scholars who would say if someone did that does it matter,
but I could see a whether or not it matters it not necessarily the same depending on who you are, and b the
Parker Library probably wants to know who’s reading their manuscripts, as a paid service that has to sustain
themselves, etc etc”

can’t download from DIAMM, main con; “so I tend to do illegal screenshots, which is what everyone does”. “I
often just need to have a reference picture that is easily accessible to me, and particularly be able to put
several different versions of a song side by side on a table. A physical tabletop. And you can’t really do that on
screen--I mean, I could have four different screens and blah blah but actually having a physical printoff of a
file is really valuable...” “So Gallica is really valuable because you can print the pages. A lot of German libraries
also let you print the pages, and like Gallica let you download the whole MS in a slightly lower resolution PDF,
which is really useful for scholarly purposes. It just means that there’s no reason for me ever to have to go to
these libraries, because I’m not doing the kind of work where I have to look at binding structure or anything
in huge detail”

“Very rarely, because--well especially I think you can download entire documents but it’s at very poor
resolution, so that’s not going to work. But maybe for particular images, say, but again that tends to be--so
just details of particular folios, I might download those. But not so often, no.”

download and then look at images on his computer: appreciates being able to “put them onto Powerpoint,
store them so that I don’t constantly have to look at them...of course when you’re lecturing, I try to make
myself do the PP slides as I go along, but I usually fail”

downloading: “one way or another, yes. I’ve got better at it. So when I was first doing PP slides I found it very
difficult, and the more you do the easier it gets, and there are some places that I know are difficult to
download material from but I’ve worked out ways of doing it, although as I say I’m by no means skilled at
using this sort of thing. I just work it out by myself, trial and error”
“I’d cite the original and then I’ll give a link to the digital version and pray that it will stay up-to-date.... I mean
the major libraries do their best”

7.3.12 Social media
“occasionally to point out that they’re there to other people. Not to find except for serendipitously, like if
someone says we’ve just digitized this, or I’ve been looking at this really interesting thing, I might--yeah. It’s
only relatively recently that I’ve been doing such a high proportion of my work using digitized MSS, of
course.... There used to be quite a lot of tweets pointedly saying this isn’t digitized--‘Today I’m looking at this,
which you’ can’t look at online’. But yeah, I guess I've used it to find things but I haven’t searched SM for a
particular manuscript. I do have a running Tweetdeck permanent search column for the WB, for instance, and
for the Prick of Conscience and for a couple of other things.”

social media “has become too large and unspecialized to be useful, and yet not global enough to be useful
either. So I can’t see myself using social media to find stuff. I don’t really use [forum] to find stuff either”--
become too general; “pretty much it has like 1500 or 2000 members, and this is all the medieval
musicologists in the world, really nearly everybody, and then you were divided into groups [by
specialization]...and people posted things about performances, about stuff they’d found, questions, and so--
but then it became too big. So then you get people who are basically just interested amateurs sort of asking
dumb questions. Which I don’t have a problem in a way with that but it makes it a different kind of thing. It
makes it more like the Wikipedia chat room rather than something very specialist. So a lot of the specialist
discussion then moved onto Facebook, onto a Facebook discussion group, because membership of that was
moderated”

would use Twitter to ask around “in fact I nearly did that yesterday. And quite often a lot of people do do that
in the accounts I follow, and that’s how I also hear about things that are there.”

“There is however one exception, which is a group on Facebook called Hebrew Paleography and Codicology,
which has people like me writing in to say does anybody know how to--so occasionally I’ll do that, and people
will help each other. So that’s quite a specialized interest group.”

7.4 APPENDIX D: LIST OF TRANSCRIPTION CODES

citation
colleague resource sharing
databases/directories
desired discovery system
desired metadata
desired viewer functions
digital vs. physical
digitization advantages
digitization availability
digitization inadequate
DIY digitization
downloading
Google
lack of confidence in own methods
librarians
metadata problems
microfilm
online catalogues
opinions on discovery systems
opinions on viewers
others' behavior
permalinks
print catalogues
resource discovery process
role of digitization in research
serendipitous encounter
shelfmark
social media
students
university/library website
unreliability of websites
7.5 Appendix E: Participant Information Sheet

Participant Information Sheet

Emma Stanford, City University London

Name of Study: Understanding resource discovery and use habits among users of digitized special collections

I would like to invite you to take part in a research study. Before you decide whether you would like to take part it is important that you understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information.

What is the purpose of the study?
This study, which I am undertaking with the support and cooperation of Bodleian Digital Library Systems and Services, will form my master’s dissertation in Library Science at City University London. The aim of the study is to build a clearer picture of how researchers learn of digitized books and manuscripts, how and why they access them, and what features and options they need the digitized resources to have. The study, including analysis and write-up, will be completed in December 2016.

