A Research Information Network Report April 2011

Reinventing research? Information practices in the humanities







Acknowledgements

This report was the result of a collaborative effort between RIN and the study authors at the 1) Oxford Internet Institute (OII), University of Oxford, 2) UCL Centre for Digital Humanities and the Department of Information Studies, UCL, 3) e-Humanities Group, Royal Netherlands Academy of Arts & Sciences (KNAW) and Maastricht University, and 4) Oxford e-Research Centre (OeRC), University of Oxford. The main authors for this report are:

Monica Bulger, Eric T. Meyer, Grace de la Flor, Melissa Terras, Sally Wyatt, Marina Jirotka, Katherine Eccles, Christine Madsen

For the full acknowledgements, see page 82.

* TABLE OF CONTENTS

Executive summary	6
Introduction	10
Case studies	20
Resource: Old Bailey Online	22
Resource: The Digital Image Archive of Medieval Music (DIAMM)	30
Department: University of Birmingham English Department	36
Department: UCL Philosophy Department	42
Field: Corpus linguistics	48
Collaborative Project: The Digital Republic of Letters	54
Findings across cases	62
Engagement with digital resources	68

Conclusions	70
Policy changes and recommendations	74
References	78
Notes	80
Full acknowledgements	82

EXECUTIVE SUMMARY

Overview

This report is the second in a series of three commissioned by the Research Information Network (RIN), each looking at information practices in a specific discipline (life sciences, humanities and physical sciences). The aim is to understand how researchers within a range of disciplines find and use information, and in particular how that has changed with the introduction of new technologies.

Humanities scholars are often perceived in very traditional terms: spending a lot of time working on their own and collaborating only informally through highly-dispersed networks. Unlike most scientists, they have no long tradition of working in formal, close-knit and collaborative research groups. Humanities scholars have also sometimes been presented as 'depth' rather than 'breadth' researchers, preferring to spend significant amounts of time with a few items, rather than working across a broader frame. In terms of information sources, text and images held in archives and libraries tend to dominate, with less of an association with new web-based technologies (although this is changing with the increasing visibility of digital humanities).

This report suggests that such perceptions may be out of date. In each of our case studies we found researchers working with new tools and technologies, in increasingly collaborative environments, and both producing and using information resources in diverse ways. There is a richness and variety within humanities information practices which must be recognised and understood if we are to provide the right kind of support for researchers.

Method

The study used six very different cases to examine the spread of practices across the humanities. Two cases involved users of a specific online resource, two involved members of humanities departments, one looked at an entire field and the final case involved participants on an international collaborative project. In each case, data was gathered using a series of interviews with researchers at various stages of their careers, followed by focus groups to explore some of the common interview themes.

Key findings

Researchers in the humanities adopt a wide variety of approaches to their research. Their work tends to focus on texts and images, but they use and also create a wide range of information resources, in print, manuscript and digital forms. Like other researchers, they face multiple demands on their time, and so they find the ease and speed of access to digital resources very attractive: some of them note that they are reluctant on occasion to consult

texts that require a trip to a distant library or archive. Nevertheless, none of the participants in our study is yet ready to abandon print and manuscript resources in favour of digital ones. Rather, they engage with a range of resources and technologies, moving seamlessly between them. Such behaviours are likely to persist for some time.

This is reflected also in how researchers disseminate their research. The overwhelmingly dominant channels are the long-established ones such as journal articles, conferences and workshops, monographs and book chapters. We found only limited use – except among philosophers - of blogs and other social media. We noted the doubts expressed in other fields about quality assurance for users of such media, but also concerns about how best to present material that will be read by non-academic audiences.

A key change in humanities research over the past 10-15 years has been the growth of more formal and systematic collaboration between researchers. This is a response in part to new funding opportunities, but also to the possibilities opened up by new technology. Over recent years there has also been a shift from the model under which technology specialists tell researchers how to do their research to more constructive engagement. Like other researchers, scholars in the humanities use what works for them, finding technologies and resources that fit their research, and resisting any pressure to use something just because it is new.

But there is little evidence as yet of their taking full advantage of the possibilities of more advanced tools for text-mining, grid or cloud computing, or the semantic web; and only limited uptake of even simple, freely-available tools for data management and sharing. Rather, they manage and store information on their desktops and laptops, and share it with others via email.

Barriers to the adoption and take up of new technologies and services include lack of awareness and of institutional training and support, but also lack of standardisation and inconsistencies in quality and functionality across different resources. These make for delays in research, repetitive searching, and limitations on researchers' ability to draw connections and relationships between different resources.

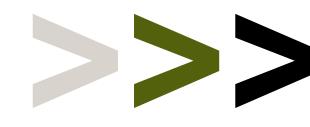
Policy Challenges

The findings of this study underscore those of the earlier one on researchers in the life sciences: if the policies and strategies of research funders, universities and service providers are to be effective in optimising the use and exchange of scholarly information, they must be sensitive to the practices and cultures of different research communities.

This report shows that researchers in the humanities are adopting new technologies, and that they are engaging with a wide range of digital resources and services, alongside printed and manuscript texts and images. There is much talk about developments in 'digital humanities', but the practices and concepts are not as yet well-established or fully-embedded as standard features of scholarly practice. Learned societies and professional associations, as well as universities, have important roles to play in helping scholars in the humanities to engage with new digital approaches, so that they become an integral part of what it means to be a humanities scholar. In order to support further progress towards that end, without diminishing the importance of traditional scholarly skills and techniques, a number of policy challenges will need to be addressed.

- 1. Researchers in the humanities need training from specialists in their domain, as well as from technology specialists, on how to use new tools and services that will facilitate their research. They also need effective ondemand support to underpin the intellectual investment that they make in learning new skills and engaging with new resources, tools and services.
- 2. Researchers are wary of digital resources that are either incomplete or highly-selective. They also worry about the long-term preservation and sustainability of resources which require regular scholarly as well as technical input if they are to retain their usefulness. All those with an interest in the

- future of humanities research universities, funders, libraries, archives, publishers need to consider how best to develop and sustain the digital infrastructure to support research and scholarship for the long term.
- 3. Building and sustaining online resources for humanities research such as Old Bailey Online and the Digital Image Archive of Medieval Music studied in this report makes heavy demands on the time and expertise of both disciplinary and technology specialists. More explicit recognition and career rewards for such work would make humanities scholars more prepared to participate in building new resources, so that they are well-designed for effective use.
- 4. For humanities scholars, the sharing of data between researchers is not such a prominent challenge as it is in the many of the sciences. A key challenge rather is to improve their ability with user-friendly tools and methods to link data housed in different archives.
- 5. Assessing the value and impact of digital resources hampered by the lack of standard practice in citing their use. Building resources and systems with digital object identifiers (DOIs), and which also include recommendations on citation practice, would help those following chains of citations, as well as those interested in assessing the value and impact of different resources.



NTRODUCTION

Background

This report is one in a series of reports by the Research Information Network (RIN) that includes case studies of information use in the life sciences (RIN & British Library, 2009), humanities (this report), and the physical sciences (forthcoming). More broadly, RIN has also issued reports on a wide variety of topics, including how academic libraries are used, how researchers discover information, how they overcome barriers to accessing research information, how researchers engage in collaborative research with industry, and how open access and open science are offering new opportunities for researchers to share and communicate their work.

The case studies on information use and exchange in the life sciences showed that researchers use a limited range of services, and were more reliant on colleagues for informal advice about appropriate services than they were on institutional service personnel such as librarians. The report also suggested that "there is a significant gap between how researchers behave and the policies and strategies of funders and service providers" (RIN & British Library, 2009, p. 5). Some groups of life scientists are aware of the need for effective data management and sharing, but compliance with funder requirements is patchy at best. Information professionals must therefore engage with the

research communities they are meant to support, and there must be a new focus on development for people to fulfill roles as information specialists to support research on a regular basis.

Scholars in humanities and life sciences share similarly high expectations of digital resources. This report examines humanities scholars' expectations of tools and their criteria for use. Scholars in both fields share surprisingly similar practices in information management, collaboration, and creation. The degree to which humanities scholars engage in these practices is uncovered in this report.

Recent studies suggest that many humanities scholars are becoming enthusiastic users of digital resources (Meyer et al., 2009), although they still tend to prefer generic informational resources such as library and archive web pages or online reference collections to digital objects which are comparable to scholarly books (British Academy, 2005). Humanities scholars need a wide range of both analogue and digital resources for their research, encompassing various ages and types of materials, including traditional materials such as printed matter and manuscripts as well as materials in newer formats such as digital images, audio, 3-dimensional scans of artefacts, and other new forms of information.

Research has only recently been conducted into the online behaviour of humanities scholars: by Warwick and Terras' CIRCAh research group (UCL) and Duff and Cherry (University of Toronto) (Duff & Cherry, 2000). While earlier work by Bates addressed similar issues, the research predates web publishing (Bates, 2002). From current work, we already know quite a lot about what humanities scholars do when online. They are aware of the potential and problems of digital resources, being concerned with accuracy, selection methods, and ease of use (Warwick et al., 2008), and require information about the original item when materials are digitised (Duff & Cherry, 2000). They expect high quality content: anything that makes a resource difficult to understand - a confusing name, a challenging interface, or data that must be downloaded - will deter them from using it (Warwick et al., 2008). Humanities scholars are therefore a particularly challenging population for whom to design, since they are easily deterred from using digital resources. Yet if the right resources, containing useful content, with a usable interface can be produced this should significantly increase uptake. To accomplish this type of development, we need to understand better user behaviours, needs, and preferences in the humanities field (Borgman, 2009).

Collaboration in the humanities

The image of the "lone scholar" has dominated the characterisation of the humanities scholar since the latter half of the twentieth century (Buchard, 1965). Empirically, however, this notion has long since been debunked. The 2008 TIDSR Survey of Humanities Scholars found that a majority of respondents had collaborated with someone from inside their department (53%) in the previous five years and that collaborations beyond their department were even more prevalent (65%) (Meyer et al., 2009). Moreover, dialogue with other scholars is a feature of the work of virtually all humanities researchers.

Willard McCarty (2005) makes a strong case that although the act of writing may be a solitary one for humanities scholars, their work environment has always been "virtually communal."

In the humanities, scholars have tended to be physically alone when at work because their primary epistemic activity is the writing, which by nature tends to be a solitary activity. (p.12)

McCarty does not confuse the solitary nature of writing, however, with the type of intellectual isolation inferred by the idea of the "lone scholar." He argues, "[i]f we look closely at this solitary work, we have no trouble seeing that the normal environment has always been and is virtually communal," (p. 12) for to

participate in the practice of scholarship within the bounds of academia, one has to participate in a conversation. McCarty asserts that to quote, cite, or reference, is to enter into a community of dialogue, however asynchronous.

However far back in time one looks, scholarly correspondence attests to the communal sense of work. So do the conventions of acknowledgement, reference and bibliography; the crucial importance of audience; the centrality of the library; the physical design of the book; the meaning of publication, literally to 'make public'; the dominant ideal of the so-called 'plain style', non sibi sed omnibus, 'not for oneself but for all'; and of course language itself, which, as Wittgenstein argued in Philosophical Investigations, cannot be private. Writing only looks like a lonely act. (McCarty, 2005, p.13)

Why and When Scholars Make Use of Digital Tools

Underpinning the present study of information practices of humanities scholars is the larger question of why humanities scholars use or adopt digital tools in their work. Studies of humanities scholars consistently find that they adopt a technology when it improves upon their current practice, for example performing functions or enabling breadth not possible without technologies (Wiberley, 1991; Wiberley & Jones, 1989; Wiberley & Jones, 1994). In

short, as Madsen (2010) found: "humanities scholars will use a technology or tool when it fills an existing need or in some way accelerates or simplifies a task essential to their work."

In 2002 Deegan and Tanner expanded this idea to look at formats as well as tools. Their study of why people favoured digital formats over analogue ones concluded that it was the immediacy and ease of access that primarily drew people to digitised works. They were particularly drawn in by tools that helped improve access in some way, by allowing zooming or changing contrast etc., to improve the visibility of the image. As Brockman found, "it takes time for any researcher to develop facility with a particular tool" (2001, p. 23), so tools must be useful if they are to be used. The converse is also true. If a tool is not useful, if it does not in some way facilitate the work of the scholar (Bates, 1996), or if it requires the acquisition of too many new skills (Brown et al., 2006), it will not be used. Novelty in and of itself is not necessarily useful. Tools must first facilitate or simplify an existing practice before creating or suggesting new ones.

What was for many years generally reported as indifference by humanities scholars to digital tools and resources was found by several researchers to be a deliberate choice to avoid that which is not useful. Talja & Maula (2003) found, for example, that humanities researchers adopt new technologies when they save them time or effort. Or, as Brockman et al. (2001)

found, adoption of new technologies is first in the form of an extension of existing research practices.

Why This Study is Important

As new types and forms of information are becoming available to humanities researchers, it is vital that we continue to develop our understanding of their needs and preferences, to inform the design of intuitive and effective digital resources. We need to understand the changes in the inputs, practice, and outputs of scholars in the humanities, as well as the characteristics that have not changed (such as the continued reliance on monographs in some disciplines) (Harley et al., 2010).

We must also recognise that as disciplines expand and sub-disciplines evolve, not all facets of the humanities can be lumped together. A medievalist may use information resources very differently from a modern historian, but both are humanities scholars. Comparisons between the humanities and with the sciences are inevitable however, and so it is important to understand both the significant differences and similarities between them rather than relying upon broad characterisations of 'two cultures.' Parallel studies to this one have been conducted for the life sciences (Harley et al., 2009), and therefore allow for identification of similar themes, such as preference for useful tools, and comparisons, such as uptake of digital resources.