Why have I been invited?
You have been invited because you make use of digitized special collections items—books, manuscripts, maps etc.—in your research. The total number of participants in the study will be between 10 and 20.

Do I have to take part?
It is up to you to decide whether or not to take part. If you do decide to take part you will be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. You will not be penalized or disadvantaged in any way for withdrawing from the study. Similarly, if there are any questions you do not feel comfortable answering, you may choose not to answer them, and you may retract your answers at any point. Participation or lack of participation in the study will not affect your grades (where applicable) or your treatment by Bodleian Digital Library Systems and Services.

What will happen if I take part?
Your participation will take the form of an in-person interview. The interview is designed to be undertaken in a single session of about an hour, but if you would prefer, they can be split into multiple sessions. The meeting(s) will take place in a library meeting room or in your office, as you prefer.

When we meet, I will explain the background and aims of the study in more detail and answer any remaining questions you might have. I will then ask you a few questions about your research background and preferred methods of gathering information and discovering resources, before diving into the main interview, where I will ask you questions about how your use of digitized special collections resources. The interview will be loosely structured, so you will be able to expand on any issues you believe to be important. I will be recording what we say during the interview using a voice recorder, and once it's done I will transcribe the recording and, if you desire, send the transcript to you for review.

At the end of the interview, you will have another opportunity to ask questions.
What do I have to do?
I will ask you to give honest answers to my questions about your research strategies, habits and preferences. You will not have to share any information you don’t want to share, or answer any questions you don’t want to answer. There are no wrong answers.

What are the possible disadvantages and risks of taking part?
One certain disadvantage is that the interview will take up your time. I will minimize this drawback by confining the interview to a single session (unless you would prefer otherwise), and I hope to offset it through the benefits of taking part (see below).
There is also a small risk that, given the small size of some research communities, your identity may be guessable from the details provided in my analysis. I will minimize this risk by excluding personal details, discussing academic background separately from resource discovery methods and preferences, and checking with you to ensure that your identity is protected in the discussion of any subjects you feel are particularly sensitive.

What are the possible benefits of taking part?
Once the interview is done, I will be happy to offer advice and instruction on how to make use of digitized resources, and to answer any questions you might have about the Bodleian’s digitization projects and plans.
On a broader scope, the information obtained in the study will feed into a series of recommendations to ensure that Bodleian digitized resources are accessible, effectively publicized, and easy to use. These recommendations may also be circulated to other institutions. By participating in this study, you will be helping us to improve the resources we provide to you and other researchers.

What will happen when the research study stops?
Once the interview is done, I will be happy to offer advice and instruction on how to make use of digitized resources, and to answer any questions you might have about the Bodleian’s digitization projects and plans.
On a broader scope, the information obtained in the study will feed into a series of recommendations to ensure that Bodleian digitized resources are accessible, effectively publicized, and easy to use. These recommendations may also be circulated to other institutions. By participating in this study, you will be helping us to improve the resources we provide to you and other researchers.

Will my taking part in the study be kept confidential?
I will make every effort to ensure confidentiality. Once I have transcribed the recording of your interview and obtained your approval of the transcript, I will destroy the original recording and eliminate any personal details from the transcript. I will also assign you a random number that I will use instead of your name when I analyze the data. In the course of analysis, I may share transcript excerpts with my dissertation supervisor or with my colleagues, but I will ensure that no identifying details are included in these excerpts. The finished dissertation will be deposited into the City University dissertation repository, at which point the interview transcripts and notes, up to now stored on my computer and on Dropbox, will be destroyed.

What will happen to results of the research study?
The study will be finished in December 2016 and submitted in January 2017, and deposited into the City University dissertation repository. It is possible that the study will be adapted into a shorter form for publication in a journal; if it is, the same standards of confidentiality and anonymity discussed above will obtain. If you would like a copy of the finished study, or details of any future adaptations, you may ask me at any time.

What will happen if I don’t want to carry on with the study?
You may decide at any point during the study (up until the study ends in December 2016) that you don’t want to carry on with it. If you do, simply tell me and I will destroy the data from your interview. You will not need to provide an explanation for your withdrawal, and you will not be penalized in any way for withdrawing.
What if there is a problem?

If you have any problems, concerns or questions about this study, you should ask to speak to a member of the research team. If you remain unhappy and wish to complain formally, you can do this through the University complaints procedure. To complain about the study, you need to phone 020 7040 3040. You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the name of the project is: Understanding resource discovery and use habits among users of digitized special collections.

You could also write to the Secretary at:

Anna Ramberg
Secretary to Senate Research Ethics Committee
Research Office, E214
City University London
Northampton Square
London
EC1V 0HB
Email: Anna.Ramberg.1@city.ac.uk

City University London holds insurance policies which apply to this study. If you feel you have been harmed or injured by taking part in this study you may be eligible to claim compensation. This does not affect your legal rights to seek compensation. If you are harmed due to someone’s negligence, then you may have grounds for legal action.

Who has reviewed the study?
This study has been approved by City University London Research Ethics Committee

Further information and contact details
Researcher: Emma Stanford, emma.stanford@city.ac.uk
Supervisor: David Bawden, d.bawden@city.ac.uk

Thank you for taking the time to read this information sheet.
7.6 APPENDIX F: PROJECT PROPOSAL
Understanding resource discovery and use habits among users of digitized special collections

Introduction

Last year, I requested a Bodleian manuscript to a reading room, as it was in the process of being digitized and I needed to check a few of the images against the original. When I opened the manuscript, I found a note from a researcher using the Bodleian, who was excited that I appeared to share his interest in an obscure 16th-century Hebrew manuscript. He was surprised to learn that the Bodleian was undertaking a massive Hebrew manuscript digitization project, and he asked me to notify him when his manuscript was available online.

This incident points up the question of how traditional models of special collections resource discovery—in which books and manuscripts are discovered by consulting printed catalogues, subject specialists or literature citations—translate to the world of digitized special collections, where requests are less localized and available resources are constantly changing. Do digitized special collections resources adequately support traditional models of resource discovery? Do they foster new models of resource discovery that are better suited to their strengths and weaknesses? To answer these questions, we must know more about how researchers discover and interact with digitized special collections: how they find them, what they do with them, and what they would like to be able to do with them.

This project will investigate the discovery habits and resource requirements of researcher users of digitized special collections through a series of interviews, with a particular focus on the users of the Bodleian’s new digitized special collections resource, Digital.Bodleian.

Aims and objectives

Through a series of interviews with users of digitized special collections, conducted in person (or if necessary by email or webchat), the project aims to create a picture of how researchers learn about the existence of digitized items, how and why they access them, and what interface features and functions they need or want. For example, I will try to find out what channels researchers use to learn about digitized resources (e.g., social media, forums, library websites, word of mouth), what research purposes they consider digitized resources to be useful for, how much time they spend viewing digitized resources, what viewing interfaces they consider to be the most useful, and what functions they wish these viewing interfaces had. If possible, the in-person interviews will include a user testing component, where I watch the interviewee locate and view specific digitized resources. The objective of my research is to produce a chart or set of flow charts of how researchers find and navigate through digitized special collections.
collections, and a series of recommendations for how to present digitized special collections resources so that researchers will be able and motivated to use them. These recommendations will be pitched toward the creators of Digital.Bodleian and similar large-scale institutional special collections image repositories. The recommendations may touch on such questions as where to advertise the collections, what finding aids to provide, how important it is to integrate digitized manuscript records with existing catalogues, and how to modify existing viewing interfaces to better suit the needs of researchers.

**Scope and definition**

For the purposes of this research, “digitized special collections” are defined as online collections of high-quality images of rare books and manuscripts, provided by libraries and potentially also by museums. These collections may take a variety of forms, from searchable websites to Flickr collections. Furthermore, they may be maintained by institutions across the globe, as geography is not a barrier to digital collection use. Large-scale microfilm imaging and transcription projects such as EEBO-TCP are excluded from this definition, however, as the focus of my research is not on data or text mining but on digital visual representations of special collections objects, of the type currently being produced in large quantities by the Bodleian and other libraries.

This project defines “researcher” loosely; the subjects of my investigation may be professors, postdoctoral or doctoral researchers, or students in taught undergraduate or graduate programmes (most likely at the end of these programmes, as early taught students tend not to use digitized primary sources). Casual, amateur or commercial users of digitized special collections are outside the scope of my research, however. Researchers who do not use digitized special collections are also out of scope. Geographical limitations mean that my in-person interviews will be conducted with researchers who work at or near the University of Oxford, but I may also engage with international users through an email questionnaire or webchat interview. Because special collections resources are mainly used by humanities researchers (and, particularly, palaeographers and art historians), most if not all of my study participants are likely to be humanists, but I do not plan to intentionally exclude other disciplines.

Although the research objective is to produce a set of recommendations that can be taken into consideration by the creators of Digital.Bodleian and other Bodleian Libraries and University of Oxford resources, the scope of the research will not be limited to users of these resources. In fact, in order to get a sense of how discovery and viewing strategies vary from one resource to another, the interviews will be designed to allow subjects to compare different institutions’ resources. The final recommendations, however, will focus on actions that can be taken by the Bodleian, as my knowledge of the Bodleian’s digitized special collections resources will allow me to recommend feasible courses of
action, and through my own work and connections I may be able to put some of the recommendations into effect.