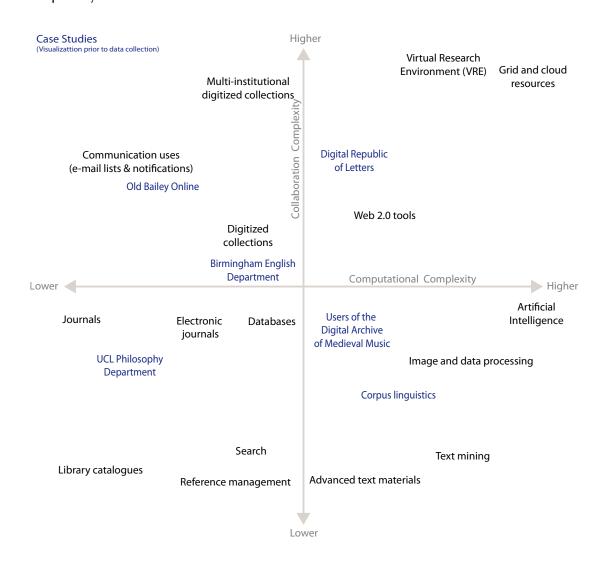
This report extends our understanding of the different types of information humanities scholars rely upon to conduct their research. We present the findings of six case studies from a broad range of humanities disciplines across a variety of UK and international academic institutions. From these studies we identify key trends in information use, and we provide recommendations that research funders and institutions may use to formulate their research policy, strategy and practice.

The complexity continuum

Developed in response to trends we previously observed in humanities scholars' use of digital resources (Meyer et al., 2009), Meyer developed the concept of a complexity continuum for this study. This model allows us to examine how humanities researchers locate, evaluate, organise, manage, transform, create and communicate their research data and information sources as part of the research process, as well as the barriers they face. The ways in which scholars mix digital and analogue

resources, and how it may be transforming work practices, are analysed through the lens of this complexity continuum, as shown in Figure 1, which will be revisited later in this report. The version here shows the general elements of the complexity continuum, and the general placement of the cases we selected as we envisioned them *prior to data collection*. As we will see later in this report, however, the range of actual behaviours is quite broadly scattered across the diagram, even within each case study.

Figure 1. Complexity continuum



Approach and methodology

The complexity continuum lets us visualise a range of information seeking, access, use, and dissemination behaviours through this two-dimensional matrix.

The *horizontal axis* indicates a range of lower to higher complexity in terms of the computational resources required to create, maintain, access, and use research information resources. This axis of computational complexity is labelled from lower to higher rather than from low to high because the perception of computational complexity changes over time. Examples include the use of electronic journals, image processing and virtual research environments.

The *vertical axis* indicates the collaboration complexity of various types of information behaviours. The lower portion of the complexity continuum represents those information behaviours that take place in isolation, or by way of relatively small or simple social arrangements. The upper portion includes those information behaviours that are linked to more complex social coordination and collaboration. Examples include collaboration in multi-institutional digitised collections, the use of email lists, and face-to-face collaboration.

The scale we have used does not imply any value judgment: information resources with lower computational complexity may be just as important and valuable, even more so, than their more complex cousins. However, from a sociotechnical perspective, the relative complexity of one set of resources in comparison to another has important implications in terms of types of users, uses, and the likelihood of uptake.

We aimed to cover a broad range of humanities research practices and therefore sought cases that would represent traditional resource use, advanced digital resource use, and cases in between. The selected cases represent different aspects of scholarship within the humanities and include resource use with a specific database, use within a traditional department, use within a field, and use within a collaborative project:

- Resource: *Old Bailey Online* [http://www.oldbaileyonline.org]
- Resource: *Digital Image Archive of Medieval Music* [http://www.diamm.ac.uk]
- Department: University of Birmingham English department [http://www.english. bham.ac.uk]
- Department: UCL Philosophy department [http://www.ucl.ac.uk/philosophy]
- Field: Corpus Linguistics
- Collaborative Project: The Digital Republic of Letters

Participants

Similar to recent exploratory studies of scholarship practices among faculty (Harley et al., 2009; RIN & British Library, 2009; RIN & NESTA, 2009), we relied upon a combination of convenience and snowball sampling. The convenience aspect of our sampling involved contacting colleagues recommended by a known contact at the beginning of our study. Snowball sampling was used to identify other potential participants within a hard-to-reach group. Typically, in snowball sampling, one contact is asked to suggest additional contacts, who are also asked to recommend contacts. In particular, respondents were asked to identify other researchers with higher and lower levels of familiarity and skill with computational resources as a way of broadening the sample. While these methods may introduce bias because they are not random, they allowed us to qualitatively explore use practices within relatively small academic communities.

We conducted 39 semi-structured interviews and four focus group discussions (with 19 participants, 4 of whom also participated in interviews) for a total of 54 participants. In addition to interviewing professors, lecturers, and researchers, we identified database developers (3), project managers (3), and academic support staff (2) actively involved in humanities research. Some scholars acted in dual roles as faculty members and database developers or programme managers. To provide a broad perspective of scholarly resource use in the humanities, we also included graduate students (10) and postdoctoral researchers (2). In total, scholars from 25 institutions in five countries participated in our study. Table 1 provides a description of participants within each case and case type.

Table 1. Participants by Case and Case Type

Case type	Cases	Interview Participants	Focus Group Participants
Resource	Old Bailey Online	8	4
	DIAMM	5	4
Department	Department University of Birmingham English Department		0
	UCL Philosophy Department	6	4
Field	Corpus Linguistics	6	0
Collaborative The Digital Republic of Letters Project		9	7*
		39	19

^{*}Note: In total, 7 scholars participated in the *The Digital Republic of Letters* focus group. Four of the participants also participated in the interview, so were counted only once toward the total participant number of 54.

Process

Scholars were invited via e-mail to participate in the study. Whenever possible, we conducted in-depth interviews face-to-face, but when distance or timing precluded this option we used Skype, often with video conferencing enabled. Interviews usually lasted one hour, but we allowed additional time for elaboration and discussion. Following the interview, we asked participants to complete an online survey.

Once interviews were complete, we conducted focus group discussions. In three of the cases, these discussions involved participants who had not been interviewed; the fourth case involved a blend of interview participants and additional discussants. Focus groups allowed us to explore more deeply themes emerging from the interviews. They provided an opportunity to speak to graduate students after interviewing faculty, or practitioners after interviewing developers.

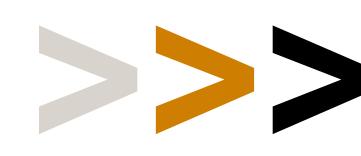
Near the end of our study, we invited scholars to participate in a mini-pilot in which we tested a web history collection methodology. Web histories provide insight into actual resource use and complement self-reports collected through

survey or interview (Bulger, 2009; Bulger et al., 2008; CIBER, 2008; Dirksen, 2007). In particular, they provide information trails that allow for deep investigation of search practices and engagement with specific resources (Bulger, 2009; Cothey, 2002). For those who participated, we provided instructions for setting up a remote proxy server. Using their Internet browser, participants could log in to a server hosted by the Oxford Internet Institute, which recorded and time-stamped their actions and saved the logs to a password-protected file. Participants logged in to the proxy server to collect their web histories over a 1-2 day period. To supplement the recording of their online actions, we asked these participants to maintain a diary of their offline research activities. Three scholars (2 from the Old Bailey Online case study and 1 from the corpus linguistics case study) participated in this exploratory portion of the study. Although this is a small number of participants, the data suggest that this method will be useful for future research, including in the forthcoming case studies focusing on the physical sciences.

Analysis

Interviews were transcribed and then reviewed by another researcher for accuracy. From each interview, we identified emerging themes within the interview transcripts and wrote summaries of each interview organised around these themes. The interview summaries informed summaries of each case, which then informed the larger report. Survey responses were analysed quantitatively using SPSS 13.0. Responses were coded and tallied. We performed frequency analysis of resource use and communication practices. Additionally, we conducted crosstabulations to explore relationships among groups and their use. Using Webometric analysis techniques developed by Thelwall (2009), we extracted information trails from the web history logs. In particular, we focused on searching and browsing practices. We counted how many unique URLs were visited, how many times a search term was modified, and the duration of each session. We additionally examined the types of resources used; for example, journal databases versus digital archives, or online reference websites. We supplemented our log file analysis with the daily diaries to explain gaps in use and to consider the possibility of multi-tasking and simultaneous hybrid use of print and digital resources.

The multiple methods employed involved both automatic and manual data collection of information behaviours through personal interviews, focus groups, and webometrics. By triangulating these different data sources, we have secured a reasonable understanding of the information practices of the humanities scholars who participated in the study. Our findings should not, however, be considered as representative of all humanities scholars, and wider generalisations would be premature. Nevertheless, we are confident that the report provides relevant insights into transformations in research practice, and their implications for researchers, institutions, and funders.



Case studies

V

How do humanities scholars, working in their offices, libraries, archives, or homes, access and use information to support their research? As with any complex set of behaviours, the answer to this question varies widely from person to person, discipline to discipline, and institution to institution. Each researcher is at the centre of a complex socio-technical interaction network¹ made up of scholars engaging with each other, with the institutions of which they are a part, and with the technological systems, both analogue and digital, which enable them to access and use information. These technologies range from the centuries-old technology of writing and the book, to the latest high-powered computers that allow researchers to analyse huge bodies of text and images.

The following case studies are meant to sample a range of these information behaviours by focusing on specific types of humanities scholars. Each case was selected to bring into focus different types of strategies that humanities scholars draw upon for using information in their research. These case studies are illustrative, not exhaustive. One could never hope to document the entire range of diverse practices across the humanities. In particular, our case selection and recruitment tends to favour scholars likely to be engaging with digital materials at some level. These case studies suggest, however, that the traditional stereotype of the dust-covered humanities scholar labouring alone in the dark archive of manuscripts may be becoming the exception, rather than the rule.

The first two cases include participants who use information resources that were selected for their widespread use (*Old Bailey Online*) and to show the use of different modes of data such as images (*DIAMM*). The second two cases comprise scholars recruited from traditional humanities departments (Birmingham English & UCL Philosophy). The fifth case comprises of members of a computationally-intense research field (Corpus Linguistics), while the final case focuses on members of a collaborative international humanities project (*The Digital Republic of Letters*). Since two of the case studies focus on a resource, another on computationally-intensive field, and a fourth on an a international collaboration, our sampling methods may have preferenced tech-savvy users (or at the very least, those not afraid of using technology). Nonetheless, these cases illustrate the broad range of information use strategies practised by humanities scholars.

OLD BAILEY ONLINE

THE

PROCEEDINGS

AT THE

Seffions of the Peace, and Oyer and Terminer,

FOR THE

City of LONDON,

AND

County of MIDDLESEX;

O N

Wednesday the 3d, Thursday the 4th, Friday the 5th, and Saturday the 6th of December, 1729, in the Third Year of His MAJESTY's Reign.

Being the First SESSIONS in the Mayoralty of the Right Honourable Sir RICHARD BROCAS, Knt. Lord Mayor of the City of LONDON, in the Year 1730

No. I. for the faid YEAR.

LONDON:

Printed for T. PAYNE, at the Crown in Ivy-Lane, near Pater-noster-Row. M.DCC.xxx.

(Price Six Pence.)

The Old Bailey Online is a database containing criminal proceedings of London's central criminal court from 1674-1913. With records of 197,745 criminal trials, the Old Bailey Online represents the largest "fully searchable edition...detailing the lives of non-elite people"2 in London from the 17th to the early 20th centuries. Of interest to historians and in particular legal historians, this resource is often used to identify trends in crime, to trace specific individuals, and to provide historical details of everyday life. In a recent survey (Meyer et al., 2009), 62% of humanities scholars from the United Kingdom reported having seen or used Old Bailey Online, ranking it the fourth highest of 15 resources.

The database allows for keyword and advanced searches. Results include keyed in summaries as well as digital images of the original documents. As a result of space limitations at the time of their original reporting, the *Old Bailey Online* proceedings are not representative of all cases, nor do they comprehensively report on each trial. Nonetheless, John Langbein,

a legal historian at Yale Law School, confirms that the proceedings published in the *Old Bailey Online* are "probably the best accounts we shall ever have of what transpired in ordinary English criminal courts before the later eighteenth century."³

Development of this online resource was funded by the Arts and Humanities Research Council (2000, 2005), the Big Lottery Fund (2001), and the Economic and Social Research Council (2005), with technical and administrative support from the University of Hertfordshire, Open University, and University of Sheffield.⁴

To examine the information behaviours of users of this resource, we interviewed eight scholars, including one of the developers of the *Old Bailey Online*. Participants include professors of history, British history, history of crime, humanities computing, and human communications. Additionally, four graduate students in the areas of 18th century British history participated in a focus group discussion.

How Old Bailey Online is used

Use of the *Old Bailey Online* is purpose-driven, with most scholars in our study reporting that they use it intermittently, depending upon project needs and research questions. Participants describe their research practice as something that occurs in bursts, usually during breaks, when they are not teaching or engaged in administrative duties. Historians report using the Old Bailey Online as a starting point that usually pre-empts Google in their initial searches. To piece together stories of single individuals, types of crimes, or everyday life in London, scholars report that a simple keyword search of the Old Bailey Online proceedings yields useful data. The social scientists we interviewed use the Old Bailey Online for specific projects that need a historical context and primarily use the keyword search function. In fact, keyword search represents the most common use of the Old Bailey Online, with most reporting that, given the volume of texts they study, browsing them would be overwhelming, timeconsuming, and likely not as productive.

Historians in our study use other collections in combination with the *Old Bailey Online*, rather than considering it a stand alone resource. *London Lives*,⁵ the second generation of *Old Bailey Online*, was used most frequently because of its links to a wealth of resources, including parish, hospital, coroner, and guild records contained across 15 datasets. The historians also report using *Eighteenth Century Collections Online* (ECCO)⁶ and the *Burney Collection of Newspapers* (British Library).⁷ Other resources frequently used

in conjunction with the *Old Bailey Online* include the *British Library Catalogue*, ⁸ *The Times Online*, ⁹ Lewis Walpole Library collection of print images, ¹⁰ National Archives websites, ¹¹ and the Wellcome Library. ¹² The scholars we interviewed also reported frequent use of traditional scholarly resources, such as online journal databases accessed through their university libraries. A majority use the search functions of their library's relevant journal article databases, such as *MetaLib*, ¹³ with few reporting use of Google Scholar ¹⁴ or Google Books. ¹⁵

Organising and annotating data

When *Old Bailey Online* users find useful information, they tend to not print. Rather, most maintain some sort of mini-database on their personal computers. For some, this database is simply a Microsoft Word file or collection of Word files. Others report maintaining lists of relevant trials using Microsoft Excel, Zotero¹⁶ (an online resource manager), or keeping a folder of each trial. One respondent received special permission from the developers of the *Old Bailey Online* to download the entire database and re-programme parts of it to meet his research needs.