Research context/literature review

In the last decade or so, improving digital camera technology, low-cost data storage and widespread use of the Internet have made it possible for libraries to make increasingly high-quality images of special collections items freely available to users worldwide. Libraries are currently undertaking digitization projects of this nature on a large scale. The British Library, for example, has digitized 600 Greek manuscripts alone; the Bodleian and Vatican Libraries have recently collaborated to digitize 1.5 million images of their manuscripts and early printed books, and the Vatican Library has further plans to digitize its entire manuscript collection.

Generally speaking, libraries have tended to adopt an “If you build it, they will come” attitude to special collections digitization, postponing the creation of rich metadata and centralized finding aids in order to focus on simply getting the images online. (See, for example, the low-metadata digitized collections produced by the Bodleian and Vatican Libraries for the Polonsky Foundation Digitization Project). It is becoming clear, however, that libraries need to start tailoring their digital offerings to the needs of their users, especially as library digitization projects are often funded by philanthropic organizations that need to see that the resources they are paying for are being used.

To satisfy their funders, libraries often use web analytics or surveys to show that a wide range of users are accessing their digitized resources, and they often turn to social media to publicize their offerings (Kelly, 2014). Schlosser and Stamper (2012) demonstrated in their comparison of Flickr and institutional image repositories that social media promotion is key to building a widely used image library, but within the academic community, searchability, metadata quality and most of all interface usability are crucial to ensuring that a particular digital library is viewed as a valuable resource (Xie, 2006). General resource discovery studies tend not to consider digital special collections holdings, however (see Madsen, 2015, in which the list of resource types includes only “printed books and journals, modern papers and archives, manuscripts, museum collections (objects and works on paper), e-books and e-journals, data sets, open access materials, pre-prints, and computer code” (p. 9)), and where digital librarians have conducted their own studies, their focus on on assessing the value of their own offerings (Kelly, 2014) means that they tend to start with the resource rather than the researcher. By interviewing researchers about their resource discovery habits across institutional collections, and by focusing specifically on digital special collections resource discovery, I hope to fill these gaps.
The newness of digitized special collections resources, and the rapidity with which existing resources are changing, make for a relatively sparse research landscape. I have found a few highly relevant resources from which I intend to work outward, however, starting with Elizabeth Joan Kelly’s “Assessment of Digitized Library and Archives Materials: A Literature Review” (2014), which surveys existing research into the usability and popularity of digital libraries. I will also draw on Hong (Iris) Xie’s research (2006), in which she asked digital library users to create their own set of evaluation criteria (including such facets as “collection quality”, “user opinion solicitation” and “system performance efficiency”), which they then used to evaluate 18 American digital libraries. Closer to home, Athenaeum21 and the Bodleian Libraries produced a report last year on Resource Discovery @ the University of Oxford, which incorporated literature reviews and interviews of users and providers of information resources at Oxford. As mentioned above, digitized special collections are notably absent from the list of resource types discussed in this report, but the literature reviews and analysis will be useful resources.

Starting with these literature resources, I will follow citations to find other resources on the subject of digital library use and discovery. I will also consult the journals D-Lib and Journal of Web Librarianship, among others, to look for related research. I will also consult more traditional library journals to look for studies of manuscript and rare book resource discovery behavior, in order to be able to compare and contrast digital and analogue discovery methods. Finally, I will look for blogs and white papers by library staff, which may provide a more up-to-date perspective on the resource discovery strategies supported and observed by libraries.

Methodology

The project will take a qualitative approach to research, structured around a set of semi-structured interviews. Because I am interested in obtaining a detailed picture of how researchers use digitized special collections resources, an in-depth interview is more appropriate than a questionnaire, and the interview must be structured flexibly enough to allow for participants to guide the conversation toward things they think are important. I anticipate that after conducting a thorough literature review I will have a clearer sense of how many subjects I will need to interview in order to make my research balanced and credible, but my initial plan will be to interview as many subjects as possible, as I expect that I will be limited more by the number of willing subjects than by restrictions on my own time and resources. I would hope, however, to conduct between 10 and 20 in-person interviews.

Where possible, the interviews will take place in person, possibly with an additional observation component in order to better understand the subject’s process for finding and viewing digitized resources. However, I anticipate that a large proportion of potential subjects will consent to an email
questionnaire or webchat interview but not an in-person interview. Once I have conducted an initial set of interviews, I will evaluate the need to supplement further interviews with electronic questionnaires or interviews. This option has the additional advantage of potentially broadening the pool of subjects to include those who are not based at the University of Oxford.

For both the in-person interview and the email questionnaire option, the main difficulty I anticipate is recruiting subjects. I am aided here by the fact that I work for the Bodleian Libraries’ digital team, and I will be able to use my connections to find subjects. I plan to investigate the possibility of providing incentives for in-person subjects, in the form of gift vouchers or similar, which my employer may be willing to pay for, as my project will aid its effort to increase user engagement with digital collections.

It is possible that subjects may feel uncomfortable speaking negatively about Digital.Bodleian and other Bodleian resources in my presence, as they are likely to know that I work on them. To be honest, this is not a phenomenon I have encountered much in the past; researchers are generally frank in criticizing the resources that are available to them. Nonetheless, I will make an effort to explain to each subject that I am interviewing them in my capacity as a researcher rather than as a Bodleian employee, that I welcome criticism of Bodleian resources, and that they will not be penalized for anything they say.

Once I collect the interview data, the structure of my analysis will be determined partly by ideas obtained from my literature review and partly by the trends that become evident in the data I collect.

Work plan

My research schedule will be somewhat cramped by the summer vacation, which extends from mid-June to early October and during which many students are away from campus and many later-career academics are difficult to reach. I plan to conduct my literature review in early summer, then to design my interview plan and questionnaire, with the aim of conducting a first round of interviews in August and September, before term starts. Depending on the volume of responses I get during the summer, I will most likely need to conduct another round of interviews in October, once more people are back on campus. November and December will then be devoted to writing up my research and recommendations on a compressed schedule. Therefore, the timetable will look like this:
### Resources

The basic equipment I will need will be my laptop computer and an audio recording device for interviews, should I choose to use a recorder rather than simply taking notes on my computer. (I will investigate using the voice recording option on my phone, but if this is of insufficient quality or has insufficient storage, I may be able to borrow a recorder from my work.) Because I will not be processing large amounts of quantitative data, I won’t need any software apart from Word for writing and note-taking and RefWorks for compiling references. I have institutional subscriptions to both of these.

I anticipate being able to conduct most of my interviews at the library or in researchers’ offices, so I don’t expect to incur any travel costs. The only significant potential cost would be incentives for interviewees in the form of gift vouchers or similar, which my department may be able to cover; if not, I will probably do without them.

### Ethics

Because I will be interviewing postgraduate and late undergraduate researchers, I do not anticipate that any of the participants in my research will be vulnerable or unable to give informed consent. Furthermore, while use of high-quality online images is a potentially sensitive subject, as there have been cases where users download and stitch together image tiles in violation of the image delivery websites’ terms of use, I will steer clear of this subject in designing my interview plan, as it is not strictly relevant to my research questions (which are focused on how researchers discover the resources and what viewing options they would like to have). Participants may also wish to keep certain details about their research habits or research backgrounds private or confidential, so I will explain to them the purpose and scope of my research, discuss confidentiality measures, and obtain their consent in each case. The completed ethics checklist, consent form and participant information sheet are attached.

<table>
<thead>
<tr>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature review</td>
<td>Literature review</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design interview plan &amp; questionnaire</td>
<td>Design interview plan &amp; questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruit participants</td>
<td>Recruit participants</td>
<td>Recruit participants</td>
<td>Recruit participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduct interviews</td>
<td>Conduct interviews</td>
<td>Conduct interviews</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyze results</td>
<td>Analyze results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write up</td>
<td>Write up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The completed ethics checklist, consent form and participant information sheet are attached.
second participant information sheet will be needed if I choose to conduct email questionnaires as well as in-person interviews.

Confidentiality

I will keep the identities of my interviewees confidential, identifying them by randomly assigned number, if at all, in my analysis. However, I do plan to ask each interviewee questions about their research background and interests, so that I can identify any needs or habits specific to a particular discipline or research level, and because some research communities are very small, including these details in my analysis may give away the identities of certain researchers. If it becomes evident that this will be a problem, I will investigate options to preserve confidentiality. For example, I might omit the potentially identity-revealing information from the final report, or speak in general terms where necessary so that the identity of a single participant can’t be traced from one section to another.
References


Participant Information Sheet

Emma Stanford, City University London

Name of Study: Understanding resource discovery and use habits among users of digitized special collections

I would like to invite you to take part in a research study. Before you decide whether you would like to take part it is important that you understand why the research is being done and what it would involve for you. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information.

What is the purpose of the study?

This study, which I am undertaking with the support and cooperation of Bodleian Digital Library Systems and Services, will form my master’s dissertation in Library Science at City University London. The aim of the study is to build a clearer picture of how researchers learn of digitized books and manuscripts, how and why they access them, and what features and options they need the digitized resources to have. The study, including analysis and write-up, will be completed in December 2016.

Why have I been invited?

You have been invited because you make use of digitized special collections items—books, manuscripts, maps etc.—in your research. The total number of participants in the study will be between 10 and 40.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. You will not be penalized or disadvantaged in any way for withdrawing from the study. Similarly, if there are any questions you do not feel comfortable answering, you may choose not to answer them, and you may retract your answers at any point. Participation or lack of participation in the study will not affect your grades (where applicable) or your treatment by Bodleian Digital Library Systems and Services.

What will happen if I take part?