A majority of the scholars we interviewed report feeling that their organisational strategies are haphazard and project-based. Many describe their organisational strategy as dependent upon the specific needs of the project, explaining that they sometimes use the *Old Bailey Online* simply to confirm dates, while during times of heavy use, they tend to employ more structured organisation.

Nearly all respondents annotate the texts or take notes, but specific practices differed.

To annotate the texts, most copy and paste relevant passages into Microsoft Word and then use the comment and highlight function. A few describe handwriting notes while reading the passages on their computer screens. Others report toggling between Microsoft Word and the files they are reading, taking notes while reading the texts online.

Data analysis

The scholars we interviewed describe their analysis as a process of deeply engaging with the primary texts, with a few additionally using textual analysis tools available through TAPoR¹⁷ or Voyeur.¹⁸ The analysis portion of their research occurs after the majority of their information collection, with use of the Old Bailey Online and supplementary resources limited to fact-checking or confirming points. During this phase, most describe a thinking space in which they alternate between stepping away from the texts and more deeply engaging with them. Some describe this process in terms of their organisational structure: at this point, they may copy and paste from their Microsoft Word file, re-organise index cards or notes, and handwrite portions of their draft or draw tables on several sheets of paper spread across their desk. One respondent describes her process: "It's easier to kind of erase columns

and move stuff, rough stuff out, and stick it in the next category, it's just, for me, it's faster with paper. And also, of course, you can spread it out and have a better look at it. You know, otherwise, you're stuck with the size of the screen." (Cognitive Psychologist in Sign Language and Deaf Studies).

Nearly every respondent described this process as finding the story within the information. Since many were using keyword search to understand trends in crime by date or by person, during this phase they piece together this information to form a coherent picture of a person's life or happenings within a specific time period.

Citation practices

For those we interviewed, citation practices were dependent upon the nature of their use of the resource. If they used the *Old Bailey Online* simply to confirm a date, none reported citing it. If it were used to locate a print resource, the print version was usually cited, a practice consistent with previous work (Meyer et al., 2009). But most felt strongly about citing digital resources in their work, especially the *Old Bailey Online*. They described *Old Bailey Online*'s citation instructions¹⁹ as helpful, something they followed, and something they passed along to their students, when applicable.

Dissemination practices

All disseminate their research primarily through traditional means such as peerreviewed journals, books, book chapters, and conference presentations. A majority expressed interest in publishing online, or perhaps plans to publish online, describing in detail studies that they felt would be appropriate for online audiences. One scholar described the need to make his research relevant to current events if he were going to publish online,20 with another describing an area of her research she found particularly interesting and therefore appropriate for an online audience. Four respondents actively maintain blogs and publish online. In reflecting upon his dissemination choices, one scholar said he now avoids book chapters because they often are not published online, so therefore are not searchable and seldom used.

Collaboration

11 of the 12 scholars interviewed report engaging in collaboration, mostly in the form of co-authoring of texts with colleagues who live within close proximity, or whom they met at a conference. Remote collaboration is usually accomplished through e-mail and phone conversations, with drafts and supplementary documents shared via e-mail and, in rare cases, using online dropboxes. Many engage in collaboration with colleagues either at their own university or within close proximity, so they arrange face-to-face meetings. All describe their collaborative work as a positive experience.

Transformations in practice

When asked about changes in their research practice, most felt that differences were due as much to increased experience as to innovations in technology, yet when describing how these changes came about, most focused on the technology. Differences in practice were mainly described in terms of ease and speed of access to information, but also ability to conduct research that was impossible before keyword search.

The older scholars we interviewed describe their earlier information collection as involving card catalogues, and later microfilm or microfiche. Many recall visiting archives and having limited time to use the primary texts, often having to handwrite important passages. Using archives meant costly travel and scheduling time to spend in the archives. Some say that cost, travel, and time presented significant barriers to their earlier research. Yet a few describe the work in the archives or at the library as more sociable than the modern alternative of working from home on their computers. Significantly, in describing shifts in their use, scholars seem to view these tools in terms of a continuum of evolving information practices, rather than abrupt, disruptive change.

Nevertheless, ease and speed of access has significantly affected their research practice, with many saying their work have been possible before, but would have involved significantly more time: I was discussing that with a colleague and what might take you several months if not years of research, you could do in hours, days, a week. So I think that means that it makes the nature of your research different because it allows you quantitative information much more quickly, which then allows you to maybe think about how you might use that information differently, because you've got so much more time.

- Historian of 18th century crime in London

When I started, you had to write things out, copy them by hand, it wasn't of course searchable or indexed in any way. And so the amount of time I now spend doing that kind of very mechanical, laborious, time-consuming work is much smaller. You can do things now in five seconds which it took you three months to do a few years ago.

Historian of 18th century crime in London

Well, now instead of just taking everything that one archive has and trying to make a story out of that, I can search through all of the archives, or through an awful lot of them, and fairly rapidly figure out what sort of documentation is out there and how I might use it. And so not only do I have a much greater variety of things at my disposal, but I know how to find them, and that makes a huge difference.

- Historian of 18th century British politics

With something like the Burney Collection... five years ago for writing an article I would need to review the newspapers. I would have gone into the British Library and done it on microfilm. Twenty years ago, I would have gone into the British Library and done it with the actual paper in front of me. Now I sit at home and I do a keyword search.

– Historian of 18th century Britain

New questions, new answers?

For most, the change of access represented by the Internet significantly changed the information collection aspect of their research. Feelings were mixed about whether this increased access to information resources enables them to ask and answer new questions. Many described their work as similar to what they had done before, only faster, while others believed the work they currently pursued would not have been possible before the Internet:

It speeds it up for the historian, but... it actually then also makes you think about how you use that information and the sorts of research questions you might ask, because the technology allows you to do different things.

- Historian of 18th century crime in London

Old Bailey Online hasn't replaced anything or displaced anything for me, but it is part of this general transformation of how I do what I do. On the one hand, it is tremendously liberating. I don't have to go read old books in onion skin anymore. On the other hand, it also puts a much higher premium on creative use of the resources. There's no reason not to be creative. There's no excuse not to get it just right. And so it's like how word processing changed the way we write. When you don't have to physically retype each page, you make a lot more changes, and so there's no excuse not to have better writing.

– Historian of 18th century British politics

Tim Hitchcock, one of the developers of the *Old* Bailey Online, expressed concern that users of digital resources may limit themselves to uses and discoveries pre-determined by the tools: "Keyword means that you discover all kinds of weird, different things, happenstantial things, when our traditional way of doing scholarship is to sit down and read a book." When scholars describe the work they pursue and work they envision using the Old Bailey Online, they often do so in terms of keyword search, thus supporting Hitchcock's concern. Most described research based on keyword search, rather than the more traditional practice of browsing primary texts, either printed in books or manuscripts. Borgman (1996) describes keyword use as a process of transforming a concept into query, which may or may not be beneficial to scholarship. For Old Bailey Online users, the keyword search seems to have a significant impact on the types of research they conduct and conceive. One 18th century social historian expressed concern about this approach, acknowledging the benefits of searching the cases in a more systematic way, but emphasising the need to learn the tools well enough to be flexible in using them.

Hitchcock further states: "My greatest frustration is that we can now answer all the questions we had in 1980 faster, much, much faster. And we can get around to publishing them much, much more quickly. But, what we haven't yet done is develop the new questions and the new paradigms that should be possible, and that we as imaginative scholars should be able to imagine."

Future technologies

Geoffrey Rockwell, lead developer of TAPoR, expressed a similar vision for advanced use of digital resources:

Most of the text tools that we have out there assume you know what question you're asking. They sit there and say 'we know it's a needle you want and we'll go find it in the haystack.' But what if you want to know what the shape of the haystack is? You don't have a needle. You're wondering what needles are here that I'd want to stick in my foot in the first place. That's where multi-variant analysis and text analysis really sing. It's there to sketch out the scope and variety that you have and allow you to explore something and figure out then what you want to pursue. For that to work, for humanists to know what to do with that means the deeper thing of changing the types of questions they ask and the way that they pursue their research.

At heart for the developers of digital resources and their advanced users is a concern that while access is easier and faster, the research questions are not very different from those asked in print. More so, their concerns lie in the unfulfilled potential of these resources. While adoption of these resources, in particular the Old Bailey Online, follows patterns of technical adoption in terms of first using a new tool within the familiar paradigm of the old tool (Straub, 2009), advanced users envision a more flexible, connected approach. They see the largest challenge as a non-technical one: changing the way the field approaches research. They believe that realising the potential of these resources and tools involves re-thinking the ways in which scholars connect concepts, formulate questions, and seek answers.

When asked what they wished they could do that they currently cannot with regard to information, most scholars wanted archives to be seamlessly connected, rather than disparate collections. Within their workflow, scholars used multiple resources, which meant inputting and modifying the same search over several databases. Thus, disconnected resources proved to be the largest barrier to accessing information. Further, not having subscriptions to specific collections, archives, or journal databases presented an often insurmountable barrier to scholars at smaller universities.

Where possible, connecting these databases together, as modelled by *London Lives*, would dramatically simplify the research process. In addition to simplifying their process, many scholars think that linking textual databases with image databases would enable them to ask increasingly complex questions about the relationships between individuals, objects, and events.



The *Digital Image Archive of Medieval Music (DIAMM)*²¹ was designed to conserve medieval polyphonic music manuscripts dating from approximately 800 to 1550. A primary impetus toward digitisation was to secure access to rare texts and preserve the quality of materials that could be lost due to natural degradation, theft, or geopolitical difficulties. Increasingly, however, musicologists and other scholars use the *DIAMM* archive as a research tool. As the web has taken on increasing importance for academic research, the *DIAMM* project has added advanced online tools to support digital image and musical notation analysis.

DIAMM users were selected as the focus of this case study primarily because they work with image-based documents. Though these documents are easily scanned into computers using modern scanning technology, they still require considerable human intervention when it comes to reading and interpretation. Unlike bodies of text which can be recognised by computers using OCR, ²² interpreting images such as music manuscripts is still beyond the powers of computer algorithms to perform at the levels of reliability required for this type of research. To understand the scholars who use this resource, we interviewed four musicologists across three continents who regularly use DIAMM, as well as DIAMM's project manager, who is a musicologist and digital imaging expert. We also held a focus group discussion for four users at a DIAMM workshop.

31

How DIAMM is used

The *DIAMM* project works with a variety of manuscript collection-holders across Europe to gain access to manuscripts and digitally photograph them. *DIAMM* adds value at this stage both as a central repository, and because *DIAMM* staff have skills in digital photography and image restoration that local collections often do not otherwise have available. The project also provides storage space for digital images which are accessible through the web to users who have created a free account with the service. These activities of resource creation and resource delivery are at the core of the DIAMM project.

Musicologists conduct much of their initial

Data collection

work in physical archives where they inspect physical manuscripts and fragments. Unsurprisingly, the musciologists we interviewed report an increasing need to use the materials away from the physical archives. They report that, when allowed, they photograph the images. Otherwise, DIAMM often contracts with libraries or museums to photograph images in the archives. The musciologists we interviewed also purchase facsimile reproductions of particularly noteworthy manuscripts such as the Eton Choirbook,23 which DIAMM has digitised and reproduced for publication. They also scan facsimiles from journals and less frequently, using microfiche.

Organising and annotating data

The musicologists in our study report relatively consistent data management practices. After accessing an image from *DIAMM*, they create notes about the documents in standard software such as Microsoft Word. They also report annotating the images, either on their personal copies or using tools such as the online annotation tool available on *DIAMM*. Musicologists in our study frequently bookmark specific images of interest within *DIAMM* so that they can easily find the image again.

Data analysis

Beyond these basic notes and annotations, different research interests in musicology require different types of data analysis. For instance, when studying the materiality of manuscripts, musicologists report using the digital images in DIAMM to examine the physical properties of manuscripts (codicology) such as the parchment it was written on, or the binding and ink pigments used. Materiality of manuscripts can also include manuscript layouts (*mise-en-page*) such as the use of typography, shape, spacing and illustrations. Musicologists pursuing palaeographical research follow the career of scribes as they move from place to place, and thus study handwriting to determine which scribes wrote what manuscripts. In interviews, they described the process as analysing the shapes of letters in detail to determine attribution. Before the availability of digital images, they would trace letters by hand. They report that instances of a letter (e.g. 'S') may now be clipped and compared using graphics-editing software.

High-resolution images enable post-processing techniques to recover lost data. For example, one respondent identifies instances of erasures or repair to damaged areas of a manuscript: "Being able to detect these changes, the scribe's erased something and rewritten something over the top of it...that has changed my scholarship in the sense that much of my work is concerned with identifying these erasures and identifying why these erasures occurred from the context of musical culture." (Musicologist specialising in late medieval music culture).

Other scholars are interested in piecing together disparate fragments of individual manuscripts using the DIAMM database. DIAMM's database houses digitized manuscripts from many European countries and enables both quick manuscript fragment comparison and letter comparison. Those using this comparative function report that it allows for an increased understanding of the relationships between different source materials. Similarly, other scholars report that DIAMM's breadth of material allows them to explore the historical and cultural contexts of manuscripts and their association with artefacts from other archives, such as receipts of sale, payment slips, etc. The *DIAMM* users in our study report that mapping these relationships allows for deeper understanding of the context in which a manuscript might have been written, as well as uncovering possible meanings of elements and musical phrases within it.

Citation practices

Musicologists report always citing *DIAMM* when using its images. A few of the musicologists we interviewed reported using

a prefatory footnote stating that some of the images in their paper were provided by DIAMM. Although citing a digital database, they still use the traditional process of referring to manuscripts in terms of where they came from and the archive that holds the shelf number and the folio. Others report citing the archive in which the physical object is held using 'with permission from' underneath the image. A few report including images can be accessed on DIAMM in their citation. But they cannot provide a specific URL, since DIAMM does not provide static addresses for each image. This lack of a static address seemed more problematic for novice users of DIAMM, since experienced users knew where to retrieve the images.