Your participation will take the form of an in-person interview, with the optional addition of a brief observation session, in which you will undertake two or three simple resource discovery tasks. The interview and observation session are designed to be undertaken in a single session of about an hour, but if you would prefer, they can be split into multiple sessions. The meeting(s) will take place in a library meeting room or in your office, as you prefer.

When we meet, I will explain the background and aims of the study in more detail and answer any remaining questions you might have. I will then ask you a few questions about your research background and preferred methods of gathering information and discovering resources, before diving into the main interview, where I will ask you questions about how your use of digitized special collections resources. The interview will be loosely structured, so you will be able to expand on any issues you believe to be important. I will be recording what we say during the interview using a voice recorder, and once it’s done I will transcribe the recording and send the transcript to you for review.

After the interview, I will ask you to complete several short tasks on a computer (e.g., looking for particular resources or images on a library website). I will not record audio or video for this section; I will simply take handwritten notes.

At the end of the interview, you will have another opportunity to ask questions. Once the interview is done, I will transcribe the recording and send it to you for review and approval.
Expenses and Payments (if applicable)

[This section to be completed once I know whether I am able to offer any rewards or incentives]

What do I have to do?

I will ask you to give honest answers to my questions about your research strategies, habits and preferences, and then, if you choose, to complete a few simple resource discovery tasks, trying to approximate what you would do in a normal unobserved research situation as closely as possible. You will not have to share any information you don’t want to share, or answer any questions you don’t want to answer. There are no wrong answers.

What are the possible disadvantages and risks of taking part?

One certain disadvantage is that the interview will take up your time. I will minimize this drawback by confining the interview to a single session (unless you would prefer otherwise), and I hope to offset it through the benefits of taking part (see below).

There is also a small risk that, given the small size of some research communities, your identity may be guessable from the details provided in my analysis. I will minimize this risk by excluding personal details, discussing academic background separately from resource discovery methods and preferences, and checking with you to ensure that your identity is protected in the discussion of any subjects you feel are particularly sensitive.

What are the possible benefits of taking part?

Once the interview is done, I will be happy to offer advice and instruction on how to make use of digitized resources, and to answer any questions you might have about the Bodleian’s digitization projects and plans.

On a broader scope, the information obtained in the study will feed into a series of recommendations to ensure that Bodleian digitized resources are accessible, effectively publicized, and easy to use. These recommendations may also be circulated to other institutions. By participating in this study, you will be helping us to improve the resources we provide to you and other researchers.

What will happen when the research study stops?

Once the study is complete and the dissertation approved, I will destroy the data.

Will my taking part in the study be kept confidential?

I will make every effort to ensure confidentiality. Once I have transcribed the recording of your interview and obtained your approval of the transcript, I will destroy the original recording and eliminate any personal details from the transcript. I will also assign you a random number that I will use instead of your name when I analyze the data. In the course of analysis, I may share transcript excerpts with my dissertation supervisor or with my colleagues, but I will ensure that no identifying details are included in these excerpts. The finished dissertation will be deposited into the City University dissertation repository, at which point the interview transcripts and notes, up to now stored on my computer and on Dropbox, will be destroyed.

What will happen to results of the research study?

The study will be finished in December 2016 and submitted in January 2017, and deposited into the City University dissertation repository. It is possible that the study will be adapted into a shorter form for publication in a journal; if it is, the same standards of confidentiality and anonymity discussed above will obtain. If you would like a copy of the finished study, or details of any future adaptations, you may ask me at any time.

What will happen if I don’t want to carry on with the study?
You may decide at any point during the study (up until the study ends in December 2016) that you don’t want to carry on with it. If you do, simply tell me and I will destroy the data from your interview. You will not need to provide an explanation for your withdrawal, and you will not be penalized in any way for withdrawing.

**What if there is a problem?**

If you have any problems, concerns or questions about this study, you should ask to speak to a member of the research team. If you remain unhappy and wish to complain formally, you can do this through the University complaints procedure. To complain about the study, you need to phone 020 7040 3040. You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the name of the project is: *Understanding resource discovery and use habits among users of digitized special collections.*

You could also write to the Secretary at:

Anna Ramberg  
Secretary to Senate Research Ethics Committee  
Research Office, E214  
City University London  
Northampton Square  
London  
EC1V 0HB  
Email: Anna.Ramberg.1@city.ac.uk

City University London holds insurance policies which apply to this study. If you feel you have been harmed or injured by taking part in this study you may be eligible to claim compensation. This does not affect your legal rights to seek compensation. If you are harmed due to someone’s negligence, then you may have grounds for legal action.