Dissemination practices

All the respondents in this case disseminate their research primarily through traditional means such as peer-reviewed journals, monographs, chapters in edited books, and conference presentations. Publishing in onlineonly journals is not seen as advantageous mainly because of pressures to publish research in highly ranked journals. However, some of the respondents also engage in less traditional modes of dissemination, such as posting versions of their papers on their personal websites or blogging about their research. Additionally, one researcher reports using Twitter²⁴ and another uses a Facebook²⁵ group to discuss and disseminate research. Most of the musicologists report that they contribute directly to the DIAMM archive through submitting their personal transcriptions of musical manuscripts.

Collaboration

All participants report that they engage in some form of collaboration, whether formally or informally. Several acknowledged that a mixture of expertise improved research where diverse skills were needed, such as translation of vocal music, codicology, or musical notation. They primarily work with personal contacts, and usually engage in face-to-face meetings with their collaborators. A majority use e-mail to share Microsoft Word files, which they mark or edit using the track changes feature. Other less common practices for document sharing include exchanging USB sticks, or using websites such as Google Docs²⁶ or YouSendIt.²⁷ Out of 9 respondents, only one reported using Skype²⁸ to collaborate remotely.

Transformations in practice

Scholars report feeling pressured to conduct research in a traditional manner: "There's this notion in the field that you can always get more out of seeing the original than seeing a digital image of it, and I do feel pressure to work more with originals than with the digital images because of the traditions of the field." The comparison between current and traditional practice encapsulates pressures reported by a majority of participants.

Transformations in practice varied depending upon when scholars began their careers. One scholar began her research before the availability of personal computing, and for her, word processing software and digital images replaced handwriting and photocopying. For all musicologists in our study, high-resolution digital images are increasingly replacing microfilm, photostats and faded photocopies of poor quality.

Respondents report that music production software has made it easier to publish new interpretations which were previously handwritten into journal articles, and that the Internet enables the piecing together of information scattered across different sources. As a musicologist specialising in philology states: "Nowadays, you can probably do 80-90 percent of everything through the highres digital images and only for the remaining 10, 15, 20 percent do you need to see the manuscript. So you can go much further (in your research) working from your own location." The ability to inspect in detail digital images of manuscripts has also allowed researchers to work in new ways. One scholar specialising in 14th and 15th century music manuscripts describes discoveries enabled by letter comparison across different manuscripts: "These are fragments that have been taken out of book bindings. So once you know that there are bindings in that library where this manuscript was broken up and used to pad the bindings, you want to go and look at other bindings from the same place....because music went out of fashion, and it was treated as scrap parchment." Overall, musicologists report that the level of detail in digitised resources provided by DIAMM allows them to examine elements within images (e.g. erasures, palaeography, materiality).

New questions, new answers?

The musicologists we interviewed say that digital images allow more highly detailed inspection of otherwise fragile physical documents. Coupled with the *DIAMM* database linking together previously disparate manuscript fragments, these advances enable existing questions to be answered more effectively. For identifying scribes by their handwriting, the pairing of high-resolution digital images allows more detailed inspection of curvatures of lettering, while the database enables cross-comparison among a collection of manuscripts. As a specialist in 14th century music describes:

The scribes of the Middle Ages worked really hard to be anonymous. If somebody started taking over the work of another scribe on page 50, the scribe would copy the handwriting of the other person. So you end up looking for very, very small, subtle, tell-tale sort of habits that you can only see with really great quality photographs, and then I'm able to see how various manuscripts connect with each other.

These advances also lead to serendipitous discoveries, as reported by a musicologist specialising in late medieval music culture:

So, for example we discovered, for example, that on the very first page of this *Chantilly Codex*, ²⁹ this very famous piece that was written in the shape of a heart, the scribe actually made a substantial erasure at one point that dramatically changed the meaning of the musical notation, and that allowed me to form a new reading of the activity of the editor's scribe at Chantilly. So the access to *DIAMM*'s images was quite instrumental in forming a very new interpretation of the participation of an individual in the history or the reception of this manuscript.

Future technologies

In the future, new advances will increase the ability to compare manuscripts in ever-greater detail. For example, new technologies such as Optical Music Recognition (OMR)³⁰ enable researchers to search through individual musical notes across many manuscripts. One researcher describes its benefits: "I can find all the pieces that are in triple time, that begin on C, that use red notes in the middle." These features allow for advance searching and comparison across manuscripts. Additionally, many expressed a desire for increased technical skills training, or support in developing them. The musicologists we interviewed echoed the view of Old Bailey Online users, that linked digital archives would enable more effective information collection and analysis. One musicologist described linked digital archives as having the potential to be "a metadatabase directed towards late-medieval song repertoire" that would provide a "confluence of information" such as bibliographic information, digital images, musical notation and scholarly translations.



University of Birmingham English Department

The English Department at the University of Birmingham has 45 members of academic staff and was graded 5 in the 2001 Research Assessment Exercise (RAE)³¹ and recorded an average 2.80 in the 2008 RAE. The department combines the study of language and literature and has a Centre for Advanced Research in English, a Shakespeare Institute, and a number of externally-funded research projects. The department is strong in Corpus Linguistics, which is addressed in our fifth case study.

All academic staff members within the department were invited via email to participate in our case study. Five faculty members, four lecturers and one senior lecturer, representing a range of research areas within the department participated. Participants' research areas include the Late Nineteenth-Century Periodical Press (1), Contemporary Literature & Postcolonial Studies (1), Early Modern Literature (1500-1800) (1), Applied Linguistics (1), and Corpus Linguistics & Discourse Analysis (2).³²

Data collection

Scholars interviewed for our study generally described beginning their data collection process with a web search, checking bibliographies, browsing their university library, and accessing hardcopy archival material, which could involve a short research trip to London, Cambridge, or another institution with large archival collections.

Most of the respondents reported using COPAC,33 an open-access catalogue that merges many university and national library databases in the UK and Ireland, and an array of individual library catalogues such as the Bodleian,34 Cambridge,35 and the British Library. The corpus linguistics staff we interviewed build a bespoke corpus, such as conversational data or online newspaper articles, in addition to making comparisons with the British National Corpus.³⁶ Other faculty members report referring to colleagues' websites to access their latest newspaper or magazine articles, short stories, or other works. They frequently use the *Literature Online* database³⁷ to conduct initial searches and then access the books through borrowing or purchase. All report using JSTOR38 and ProQuest.39

Those who use historic data report using digital images of manuscripts or newspapers. They may take photos at the library themselves, with permission, or access images through digital archives such as the 19th century British Library Newspapers Collection⁴⁰ and *Early English Books Online*. Access to the text-based transcriptions of images, in addition to the digital images, allows them to search large amounts of textual information.

Organising and annotating data

Those who work with digital images often use simple graphics applications such as Microsoft Photo Viewer.⁴¹ They may also transcribe the text within images into plain text files. For the two scholars who use corpusbased methods, data such as interviews or newspaper articles may be marked up using 'parts of speech tagging,' which identifies portions of sentences. They process these tags using text-mining and analysis tools such as Wordsmith and customised software to identify word frequency, collocates, and concordances within the data.

Data analysis

Those who use written texts spend much of their time reading and interpreting meaning. They report exploring a range of modes within texts. For example, those using corpus-based methods examine concordance lines by creating a Microsoft Excel spreadsheet to check and reorder data. Further, they compare their data with either the British National Corpus or the Bank of English42 to examine what may be novel language use within a particular social group or written genre. For those investigating images, such as in newspapers, the presentation of information is interpreted through layout. Others seek to determine the origins of individual manuscripts and may look for watermarks to determine the publisher or indications of a particular scribe.



Citation practices

When using corpora such as the *British National Corpus* or the *Bank of English* the staff we interviewed usually mention the online resource in their introduction, indicating use as a reference corpus. All staff report citing printed works such as books or newspapers, or data such as interview transcriptions, but not all cite the website where they accessed the resource.

Participants reported that the citation of nonstandard resources can be problematic. As one interviewee noted, people are now beginning to cite Twitter comments. The confusion when citing more traditional materials seems to arise when these exist in both online and hardcopy formats. The general consensus appears to be to cite the source material and not the databases used to collect those materials.

Dissemination practices

Participants report disseminating their research primarily through traditional means such as peer-reviewed journals, monographs, book chapters, and conference presentations. They report publishing fewer chapters in books because some colleagues consider them less prestigious than articles in top journals or single-author books.

Two participants published in online journals and one is a co-editor for such a journal. Two maintain a blog, while others wish to publish in open-access journals and online, digital publications.

Those who develop bespoke software for corpus linguistics disseminate it through personal websites. Other faculty make their transcriptions of manuscripts available on scholarly websites.

Collaboration

Our participants all collaborate in research at least some of the time, but frequency varies from often to rarely. They collaborate primarily through face-to-face and email communication, although for one project a scholar used a wiki to circulate drafts where students involved in the project could comment online. They use Microsoft Word's track changes feature as their primary collaboration tool, but also use the telephone, Skype, and Google Docs for co-editing journals and books. Identifying collaborators occurs through traditional means. One participant describes this as "knowing who's in your field, and what project you're working on, getting together with people at conferences" and through personal networks. In this way, scholars with different areas of expertise may come together to tackle new research questions.

Transformations in practice

Transformations in practice varied depending upon when those we interviewed began their careers. Some could have never imagined the Internet early in their careers, while others started their careers in the age of the World Wide Web. Each of them reported that they currently spend less time in the library stacks than they did previously. Some of them use Google Books to preview texts before deciding to seek a printed copy, either from the library or by personal purchase. The increased access to online texts and digital images such as 19th century newspapers and literary manuscripts dating back to the 15th century has greatly improved the speed with which scholars may begin to answer their research questions. And for the two participants who use corpus-based methods in their research, the introduction of digital concordance packages and the availability of online resources such as newspaper data has changed the research landscape.

New questions

All participants felt that the increase of online resources has substantially improved their capacity to conduct research. One specialising in contemporary literature provided the following examples: "Just simple searching across large tracts of press was impossible before" and "because there's so much more data available than 10 or 15 years ago, you can look at more large scale issues." She framed the use of digital resources not as a matter of enabling her to ask new research questions but rather as a means of significantly accelerating access: "I'm not sure about whether they're necessarily new questions. I mean, some of the things could have been answered but through a very, very long process that I can do very, very, quickly now. I think that is the difference."

Others noted that the diversity of data allowed them to expand the breadth of their research questions. One specialising in early modern literature described a research question that drew upon a large digital corpora: "The aim of my work on Donne is to try to place him in the larger cultural field in the period, so I want to know who else wrote things that are similar, and what he was influenced by. The Early English Books Online produces on-screen images, on your desktop, of all the books that are in the short title catalogue from the 1400s up to the 1700s. Now these normally are things that would only be held in rare book repositories." Additionally, as this scholar describes, having access to digitised manuscripts and newspapers has also made it possible to reduce travel time and costs: "[Before] it would take a 3-day research trip - and even then, maybe, these books might be dispersed across different libraries. Now, I can do word searches within the full text."

Future technologies

A majority of those we interviewed described greater access to source material in digital image format as a significant need. Some reported having access only to plain text files, rather than images, of either newspapers or literary works. For both palaeographical and codicological purposes, access to the layout of newspapers and features of a manuscript such as the scribe's unique writing style, or watermarks from the different publishing houses, are important features for their research. Participants were consistent in asserting that increased accessibility to digitised materials such as manuscripts, books, and newspapers would improve research outputs. For these scholars, digitised resources include searchable digital images and text files. They also emphasised the need to reconcile copyright issues so that more materials may be made available in digital formats.

41

UCL

PHILOSOPHY DEPARTMENT

The UCL Philosophy Department was ranked first in the UK Research Assessment Exercise (RAE) in 2008. Members of staff cover a range of research areas, including Philosophy of Mind, History of Mathematics, Ethics, and Metaphysics. The department was selected for the study because of its high ranking within the UK and UCL's extensive online resources.

All academic staff members were invited via email to participate in our study, and all graduate students were invited to participate in our focus group discussion. Six faculty members and four graduate students, representing a range of research areas, participated in our study.

Data collection

Faculty and students were enthusiastic about the breadth of online resources available through the UCL library. For the most part, staff embark on research projects by in-depth reading, often of primary texts. Most faculty own copies of their preferred translations, but favorably describe digitised archives, such as Thesaurus Linguae Graecae⁴³ or Clay *Mathematics Institute.*⁴⁴ All report using UCL's journal databases. Most report using Google for searching, though it is typically not their first resource choice. Online e-book websites such as *Project* Gutenberg, 45 the Internet Archive, 46 and Google Books are commonly used to download entire texts, or to read snippets in order to determine if it is necessary to find or purchase the entire book.

While some of the philosophers we interviewed use Wikipedia, others prefer discipline-specific reference sources, such as the Stanford Encyclopedia of Philosophy.⁴⁷ Many use Phil Papers, 48 an online aggregator of philosophy pre-prints, which is similar to arXiv, the pre-print repository commonly used in physics. In philosophy, posting a manuscript draft online in advance of its publication is common practice. Phil Papers aggregates these papers and allows for keyword search on author name or subject area. It is regarded as highly reliable, with graduate students subscribing to e-mail alerts to learn of new postings.

43

Organising and annotating data

For most participants, research originates with a primary text that they read several times. As one described, "In philosophy, research is more thought experiment than physical experiment." Workflow varied among respondents, yet many reported research as a slow process of iteration: reading texts, thinking, and re-reading. Respondents report a mixed use of different resources, using whichever is most appropriate to the task. One described using printed texts when attempting to cross-check page numbers, but digital copies when searching for a specific term. Most save and print texts they find relevant, with a majority making annotations or writing notes.

Data analysis

The process of analysis and drawing connections between ideas is necessarily slow. Initial surveys of additional translations, commentaries, and criticisms are frequently re-visited as philosophers form an understanding of the topic. Many use notes and filing systems as a means to remind them of ideas, author names, or key articles.

Most striking about philosophers' workflow is its inherently social nature. As one graduate student described: "A lot of how philosophy is done is in conversation." Before the Internet, books and papers were passed around to colleagues for review prior to publication. This practice continues, making strong use of the Internet's affordances. As one graduate student said: "Lots of published work

in philosophy starts off being given in seminars or as conference papers, and as a result, philosophers receive a lot of feedback on their work before it is published." Those we interviewed often post unpublished work on their university website while awaiting publication of a final draft. These works are heavily commented upon, by both invited and unsolicited peer response, which can then inform revisions. Pre-prints are believed to influence other research in a more timely manner than a published article. Papers posted online are expected to be published in a refereed journal; online posting of papers is in no way considered an alternative to publishing in a journal. In published works, philosophers regularly acknowledge those who have read and provided feedback, a practice that predates the Internet.

Citation practices

Most of those we interviewed did not cite online resources when they used them, instead citing print versions, where available. The prevailing rationale was that if a printed version existed, there was no need to indicate that the resource had been accessed electronically.

Dissemination practices

Our participants disseminate their research primarily through traditional means such as peer-reviewed journals, books, book chapters, and conference presentations. In discussions about invited book chapters, many felt that time was better spent on publication in highly-ranked journals, to meet RAE expectations. Online publication, beyond posting unpublished papers, was controversial.

Some junior scholars were uncertain how online publications would be valued. One faculty member said: "I still need to improve my publications record. I think once I've managed to get a couple of things in traditional journals I will probably try to move towards the commercial-free, open, Internet journals." Yet many referred to blogs in their descriptions of useful resources. One predicted that in the future, blogging would be more acceptable: "The barrier between real publications as we used to understand them and mere documents on the web is beginning to dissolve...or is becoming more permeable."

Collaboration

As part of the social nature of the field, philosophers share drafts with colleagues for comment. But collaboration in gathering and analysing data is unusual. Only one of those interviewed actively engaged in collaboration and her experience fell within the category of sharing drafts and commenting on each other's work. Some of the philosophers we interviewed reported ideas for collaboration, but said it wasn't encouraged within the field.

Transformations in practice

Apart from ease of access in terms of time and travel, transformations in practice relate more to resources that are no longer used than to the adoption of new ones. Thus several participants reported no longer using previously popular reference sources such as the *Philosopher's Index, Encyclopedia of Philosophy*, and *L'Année Philologique*, preferring instead

their more portable, easily searchable, and more frequently updated online counterparts. In addition, archival materials are usually accessed online.

The graduate students used more digital resources than did staff. Thus, while one staff member mentioned regularly using 3-5 digital resources, a graduate student reported 10-15. Most subscribe to e-mail alerts through *Phil Papers*, selected journals, Google Scholar, and *ZETOC*.⁴⁹ One graduate student described seeking information about new publications as "just a habit," while another outlined a comprehensive strategy using RSS feeds, email alerts and other services: "I read all the main blogs and all the tables of contents and things like that, so I get book alerts, follow online discussions."

Another student felt overwhelmed by all the information options: "I think there's a general phenomenon with electronic things and technological advancements... If I was to get involved in it, I'd have to master lots of processes to get the most out of it to make it worthwhile doing it at all. And I'm slightly frightened by them. So I guess it comes piecemeal, because I do things now that I've never have done. Three years ago, I found them quite frightening, and I guess that will change over time."

A second-year student believed that the ease of access afforded by digital resources results in higher expectations: "I find from a student's perspective, the information is just much more available, obviously, and there's sort of less excuse not to actually get hold of the paper in an online form, just read it, and to follow up with these different leads because everything is so available."

New questions, new answers?

Nearly all respondents felt that although access to information is faster and easier, it did not result in conceiving or answering new questions. One faculty member states:

... availability of digital texts has encouraged me to pursue interests in rather obscure areas... simply because I get more done in less time than previously."

Yet another felt that the technology innovations complemented his interdisciplinary pursuits, as

when you're interdisciplinary, research generates its own agenda and new connections generate new types of questions altogether. There are new questions, new subject matter, new issues, new conceptual issues...

Overall, philosophers shared a belief that, fundamentally, their research questions had not changed, despite advances in access.

Future technologies

The most common complaint among philosophers was the lack of sufficient annotation tools, in particular for PDF files and for e-readers. All expressed frustration with the need for additional files such as a Microsoft Word document to write their notes. They expressed a wish for simple functionality, such as the highlighting and comment features available in Microsoft Word. As one graduate student pointed out, while the functionality is available, it is clunky.

When asked what they wished they could do with information that they currently are not able to do, a few expressed interest in training. Some felt that their lack of facility with, for example, annotation tools, represented limited competence on their part, rather than a limitation with the technology.

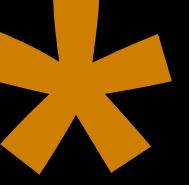


corpus linguistics



Corpus linguistics researchers' methods rely upon the availability of machine-readable texts in the "analysis of specially-designed collections of texts by computer" (Anderson, 2008). Thus, the field is born digital, in contrast to our case studies in areas such as philosophy and English. The *Open to All* study published by RIN/NESTA in 2010 explored the use of open access resources in relation to research involving text corpora (RIN/NESTA, 2010). In this study, we expand upon that work to explore more deeply the breadth of information practices within the field of corpus linguistics.

In corpus linguistics studies, the texts comprising the corpora may be written or spoken and are derived from a variety of sources including newspapers, literature, spontaneous talk, or broadcast. Using computational tools, keywords within texts are analysed with the goal of finding patterns. Developing and analysing the digital corpus requires a unique set of skills. One researcher described the 'ideal corpus linguist' as being "three things in one: a linguist, someone who knows about linguistics [and] could draw conclusions from the data analysis, a statistician, and a computer programmer, who could write code. There are a few people who tick all three boxes and I know a lot of other people who tick two and maybe a bit of the third"(Corpus linguist, Research Assistant/ PhD student).



Participants

We interviewed six corpus linguistics researchers from a variety of fields including Journalism, Education, and Linguistics.

Data collection

Our case study participants use text-based methods to examine a variety of socio-linguistic phenomena. Data collection varied greatly, depending on individual research need. Some scholars used databases popular in their field such as the *British National Corpus, Project Gutenberg, Early English Books Online, Literature Online* or the *Scottish Corpus of Texts & Speech.* Others developed their corpus using interview, conversational, or newspaper data.

Organising and annotating data

Corpus linguists use complex datasets and a range of methods to accomplish their research goals. They assemble a corpus to compare to other corpora or to address a specific research question. In designing a corpus, they may, for example, compare literary texts with each other, or examine how people and events are portrayed within specific genres (e.g. newspapers). They also build a storage system by categorising text files in appropriate directories with detailed notes, and undertake a process of gaining permission to use the corpora, which often involves dealing with both copyright and data protection issues.

Assembling a corpus often involves text capture, working primarily with text files, Optical Character Recognition (OCR) images, PDFs or HTML pages, but sometimes inputting text themselves, normally for conversational data. If using an OCR image, they convert it to text for analysis. The corpus linguists in our study engage in detailed mark-up of the texts, delineating text features such as 'chapters', 'paragraphs', 'conversational speakers' and 'parts of speech' (or grammatical tagging, e.g. noun, verb, adverb, past, present, future). In addition, they undertake a process of annotation assigning a 'part of speech' to words or multiword units (also known as grammatical tagging, e.g. noun, verb, adverb, past, present, future). Corpus linguists also mark up sentences or other text parts for particular semantic or pragmatic attributes.

Plain text files are essential to corpus linguists' work. However, digital texts are often available only in PDF, digital image format, OCR, or presented as HTML. Corpus linguists frequently invest time into re-formatting and cleaning to transform unusable digital files into machine-readable text, before they can undertake any kind of analysis.

Data analysis

Corpus linguists frequently use bespoke software tools or off-the-shelf packages like *Wordsmith*,⁵¹ a tool that recognises word patterns in large text files. New web-based data analysis tools are also being used, including *Sketch Engine*,⁵² a text query system that generates visual relationships between words; and *Wmatrix*,⁵³ a tool that performs corpus comparisons and provides annotated output. Regardless of software choice, Corpus Linguists mainly use four techniques to analyze data.

Word list: this includes a list of all words that occur in a text or corpus and displays of word frequency.

Frequency analysis: the concept of frequency underpins much of the analytical work in corpus linguistics where comparisons are made between the appearance of words in a corpora. It is also used to count grammatical forms such as past-tense, present-tense and future-tense.

Collocation analysis: examines high-frequency keyword combinations; either adjacent (e.g. strong tea/powerful tea), or non-adjacent (i.e. within 4-5 words to the left and/or right of the word investigated). Words or terms that co-occur more often than would be expected by chance are examined.

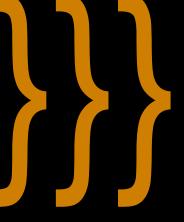
Concordance analysis: a list of specific keywords or collocates displayed within the context for which they were used. The keyword is usually listed within the context of the five words that precede and succeed it.

Citation practices

Researchers cite the name of the corpus used (e.g. *British National Corpus, Project Gutenberg*, etc) in the introduction of a paper and some may also include the web address. Others reference the newspaper used in a footnote, or after each example, in addition to mentioning the means by which articles were acquired (e.g. *LexisNexis*⁵⁴). For purposes of replicability, some researchers also publish their search terms and the date on which they were used.

Dissemination practices

All participants said that they disseminate their research primarily through peer-reviewed journals, books, book chapters, and conference presentations. Some use their university's e-Prints facility to post journal articles and conference presentations. Two participants have published in online-only journals and one commented on the unfairness of judging those as somehow second-rate even though they are peer-reviewed. Only one respondent, a PhD student, reported using Twitter and a personal blog to discuss research topics.



Collaboration

All the participants engage in collaborations, which are conducted for the most part faceto-face and through email. Network drives are the primary means for sharing corpora. Collaborations often cross disciplinary boundaries, and involve work with historians, clinicians, media theorists and others who are seeking linguistic answers to specific questions in their data analysis. As one participant stated: "I've tended to work with people who are coalescing around a particular issue from different directions. "This cross-fertilisation of expertise allows participants to derive much more complex and novel outputs when they tackle research questions from a variety of methodological as well as theoretical standpoints.

Transformations in practice

Corpus linguistics depends heavily on technological innovation and practice adapts to new tools and possibilities. As a corpus linguistics PhD student describes: "Even in pure science people do what technology available at the time allows them to do or gives them ideas." Scholars must be open to new technologies and adapt when they face barriers. A consistent pattern in respondents' choices of tools is accessibility and affordability. They tend to use the simplest tool for the task, often choosing free, task-specific tools over heavier, costly options (as is evident in the popularity of open databases and web-based freeware).

New questions

The availability of large corpora of newspaper data has made new questions possible. For example, one project has gathered 200,000 articles consisting of over 140 million words to trace daily fluctuations in the use of the word 'Muslim' over a 12-year period. Analysing daily fluctuations raises new methodological challenges:

by day, you confront what I'd call time series data that's rather messy. You have a sampling point everyday if you want it and mentions can go up and down at an alarming rate. How do you decide what is a peak or a trough of interest or business as usual?

The researchers therefore consulted with a distinguished statistician who showed them "some clever maths to smooth out the data and show us what we wanted to see." Using this analysis method, the researchers can show a relationship between the frequency of the term and the political leanings of newspapers they sampled.

Future technologies

Two participants were particularly interested in using the Enroller project (An Enhanced Repository for Language and Literature Researchers) web portal when it becomes available. 56 This portal will provide access to distributed language and literature resources, "enabling research between and across different types of literary and linguistic data while harnessing powerful e-Science services and data analysis tools." Additionally, solutions to systematically examining spelling variations in historical text mining are being developed. Some scholars expressed interest in using visualisation tools to examine how word use changes over time. New tools are constantly being developed to meet the needs of scholars' research questions and they are becoming increasingly available online, rather than developed as stand-alone applications.

Some researchers are beginning to use the World Wide Web itself as a source of corpora to be studied. One researcher analysed discussion forum data within a special interest group website. Another is harvesting the front page of two newspapers, The Guardian and The Daily *Telegraph,* to track the representations of specific social groups. As one participant described: "The web has actually proved an enormous boon for corpus linguists. Of course it is the bane at times as well, but it's also a boon. And it's a good basic resource from which to begin to mine data of all sorts. The transformative element is everything appearing on the Web, all that electronic text there. Last year [a PhD] student] was looking at hate speech and found all sorts of websites where people from some right wing organisation posted messages to one another and it was the perfect fit between researcher and data."



THE DIGITAL REPUBLIC of LETTERS

The Digital Republic of Letters⁵⁷ is an international collaboration of humanities scholars, project managers, database programmers, and data visualisation specialists developing an online research collaboratory of 17th century correspondence. The correspondence include works by famous scholars such as Descartes as well as ordinary citizens. Unlike a straightforward digitisation project, *The Digital Republic of Letters* aims to move beyond a repository and support humanities scholars' collaboration online. Individual projects involved in the collaboration are funded by substantial grants from public bodies⁵⁸ and from private foundations. The group consists of nine organisations in four countries, with around 25 contributors at any given time. It was selected for this study because the breadth of its collaboration and research would have been extraordinarily difficult to accomplish pre-Internet.

Participants

Nine members of the project were interviewed individually for this study. Following the interviews, we conducted a focus group discussion with four of the original interview participants and three additional group members. In addition to examining collaboration strategies, we also explored individual information use and ways in which the project affected members' research practice.

How the collaboration works

The Digital Republic of Letters is organised under overarching research and methodological goals, with individual projects managing their participants and deliverables. Work ranges from scholarly editing of correspondences, such as those from Grotius,59 Huygens,60 or Descartes,61 to developing databases, to creating visualisations of relationships between frequency of correspondence and geography. Project members engage in frequent e-mail communication and, where possible, regular face-to-face meetings with local collaborators. There is a rigorous schedule of deliverables. Each quarter, collaborators attempt to meet in person to demonstrate new features, developments, and research findings. Additional collaborators are invited to share expertise. For example, members of the InfoVis62 data visualisation lab at Indiana University Bloomington traveled to the Huygens Institute⁶³ in the Netherlands to train their technical team, and then the Huygens team traveled to Indiana to engage in a similar exchange.

The project focuses on developing a digital archive, tools for analysis, and a collaboration space. Within each category, project

members address challenges related to both scholarship and technical development. For example, in developing a digital archive, project members sought agreement on standards such as language use, metatags, access, and information presentation. Discussions among the project's intellectual historians, programmers, librarians, and project managers inform these decisions.