**Who has reviewed the study?**

This study has been approved by City University London [insert which committee here] Research Ethics Committee

**Further information and contact details**

Researcher: Emma Stanford, emma.stanford@city.ac.uk  
Supervisor: David Bawden, d.bawden@city.ac.uk

Thank you for taking the time to read this information sheet.
CONSENT FORM

Emma Stanford
City University London

Title of Study: Understanding resource discovery and use habits among users of digitized special collections

Please initial box

1. I agree to take part in the above City University London research project. I have had the project explained to me, and I have read the participant information sheet, which I may keep for my records.

I understand this will involve
- be interviewed by the researcher
- allow the interview to be videotaped/audiotaped
- complete questionnaires asking me about my resource discovery and resource use preferences
- use a computer to complete simple resource discovery tasks

2. This information will be held and processed for the following purposes: analysis of resource discovery and use trends and the generation of recommendations for future resource discovery and dissemination plans.

I understand that all personal details will be omitted from the data and that discussions of potentially identifying details will be kept as general as possible in order to protect my identity from being made public.

I understand that I will be given a transcript of data concerning me for my approval before it is included in the write-up of the research.

3. I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalized or disadvantaged in any way.
4. I agree to City University London recording and processing this information about me. I understand that this information will be used only for the purpose(s) set out in this statement and my consent is conditional on the University complying with its duties and obligations under the Data Protection Act 1998.

5. I agree to take part in the above study.

____________________ ____________________________ _____________
Name of Participant  Signature    Date

____________________ ____________________________ _____________
Name of Researcher  Signature    Date

When completed, 1 copy for participant; 1 copy for researcher file.
Ethics Review Form: LIS Masters projects

In order to ensure that proper consideration is given to ethical issues, all students undertaking the LIS dissertation project must complete this form and attach it to their dissertation proposal. Consult your supervisor if anything in this form is unclear or problematic. There are two parts:

**Part A: Ethics Checklist.** All students must complete this part. The checklist identifies whether the project requires ethical approval and, if so, where to apply for approval. Students who answer ‘yes’ to any of questions 1-18 should consult their supervisor, as they may need approval from the ethics committee.

**Part B: Ethics Proportionate Review Form.** This part is an application for ethical approval of low-risk research. Students who have answered “no” to questions 1 – 18 and “yes” to question 19 in the checklist must complete this part; students who have answered ‘no’ to all the questions 1-19 may ignore this part. The supervisor has authority to approve this application.

### Part A: Ethics Checklist

<table>
<thead>
<tr>
<th>If your answer to any of the following questions (1 – 3) is YES, you must apply to an appropriate external ethics committee for approval:</th>
<th>Delete as appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Does your project require approval from the National Research Ethics Service (NRES)? (E.g. because you are recruiting current NHS patients or staff? If you are unsure, please check at <a href="http://www.hra.nhs.uk/research-community/before-you-apply/determine-which-review-body-approvals-are-required/">http://www.hra.nhs.uk/research-community/before-you-apply/determine-which-review-body-approvals-are-required/</a></td>
<td>Yes/No</td>
</tr>
<tr>
<td>2. Will you recruit any participants who fall under the auspices of the Mental Capacity Act? (Such research needs to be approved by an external ethics committee such as NRES or the Social Care Research Ethics Committee <a href="http://www.scie.org.uk/research/ethics-committee/">http://www.scie.org.uk/research/ethics-committee/</a>)</td>
<td>Yes/No</td>
</tr>
<tr>
<td>3. Will you recruit any participants who are currently under the auspices of the Criminal Justice System, for example, but not limited to, people on remand, prisoners and those on probation? (Such research needs to be authorised by the ethics approval system of the National Offender Management Service.)</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>

### Part A: Ethics Checklist

<table>
<thead>
<tr>
<th>If your answer to any of the following questions (4 – 11) is YES, you must apply to the Senate Research Ethics Committee for approval (unless you are applying to an external ethics committee):</th>
<th>Delete as appropriate</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Does your project involve participants who are unable to give informed consent, for example, but not limited to, people who may have a degree of learning disability or mental health problem, that means they are unable to make an informed decision on their own behalf?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>5. Is there a risk that your project might lead to disclosures from participants concerning their involvement in illegal activities?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>6. Is there a risk that obscene and or illegal material may need to be accessed for your project (including online content and other material)?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>7. Does your project involve participants disclosing information about sensitive subjects?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>8. Does your project involve you travelling to another country outside of the UK,</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Question</td>
<td>Yes/No</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>where the Foreign &amp; Commonwealth Office has issued a travel warning?</td>
<td></td>
</tr>
<tr>
<td>(<a href="http://www.fco.gov.uk/en/">http://www.fco.gov.uk/en/</a>)</td>
<td></td>
</tr>
<tr>
<td>9. Does your project involve invasive or intrusive procedures? For example, these may include, but are not limited to, electrical stimulation, heat, cold or bruising.</td>
<td>Yes/No</td>
</tr>
<tr>
<td>10. Does your project involve animals?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>11. Does your project involve the administration of drugs, placebos or other substances to study participants?</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>