Most participants do not view this work as separate from their research. While participants often focus on a particular scholar, time period, or region, they also demonstrate specialisation in the methods of their craft, and apply them to the development of analytical tools and archives. For example, some project members develop tools for visualisation while others use these tools to visualise their data. While not seamless, participants' project development and research work appear to share complementary aims. When respondents discussed barriers to completing their research work, they spoke of what wasn't technically available, or the limitations of current tools, rather than the time they devoted to project coordination and communication.

The interviews uncovered a range of experiences, but more importantly, a shared willingness to work slightly beyond members' comfort zone. IT staff are tackling humanities questions from a technical perspective, while humanities scholars experiment with visualisations, regardless of a self-reported steep learning curve. Similarly, other scholars have assumed an administrative role, even though this commitment entailed a reduction in their research time.

Data collection

Historians involved in this project cited digitised primary texts as the key resource(s) available to them. They are heavy users of digitised archives including Eighteenth Century Collections Online, Early English Books Online, Gallica,⁶⁴ Dutch Union Catalogue of Correspondence, 65 Uppsala Waller⁶⁶ Collection, Hartlib Papers,⁶⁷ and ARTFL.68 Most use Google, Google Scholar, and Google Books as starting points, and all of them access journals through their library's databases, most frequently mentioning JSTOR. Many have developed an archive, database, or analysis tool that they also use regularly in their research. Several members of the project cite discussions with colleagues-most often through sharing resources via e-mail-as an important component of their research. In fact, after primary texts, colleagues are seen as the most important resource.

Organising and annotating data

Organisational strategies vary among the project members, with the more technically savvy members using document management systems and databases, or LaTeX for word processing. They feel confident about their organisational practices, even when they describe themselves as disorganised or haphazard.

Most avoid printing unless the document is a key resource or difficult to read onscreen (e.g., is lengthy, has multiple columns, or is in a hard-to-read font). The choice not to print does not appear to be a result of the affordances of reading onscreen, but rather a choice to conserve resources. They do not make much use of annotation, such as highlighting key pieces of text or writing notes in the margin. But most use copy/paste to manage a document file containing quotes and useful notes that they later re-organise or copy/paste into their article drafts.

Data analysis

Data analysis is a cumulative process, often occurring over several years or even decades. For many, this process begins with editing primary texts and developing a theory around a particular aspect of the work and the relationships between individuals, concepts and events. It is at this point that many respondents sought more advanced means to realise their research objectives. For one scholar, exploring relationships meant manually comparing maps, historical records, and correspondences. Other project members reported that they developed preliminary databases to explore the relationships, while some learned to use visualisation tools for the same purpose.

Many use visual terms to describe their understanding of these relationships, describing the need to "see," "visualise," or "map" them. They describe the work they are doing as "new" and something that "doesn't exist yet." Their analytic processes consider the multi-dimensionality of their subject, spanning geographic, social, and time boundaries. Project members also describe serendipitous moments with their large datasets. For example, one scholar assumed that Descartes would frequently use the term 'observation' in daily correspondences regarding his experiments. Surprisingly, when he ran the

term through his database, 'observation' only occurred in relation to five personal names, all of whom were astronomers. He describes the experience: "But why these? I understood why these would come up, but not why in this order, why not, why on top, and I only could interpret that when I had finished my study, because then I noticed that the term observation is used quite loosely in most contexts. It doesn't have a very fixed meaning. It could mean 'observation,' it could also mean 'experience.' But not so in astronomy, so that's why these astronomers appeared, because these were closely, most strongly associated with the search term 'observation.' That's what kind of extraordinary results it can give, but you have to be able to interpret them, be able to expect, well, unexpected results." During the analysis process, many project members describe using the databases to test theories and, as the Descartes scholar describes, being prepared for unexpected, sometimes serendipitous, results.

Citation practices

Project members report citing digital resources when they are 'born digital,' but for journal articles or books that originate in print, they cite the print version. Many acknowledge the need for new standards in citation practices. One describes feeling "sneaky" when not mentioning the digital resource: "because I use *ARTFL* so much, and I work with them, I had come to recognise the editions that they've used, and then I'll see other scholars citing these weird early 20th century editions of 18th century books, and I'll say, wait, I know where you got that, that's from *ARTFL*! But, you know, of course they don't cite it. And it just seems like we're sort of going around acting

as though there's something wrong about this, and that I feel like we should start to change that culture, too." Project members who develop databases and digitised archives report including citation guidelines in their interfaces.

Dissemination practices

Dissemination occurs through project members' collaboration with colleagues, development of digitised archives and databases, and publishing in traditional and online venues. Collaboration with colleagues involves detailed content and methodological discussions, often establishing or challenging standards in their field. As a collaborator from Stanford explains: "Because we're trying to cross disciplines – in some cases those are really big gaps in others not so big – we're still having to spend more time clarifying, communicating." As mentioned earlier, knowledge dissemination in this case includes sharing methodological expertise, but also training in using the technologies. One of the project managers frequently demonstrates the capabilities of the archives and databases at conferences and feels this is a critical part of the project's dissemination.

Many project members also contribute to database development and digitization beyond the *The Digital Republic of Letters* project. Typically, they become involved with development/programming because of a perceived gap in either content or analysis functionality. A scholar from Utrecht described the digitisation of Grotius' correspondence: "Henk [Nellen] ⁶⁹ started after the Second World War editing the Grotius letters. Only editing. And now of course, they have made

an online edition of that." Through the digitised archives, Nellen's work is available for any interested scholar to access.

Likewise, other project members wanted to provide tools to help draw connections between where and to whom the letters were sent, as demonstrated by Stanford's *Mapping the Republic*⁷⁰ project. Knowledge dissemination in this case takes the form both of content and of the ability to manipulate that content.

Project members publish primarily in traditional venues such as articles and books, and present at conferences. While most do not blog, use Twitter, or use Web 2.0 technologies to communicate their work, a quick search of their names indicates that they have a strong web presence. This results mainly from their involvement in online projects that are reported by the popular press, universities, and bloggers. Project members' critical approach towards currently-popular communication platforms is worth noting: not using Twitter or blogs is a well-informed decision. One states: "I find it useful in a way, but it's also worrisome. It's almost like who is most popular right now, who is getting the most tweets and re-tweets, and it's a distortion of research. I think it can be distracting, especially to grad students, when they're trying to navigate, when they're needing to learn, adopt, and use these new technologies, but at the same time learn to discriminate among technologies that are more for social things but are being used in the name of research. The lines are too blurry." A colleague in the Netherlands expresses concern that these technologies are time-consuming:

I prefer not to play on too many different platforms. You can blog, you can twitter, and I know it's all there, but I don't like it. I find it too time-consuming to work in five or four different sort of environments so I try to limit it to... disseminate in just papers and conferences and much of my work is in email.

Collaboration

Consistently, project members referred positively and enthusiastically to the collaboration, the topic of study–17th and 18th century scientific letters–and noted that was a model for the type of collaboration they were now practicing: "I think it's partly the very nature of the correspondence, because the wonderful thing about letters is they connect everything with everything thematically as well as geographically and perhaps even socially."

We have one information technology revolution responding to another one, the other one being the development of postal systems, which allow relatively ordinary individuals without enormous expense to scatter archives of their own written material across whole continents, imposing enormous difficulties to scholars, because you can't possibly go poking around in every archive where your material might turn up.

The structure of the team includes two project managers, and IT staff with an interest in humanities carefully selected to complement the goals of the team. A project manager in the Netherlands describes their search for a developer:

...although you're looking for somebody who is experienced as an IT developer, you also need somebody who has a good feeling for humanities, of language research, or history, and those are quite hard to find...because you need the development expertise, but also somebody who can understand why you want to develop something, what's the deeper meaning...

One of the developers in the U.S. describes his work: "We have some practical needs... where we'd like to be able to generate cross-references or indexes or anything else...but we also have some more research-y kinds of goals, so if somebody is working on a particular topic, it would be really nice if they could get a visualization of all the other topics in the neighbourhood and what other kinds of resources they might want to look at." The work thus reflects a trend discussed at the *Digital Humanities 2010*⁷¹ conference, that humanists often take an active role in the development of technologies to support their research. One collaborator states.

In humanities, there's an intellectual argument as to how you choose to use technology, how you present it. For our project, we treat the development of the technology and the more traditional humanities research as equally important.

Solid financial resources and strong structural vision underpin the highly productive collaboration.

A recurrent theme in the interviews was the experimental nature of the project. Work of this nature and scale is still relatively unusual in the humanities, and many referred to similar ties with work in the sciences.

Transformations in practice

Ease of information access results in fewer trips to the library, and participants use library books less frequently than before. But, as one scholar describes, digital collections and online books may "widen the audience for the texts, which is very important, but it has nothing to do with digital revolutionising the humanities. It's something practical. It is a publication strategy but it's not a research strategy." When describing changes to their research practices, participants describe ease of access as a convenience, but not necessarily a factor that has brought great changes in their research methods.

On the other hand, advanced methods of analysis afforded by databases and text mining tools are having a significant impact: participants can now undertake the kinds of analysis that were technically impractical in the past.

New questions, new answers?

When describing their early careers, many felt a sense of frustration, because they had research questions and a sense of how to answer them, but the means did not vet exist. For example, one scholar remembers his early attempts to examine a large dataset: "I think I was 23... I wanted to do research on the infrastructure of the Netherlands in the 18th century and the best way of doing that was trying to reconstruct the membership of learned society. Of course, then you have too many people [to study] and so I went to the mathematical institute to see if they could help me out. Now they thought the questions were interesting and they could help me out but did I have the money to buy computer time? Of course it was the pre-desktop computer period and did I have this money to buy punch cards? Did I have money just to pay for punch card typers? I didn't have that kind of money, so I continued doing it just with my small cards but then I knew this had to be changed." This example illustrates the technical, conceptual, and financial barriers some project members faced early in their careers.

Unlike other groups in which respondents insisted that their questions had not changed, the *The Digital Republic of Letters* members believe that by enabling more complex analysis, new technologies allow for new questions. "We're going from simple word retrieval to text-mining, so rather than treating a digital version of a book just as a digital version of a book, seeing it as an opportunity to ask new questions and to read it in an entirely different way."

The scale of analysis has thus expanded from a single archive, author, or text to a

broader corpus. By broadening their scope, they can seek connections between people or events that they would have been unable to pursue in the past. By visualising these connections, they can answer new questions about relationships and knowledge exchange.

The types of questions they address involve the movement of letters throughout the world. In particular, the collaboration between the Huygens Institute and Stanford University traces letters from Dutch philosophers in the 17th century and maps points at which correspondences were exchanged, creating correspondence networks.⁷²

Future technologies

Participants identified a number of limitations of current technologies, and looked forward to progress on practical matters such as a universal calendaring program, and advanced functionality such as incorporating visualisations of related topics in database searches. Some simply wanted more archives to be digitised. Others called for more uniform standards for digital archives and crossreferencing capabilities across datasets. They recommended services and resources, such as ARTFL, the Upsalla Waller Collection, and the Stanford Encyclopedia of Philosophy, that they felt represented models for the future. One scholar recommended a "dashboard that allows options for text mining, analysis, and visualisation." Another recommended smarter search tools: "If you're just learning about some topic you might want to be pointed towards a higher level article that will give you an overview before you burrow into some of the more specific specialist kinds of topics."

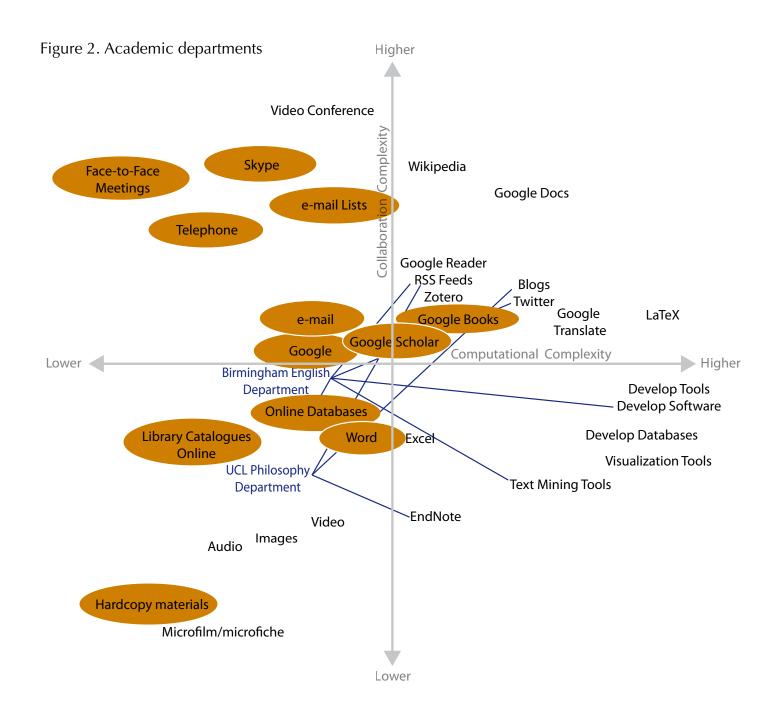
FINDINGS ACROSS CASES

Complexity continuum

Earlier in this report (see page 14) we introduced the idea of the complexity continuum as a way of conceptualising the relationship between technology and collaboration in a variety of disciplines. Before collecting the data for this project, we made some initial assumptions about technologies that might be used by humanities scholars, and their location on this grid. Now, we can return to the complexity continuum, and re-examine the cases in light of the data collected, noting that complexity here refers to computing and collaboration, not to the intellectual challenges inherent in humanities research.

The following diagrams show the types of information sources mentioned across the six cases. Their positioning is based on a qualitative judgment, so these visualisations should be seen as illustrations of our understanding of these resources and their use. First, looking at the data from the two departments we studied (Birmingham English and UCL Philosophy), presented in Figure 2, we can see a number of resources shared by the cases (represented by the orange clouds). The lines of the diagram link each to some of the resources that are less common across all six cases.

For the Birmingham English department, along with the full range of simple collaboration tools such as e-mail and Skype, scholars are using RSS feeds and Twitter for current awareness. They are also engaged with more complex computational activities including developing software and using text mining tools. Likewise, the UCL Philosophy department, which is less likely to engage with collaborative computational information uses, nevertheless relies on resources such as blogs and newsfeeds to engage with other researchers.



In Figure 3 the two resource-centred cases are somewhat further to the right on the computational complexity axis. Unlike subjects in the department-based cases, users of *Old Bailey Online* and *DIAMM* are developing tools, software, and databases and are using visualisation tools. Digital humanities initiatives, including digitising resources, building databases and data interfaces, and creating tools for broader use by humanities scholars are becoming increasingly common. They often involve both computer scientists and humanities scholars with domain expertise, such as is the case with both *Old Bailey Online* and *DIAMM*.

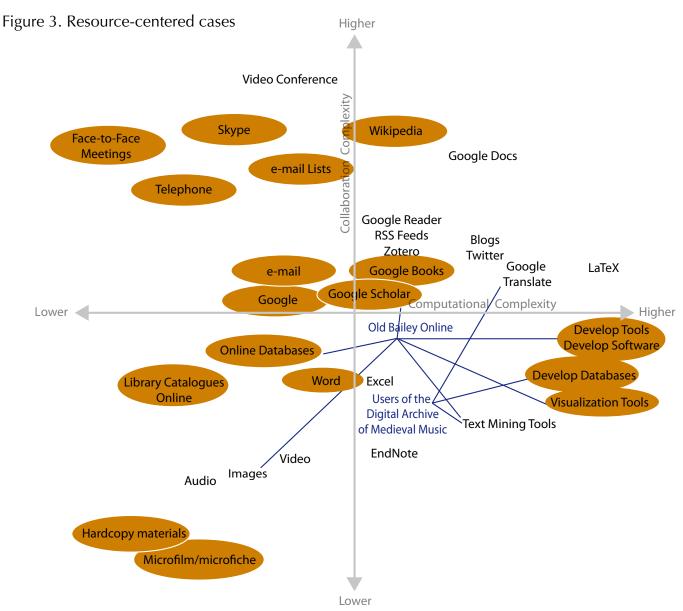
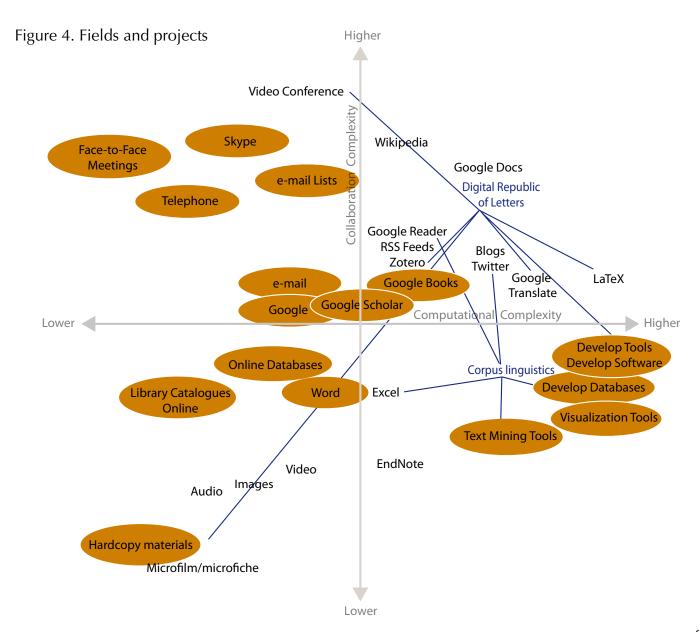
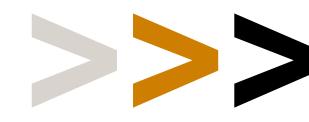


Figure 4 shows, as might be expected, that corpus linguistics and the large-scale *Digital Republic* of *Letters* project are further along the computation complexity axis to the right than the other cases. The *Digital Republic of Letters* is the case furthest to the top along the vertical collaboration complexity axis. Both of these cases are using computational resources to develop complex tools and databases. In both cases, however, more traditional resources, such as hardcopy materials and library catalogues, remain important. This suggests that researchers are not moving *from* less complex information uses *to* more complex ones, but are *broadening* their information ecosystems.



This broadening of the information ecosystem has important consequences for information providers. As humanities scholars deal with what has been called information overload or the data deluge, how will they juggle the resources competing for their attention? They are adept at finding and using the kinds of information sources they need to advance their research, and some are engaging in complex, cutting-edge projects. Whether these pioneers are paving the path for large numbers of humanities scholars to follow, or whether they are exploring a path that will diverge from the majority of humanities scholarship, however, is not yet clear.

Researchers' information uses cover a broad range of behaviours and technologies. While we have centred individual cases at points along the spectrum that typify the responses in each case, in reality the information uses in all the cases show a great deal of variety and complexity. It is also important to note, however, that there is little evidence of much involvement in activities in the top-right hand portion of the complexity diagram, indicating that the cases we studied had little evidence of highly collaborative computational approaches to research. In follow-up work, we are examining the physical sciences, and early indications are that there is much more activity in this top-right sector. These differences suggest that while on the one hand, the humanities are more complex and varied than once thought, they are still not engaged with leading-edge technological developments to the same extent that some other fields appear to be.



ENGAGEMENT WITH DIGITAL RESOURCES

Browsing

We gave participants in this study a list of 15 possible methods they might use to learn of new, relevant research, and also allowed them to list additional ones. In their responses, participants demonstrate hybrid use of their library's online and print offerings, but a distinct preference for using online journals compared with browsing print versions. Regardless of mode, library use was high overall, with 62% using keyword searches and 55% browsing library materials online. A similarly high percentage (59%) visit the library to browse materials in person. Although a higher percentage browsed online journals (76%) than the printed versions (48%), such responses indicate that use of print resources remains a significant part of humanities scholars' work, alongside online activities.

Beyond the use of journals, we found the tendency toward browsing depended upon the tool. For example, users of the Old Bailey Online often describe their research in terms of keyword search; many track specific people or a specific type of crime over time, using keyword functionality. In contrast, those using databases which lacked a reliable keyword search or consisted mostly of images tend to rely more on browsing. Scholars using DIAMM, a visual archive of music manuscripts, tended to examine single manuscripts, rather than browsing the collection.

As Harley et al. found (2009), Google use is widespread among humanities scholars. A majority of respondents use Google (79%) and/or Google Scholar (66%) as a starting point

to locate relevant research. But traditional methods, such as citation chaining (83%) and learning from peers and experts (95%) remain the most significant ways of finding resources. The scholars in our study also subscribe to e-mail lists (66%), with many using RSS feeds (31%), social networks (48%), or other notifications (42%). These numbers reflect a consistently mixed use of traditional and newer information resources and technologies and suggest a thoughtful engagement with the new technologies that best complement their research needs.

Many participants in our study frequently use digital archives; while views are mixed, most find that they increase ease and speed of access. Similar to the findings of the CLIR report (2009), some of our participants expressed concern about the selective nature of some digital archives, wondering what may be missing or given priority. Members of *The Digital Republic of Letters* project were especially concerned that choices to digitise some texts rather than others may result in texts being overlooked or lost: easy access to digital resources might make scholars reluctant to seek manuscripts and print in remote archives.

Many scholars reported fewer trips to the library, citing the inconvenience of returning recalled books or simply the trouble of leaving one's home or office. Frequently, when a resource was not immediately available, scholars sought additional information to evaluate its usefulness compared with the time necessary to gain access to it.

Search

Scholars' web histories show that they spend significant time dealing with the complexities of searching for relevant material. For example, many historians have to deal with inconsistent spellings in public documents as well as newspaper reports. Susan Sommers of St. Vincent College provided the following example of spelling variations she had to use in her research for information about Thomas Dunckerley:

Dunckerley Dunkerly Dunckrley Tunikerly Tunikley Tunikly Dunckesley Donnersly

These nine different spellings present a challenge to those developing databases and keyword searches, and they can affect the accuracy of search results using both Google and more specialist discipline-specific databases. Doctoral students using *Old Bailey Online* thus voiced concerns that they might miss information about particular individuals because of such spelling variations. Users have to modify their search several times, adding last names, removing last names, or using another detail, such as occupation, to narrow or broaden their search.

As Sommers notes, titles and initials can also cause problems, as illustrated by the following example:

E. Silby
E. Silbey
Ebenezer Silby
Ebenezer Sibley
Dr. Ebenezer Sibly
Dr. E. Sibley M.D.
Professor Sibly
Astro-Philo Sibly
Dr. Sibly

All this can make the search process timeconsuming: scholars spend much of their time modifying search terms in order to achieve greater reach and accuracy.

Conclusions

Humanities scholars produce and use information resources in diverse and rich ways. While sharing a focus on texts, our scholars represented a broad spectrum of research approaches and of technology use and awareness. Through our case selection, we covered a spectrum of engagement with new technologies. The Old Bailey Online and DIAMM cases are based around digital resources, and thus their users are likely to be comfortable with digital technologies, if only as a result of engaging with those particular resources. The Digital Republic of Letters collaborative project required its participants to engage directly in creating a digital resource. The discipline- and department-based cases captured a wider range of engagement, including more reluctant, less experienced and occasional users of digital resources and tools.

Research process

Most of the scholars to whom we spoke describe their research practice as something that occurs in bursts, during breaks, when they are not teaching, or engaged in administrative duties. This underscores the multiple time demands placed on scholars, who must often be creative in finding time to spend on serious research. It also explains, at least in part, the attractiveness of digital resources which are readily available; and a reluctance on occasion to consult text that require a trip to a library or archive, especially if that is some distance away.

Workflow

Most of our respondents start their searches for new information with Google, but many still consult printed texts, from their bookshelf or university library. Even when staying in an online environment, Google often serves merely as a starting point, from which scholars move to other resources such as specialised databases, archived primary texts, or online book collections. Management of data is usually accomplished on their computer desktop, with storage of PDF and image files the mostcommonly reported practice. Sharing of data or documents is most frequently accomplished through e-mail, though in the corpus linguistics case, two scholars report delivering data by hand. Analysis involves literary interpretations, image analysis, text mining, and data visualisations, among other practices. Further, many researchers contribute to the development of digital archives, databases, or analysis tools.

Collaboration

Humanities scholars have for long situated themselves within larger intellectual communities, as evident in conferences, workshops and seminars. Collaboration in this sense involves presenting work to peers and receiving feedback, a practice that pre-dates the Internet and is consistently part of developing ideas prior to submission for publication.

But many of our respondents are now collaborating with others more systematically. This growth in collaboration results in part from new funding opportunities, but also from the possibilities opened up by new technologies. In *The Digital Republic of Letters*, the number of texts available requires input from scholars working in different locations (where the letters are) as well as considerable technical support. They collaborate in order to address their key research questions about relationships and the dissemination of scientific knowledge in the 17th century.

Dissemination

Disseminating research through channels such as highly-regarded journals, conferences, monographs and book chapters remains the dominant approach, but over a third of our scholars have also published research online. However, many felt the RAE and other assessment systems did not give proper recognition to non-traditional outputs, and this limited their take-up.

Conversely, a few respondents explained they had not yet blogged because they did not have content interesting enough to an online audience including non-academics. Avoiding online publication is often characterised as arising from concerns about reputation or quality, but in these cases, the concern is about presenting content appropriately to a non-academic audience.

Mixed practices

Speed and convenience is the main reason for using digital resources. But many scholars use resources such as Google Books to evaluate the relevance of a text before borrowing it from a library or purchasing it, thus blurring the boundaries between the use of online and offline resources.

Our findings resonate with those of a CLIR (2009) report which states that "many will consider the electronic surrogates as an important research aid, but not necessarily a replacement for the original texts." None of our scholars was ready to abandon all print resources in favor of digital texts. Rather, they engage thoughtfully and critically with a range of resources technologies.

Thus they use generic as well as disciplinespecific tools and sometimes repurpose tools to suit their needs. For the most part, they use what works, finding technologies that fit and complement their research, and resisting any pressure to use something just because it is new. Thus, in the UCL Philosophy department, scholars described making proactive use of an online grammar tool, but also printed reference materials not yet offered online, moving seamlessly between them. Such behaviours are likely to persist for some time, depending on the nature of the research, and the stage of the research process.

Hybrid scholars

The model of computer scientists telling humanists what they can or should do in using new technologies seems to have passed. It has been replaced by a two-way collaboration between technologists and scholars, to the advantage of both.

Many scholars take ownership of the digital tools they use and are active contributors to the conception, development, testing, and implementation of new ones. There is a sense of malleability when advanced users describe their use of digital tools. At Digital Humanities 2010, a presenter asked whether engineers are bad humanists, or humanists are bad engineers. In our interviews, we found that neither was the case. In The Digital Republic of Letters project, the project leaders selected programmers with literary backgrounds and humanists who understood programming. Together, through many face-to-face meetings and sessions to share expertise, they are developing a digital archive and collaborating on the types of research made possible by this new resource.

Transformations in use

Traditional forms of output and of citation may slow down the wider diffusion and recognition of enhanced publications and digital resources. Many scholars seem to prefer to cite paper sources even when they have used an online version, largely to ensure a correct and durable record. Such practices reinforce the dominance of the printed version, and underscore the difficulty of assessing and documenting the impact digital resources are having on scholarship and learning, a key concern for many institutions and funders in recent years.

Our scholars have yet to take full advantage of the possibilities for using large-scale databases, advanced text mining, and other tools in order to address new research questions. They typically describe using digital tools as pre-determined by their design. Many users of *Old Bailey Online* thus describe their current and future research in terms of the keyword searches they are able to undertake. More advanced users, on the other hand, and those engaged in *The Digital Republic of Letters* case, envisage uses beyond current tool capabilities, suggesting ways to realise research goals through tool development, rather than the other way around.

Our scholars also showed limited uptake of advanced tools for data management and sharing. Simple, free file-sharing tools such as Dropbox⁷³ were not mentioned as part of regular practice. Moreover, there was limited awareness of the potentials of infrastructural developments, such as grid and cloud computing, the semantic

web, high performance computing, and what these might mean for the research process. They also reported limited use of social media or web 2.0 tools, although doctoral graduate students in the UCL Philosophy Department reported frequent use of blogs, RSS feeds, and social media to learn of and disseminate new work.