If your answer to any of the following questions (12 – 18) is YES, you should consult your supervisor, as you may need to apply to an ethics committee for approval. Delete as appropriate.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes/No</th>
</tr>
</thead>
<tbody>
<tr>
<td>12. Does your project involve participants who are under the age of 18?</td>
<td></td>
</tr>
<tr>
<td>13. Does your project involve adults who are vulnerable because of their social, psychological or medical circumstances (vulnerable adults)? This includes adults with cognitive and / or learning disabilities, adults with physical disabilities and older people.</td>
<td></td>
</tr>
<tr>
<td>14. Does your project involve participants who are recruited because they are staff or students of City University London? For example, students studying on a particular course or module. (If yes, approval is also required from the Project Tutor.)</td>
<td></td>
</tr>
<tr>
<td>15. Does your project involve intentional deception of participants?</td>
<td></td>
</tr>
<tr>
<td>16. Does your project involve identifiable participants taking part without their informed consent?</td>
<td></td>
</tr>
<tr>
<td>17. Does your project pose a risk to participants or other individuals greater than that in normal working life?</td>
<td></td>
</tr>
<tr>
<td>18. Does your project pose a risk to you, the researcher, greater than that in normal working life?</td>
<td></td>
</tr>
</tbody>
</table>

If your answer to the following question (19) is YES and your answer to all questions 1 – 18 is NO, you must complete part B of this form.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes/No</th>
</tr>
</thead>
<tbody>
<tr>
<td>19. Does your project involve human participants? For example, as interviewees, respondents to a questionnaire or participants in evaluation or testing.</td>
<td></td>
</tr>
</tbody>
</table>
### Part B: Ethics Proportionate Review Form

If you answered YES to question 19 and NO to all questions 1 – 18, you may use this part of the form to submit an application for a proportionate ethics review of your project. Your dissertation project supervisor will review and approve this application.

#### The following questions (20 – 24) must be answered fully.

<table>
<thead>
<tr>
<th>Question</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will you ensure that participants taking part in your project are fully informed about the purpose of the research?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Will you ensure that participants taking part in your project are fully informed about the procedures affecting them or affecting any information collected about them, including information about how the data will be used, to whom it will be disclosed, and how long it will be kept?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>When people agree to participate in your project, will it be made clear to them that they may withdraw (i.e. not participate) at any time without any penalty?</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Will consent be obtained from the participants in your project, if necessary? Consent from participants will only be necessary if you plan to gather personal data. “Personal data” means data relating to an identifiable living person, e.g. data you collect using questionnaires, observations, interviews, computer logs. The person might be identifiable if you record their name, username, student id, DNA, fingerprint, etc. If YES, attach the participant information sheet(s) and consent request form(s) that you will use. You must retain these for subsequent inspection. Failure to provide the filled consent request forms will automatically result in withdrawal of any earlier ethical approval of your project.</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Have you made arrangements to ensure that material and/or private information obtained from or about the participating individuals will remain confidential? Provide details:</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>

#### If the answer to the following question (25) is YES, you must provide details

<table>
<thead>
<tr>
<th>Question</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will the research involving participants be conducted in the participant’s home or other non-University location? If YES, provide details of how your safety will be ensured:</td>
<td>Yes/No</td>
</tr>
<tr>
<td>Attachments (these must be provided if applicable):</td>
<td>Delete as appropriate</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Participant information sheet(s)</td>
<td>Yes / No / Not applicable</td>
</tr>
<tr>
<td>Consent form(s)</td>
<td>Yes / No / Not applicable</td>
</tr>
<tr>
<td>Questionnaire(s)**</td>
<td>Yes / No / Not applicable</td>
</tr>
<tr>
<td>Topic guide(s) for interviews and focus groups**</td>
<td>Yes / No / Not applicable</td>
</tr>
<tr>
<td>Permission from external organisations (e.g. for recruitment of participants)**</td>
<td>Yes / No / Not applicable</td>
</tr>
</tbody>
</table>

**If these items are not available or not applicable at the time of submitting your project proposal, preliminary approval through proportionate review can still be given. This will be subject to you submitting the items to your supervisor for approval at a later date. Approval must be obtained prior to the research commencing.

**Templates**

The University provides templates which should be used as the basis for your participant information sheets and consent forms. These are available from the links below but must be adapted according to the needs of your project before they are submitted for consideration.

Adult information sheet:

http://www.city.ac.uk/_data/assets/word_doc/0018/153441/TEMPLATE-FOR-PARTICIPANT-INFORMATION-SHEET.doc

Adult consent form:

http://www.city.ac.uk/_data/assets/word_doc/0004/153418/TEMPLATE-FOR-CONSENT-FORM.doc