Barriers to use

Barriers to more widespread and effective use of digital resources include:

- lack of awareness of tools, and of the potential of standard software;
- lack of standardisation of online databases and archives;
- inadequate annotation tools;
- difficult and unstable access to remote resources;
- lack of institutional training and support;
- irregular use of resources, requiring scholars to engage repeatedly in a learning curve – especially for those who pick up tools and/ or resources for particular projects and then do not use them again for a year or more, by which time the interface may have changed.

Difficulties in data linking are perhaps the most significant barrier across cases, arising from the lack of standardisation, and the inconsistencies in quality, reliability, and functionality across different resources. They result in delays in research, repetitive searching, and limited ability to draw connections between sources. For users of *Old Bailey Online*, however, the new *London Lives* database, linking records from 15 datasets, is regarded as an exemplar, enhancing their research capabilities.

POLICY CHANGES & RECOMMENDATIONS

The previous report in this series stressed the diversity of research in life sciences, and this report does the same for humanities. But while some of the life scientists were engaged in "vigorous and extensive information sharing and re-use" the humanities scholars studied in these cases, while increasingly collaborative, are not as intensively focused on sharing and re-use. While some might argue that this is because humanities scholars lag behind the scientists, it seems rather that humanities researchers are simply following a different path. The evidence suggests that, when computational approaches will enhance humanities research, humanities scholars are often enthusiastic adopters of appropriate technologies.

On the other hand, humanities scholars are less likely than life scientists to be tinkerers, comfortable with changing and adapting technology to meet their needs. When presented with off-the-shelf resources, including online resources which have been built by resource providers, they critically evaluate those resources and choose whether to use them in their work. They are far less likely to try to adapt the technology or to build new features, unless they have existing partnerships with computer engineers or other technology experts.

Thus, in this report we can echo the key conclusion in the life science report: "The policies and strategies of research councils and information services providers must be informed by an understanding of the exigencies and practices of research communities if they are to be effective in optimising the use and exchange of information, and in ensuring that this is scientifically productive and cost-effective... A one-size-fits-all information policy will not be productive or effective" (p. 51). Our case studies in the humanities underscore this point – the important information resources and the ways they are used in the humanities are markedly different from those in the life sciences. Strategies for engaging humanities scholars with new types of information and new ways of working with it will require sensitivity to their existing cultures and practices.

Research Councils and other research funders

Our primary recommendations address data linking and training.

Data linking

Whereas the life science report identified data sharing and exchange as a key issue for Research Councils to consider, our humanities scholars consistently spoke of a different approach. One of their main concerns is to be able to link data housed in various archives to support the kinds of research in which they engage. The tools and methods for creating and using linked data must be user friendly, and much of the technology underlying the linking can be hidden in the default interfaces, but should be accessible to the technology partners that humanities scholars work with using tools such as APIs, RDF, and other new technologies that will be developed in the coming years.

Support for researchers

We also found that humanities scholars need training and support in their uptake of digital resources. In addition to investing time in learning tool-specific skills, scholars make an intellectual investment when engaging with a new tool or database. The *Old Bailey Online* case illustrates this point: scholars described

their research in terms of the capabilities of the resource, demonstrating that interaction between scholar and resource influences the types of questions they can ask. For those scholars, changing systems costs more than just time.

On-demand support to trouble-shoot technical problems is clearly needed. However, support in the form of training and peer-modelling is needed too. The Digital Republic of Letters serves as an obvious model because the infrastructure provides both dedicated technical support and peer training. Training could address availability, use, and creation of digital tools and resources, and humanities scholars are far more likely to respond to training that is presented by experienced domain specialists rather than by technology trainers with more generic interests. Seeing specific applications of technology, and how they have been used by their peers to pursue real, current research, is far more valuable than simple how-to courses which focus on generic uses of a tool.

Higher education and research managers

As noted in the life sciences report, academic life is built around career development and reward systems. Resources which allow scholars to be more productive, to use materials more easily, effectively, and quickly, are therefore likely to appeal to researchers with heavy demands on their time. Conversely, resources that require extensive time investment, have a steep learning curve, or offer vague and distant potential payoffs are less likely to be adopted. However, developing appropriate resources requires the participation of domain specialists.

How can these efforts be rewarded? While many institutions and funding bodies suggest that they value collaborations across disciplines, when evaluating the productivity and impact of scholars, the measures are still tied almost exclusively to narrow disciplinary concerns. Explicit recognition and rewards for contributions to building, using, and enhancing new shared resources will increase the likelihood that humanities scholars will participate in building new resources, and thus the likelihood that the resources themselves are designed so that they can be used effectively.

Publishers

Publishers play an important role as providers of and partners in many of the digital information resources on which humanities scholars rely. However, assessing the impact of digital resources is hampered by the tendency of humanities scholars to cite only the paper version of a resource. A majority of our scholars are not consistent in citing the digital resources they use, because of concerns about the legitimacy of online resources, and about disappearing links. There is also a lack of agreed citation standards for long-term persistent referencing and easy access. This is being addressed by the use of DOIs,74 but researchers are still not consistently using them. Building systems that include recommended citations and DOIs will help both scholars who are trying to follow a chain of citations, and those interested in understanding the impact of research.

Library and information service providers

Our interviews uncovered concerns about the sustainability of digital resources, and in particular about incomplete digital archives. They also revealed frustration about the lack of linking between archives. Geoffrey Rockwell, a lead developer of TAPoR observed that resources are often developed without an infrastructure to sustain them. University libraries and information services could help by providing a structure that maintains digital tools and archives and promotes development of innovative resources.

Learned societies and professional associations

Learned societies and professional associations communicate state-of-the-art practice to their members through conferences, journals, websites, and other communications. Many provide sessions on topics such as digital humanities, but these are not yet embedded as part of scholarly standard practice. Societies and associations that bring digital approaches into the mainstream of their conferences and publications will send the message that such approaches are increasingly an integral part of what it means to be a humanities scholar, without diminishing the importance of traditional skills and approaches to research.

References

Anderson, W. (2008). Corpus Linguistics in the UK: Resources for Sociolinguistic Research. *Language and Linguistics Compass*, *2*(2), 352-371.

Barrett, A. (2005). The information-seeking habits of graduate student researchers in the humanities. *The Journal of Academic Librarianship*, *31*(4), 324-331.

Bates, M. (2002). The cascade of interactions in the digital library interface. *Information Processing and Management,* 38(3), 381-400.

Bates, M. J. (2005) Berrypicking. In K. E. Fisher, S. Erdelez, & L. McKechnie, (Eds.), *Theories of Information Behavior* (58-62). Medford, NJ: Information Today.

Bates, M. J., Wilde, D. N., & Siegfried, S. (January 1993). An analysis of search terminology used by humanities scholars: The Getty online searching project report number 1. *Library Quarterly*, 63, 1-39.

Bates, M. J., Wilde, D. N., & Siegfried, S. (Winter 1995). Research practices of humanities scholars in an online environment: The Getty online searching project report no. 3. *Library & Information Science Research*, 17(1), 5-40.

Borgman, C. (1996). Why are online catalogs *still* hard to use? *Journal of the American Society for Information Science*, 47, 493-503.

Borgman, C. (2009). The Digital Future is Now: A Call to Action in the Humanities. *Digital Humanities Quarterly, 3*(4).

British Academy. (2005). E-resources for Research in the Humanities and Social Sciences - A British Academy Policy Review. Retrieved from http://www.britac.ac.uk/reports/eresources/report/

Brockman, W. S., Neumann, L., Palmer, C. L., & Tidline, T. J. (2001). *Scholarly work in the humanities and the evolving information environment*. Washington, DC.: Digital Library Federation Council on Library and Information Resources.

Brown, S., Ross, R., Gerrard, D., Greengrass, M., & Bryson, J. (2006). RePAH: A User Requirements Analysis for Portals in the Arts and Humanities. Retrieved from http://repah.dmu.ac.uk/report/pdfs/RePAHReport-Complete.pdf

Bulger, M. (2009). *Online Literacy and the Trouble with Information*. Ph.D., University of California, Santa Barbara, Santa Barbara. Retrieved from http://pqdtopen.proquest.com/#abstract?dispub=3350331

Bulger, M., Mayer, R. E., Almeroth, K. C., & Blau, S. D. (2008). Measuring learner engagement in computer-equipped college classrooms. *Journal of Educational Multimedia and Hypermedia*, 17, 129-143.

CIBER. (2008). Information Behaviour of the Researcher of the Future. Retrieved from http://www.bl.uk/news/pdf/googlegen.pdf

CLIR. (2009). Working Together or Apart: Promoting the Next Generation of Digital Scholarship (Report of a Workshop Cosponsored by the Council on Library and Information Resources and The National Endowment for the Humanities). Report. Washington, DC: Council on Library and Information Resources. Retrieved from http://www.clir.org/pubs/reports/pub145/pub145.pdf.

Cothey, V. (2002). A longitudinal study of World Wide Web Users' information-searching behavior. *Journal of the American Society for Information Science and Technology*, 53(2), 67-78.

Deegan, M., & Tanner, S. (2002). Digital futures: strategies for the information age. London: Library Association.

Dirksen, V. (2007). Social Imaginaries of Technology and Work: A Connective Ethnography. Ph.D., University of Amsterdam, Amsterdam.

Duff, W. M. & Cherry, J. M. (2000). Use of historical documents in a digital world: comparisons with original materials and microfiche. *Information Research*, 6(1).

Fry, J. (2006). Scholarly research and information practices: a domain analytic approach. *Information Processing and Management, 42*(1), 299-316.

Green, R. (2000). Locating sources in humanities scholarship: the efficacy of following bibliographic references. *The Library Quarterly, 70*(2), 201-229.

Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S., & King, C. J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines*. Report. Berkeley, CA: Center for Studies in Higher Education. Retrieved from http://www.escholarship.org/uc/item/15x7385g/.

Kling, R., McKim, G., & King, A. (2003). A Bit More to IT: Scholarly Communication Forums as Socio-Technical Interaction Networks. *Journal of the American Society for Information Science and Technology*, *54*(1), 46-67.

Lehmann, S., & Renfro, P. E. (1991). Humanists and electronic information services: Acceptance and resistance. *College & Research Libraries*, 52 (September).

Liu, A. (2002). The future literary: Literature and the culture of information. In K. Newman (Ed.), *Time and the Literary*. New York: Routledge.

Liu, A. (2004). The Laws of Cool: Knowledge Work and the Culture of Information. Chicago: University of Chicago Press.

Madsen, C. (2010). Communities, innovation, and critical mass: understanding the impact of digitization on scholarship in the humanities through the case of Tibetan and Himalayan studies. D.Phil., University of Oxford, Oxford. Retrieved from http://ora.ouls.ox.ac.uk/objects/uuid:928053ea-e8d9-44ff-9c9a-aaae1f6dc695

McCarty, W. (2005). Humanities computing. Hampshire: Palgrave Macmillan.

Merton, R. K. (1968). The Matthew Effect in Science. Science, 159(3810), 56-63.

Meyer, E. T. (2006). Socio-technical Interaction Networks: A discussion of the strengths, weaknesses and future of Kling's STIN model. In J. Berleur, M. I. Numinem & J. Impagliazzo (Eds.), *IFIP International Federation for Information Processing, Volume 223, Social Informatics: An Information Society for All? In Remembrance of Rob Kling* (pp. 37-48). Boston: Springer.

Meyer, E. T., Eccles, K., Thelwall, M., & Madsen, C. (2009). Final Report to JISC on the Usage and Impact Study of JISC-funded Phase 1 Digitisation Projects & the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR). Retrieved from http://microsites.oii.ox.ac.uk/tidsr/system/files/TIDSR FinalReport 20July2009.pdf

Overhage, C. F. & Harman, R. J. (1965). *Intrex: Report of a planning conference on information transfer experiments*. Cambridge, MA: The MIT Press.

Palmer, C. L. & Neumann, L. J. (2002). The information work of interdisciplinary humanities scholars: exploration and translation. *The Library Quarterly, 72*(1), 85-117.

Research Information Network & British Library. (2009). *Patterns of information use and exchange: Case studies of researchers in the life sciences*. Report. London: Research Information Network. Retrieved from http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie. Research Information Network & National Endowment for Science, Technology, and the Arts. (2009). *Open to All? Case Studies of Openness in Research*. Report. London: Research Information Network.

Stone, S. (1982). Humanities Scholars-Information needs and uses. *Journal of Documentation*, 38(4), 292-313.

Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. *Review of Educational Research*, 79, 625-649.

Talja, S., & Maula, H. (2003). Reasons for the use and non-use of electronic journals and databases. *Journal of Documentation*, 59(6), 673-691.

Thelwall, M. (2009). *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*. San Rafael, CA: Morgan & Claypool.

Tibbo, H. (1994). Indexing for the humanities. Journal of the American Society for Information Science, 45(8), 607-619.

Warwick, C., Terras, M., Huntington, P., & Pappa, N. (2008). If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data. *Literary and Linguist Computing*, 23(1), 85-102. doi: 10.1093/llc/fqm045

Warwick, C., Terras, M., Huntington, P., Pappa, N., & Galina, I. (2006). *The LAIRAH Project: Log analysis of digital resources in the arts and humanities final report to the Arts and Humanities Research Council*. Arts and Humanities Research Council. Retrieved 26 May 2008 from: http://www.ucl.ac.uk/slais/claire-warwick/publications/LAIRAHreport.pdf

Warwick, C., Terras, M., Galina, I., Huntington, P., & Pappa, N. (2007). Evaluating Digital Humanities Resources: The LAIRAH Project Checklist and the Internet Shakespeare Editions Project. In L. Chan & B. Martens (Eds.), *ELPUB2007*. *Openness in Digital Publishing: Awareness, Discovery and Access - Proceedings of the 11th International Conference on Electronic Publishing held in Vienna, Austria 13-15 June 2007*. Vienna: OEKK-Editions, 297-306.

Watson-Boone, R. (1994). The information needs and habits of humanities scholars. Reference Quarterly, 34(2), 203-216.

Wiberley, S. E. & Jones, W. G. (2000). Time and technology: A decade-long look at humanists' use of electronic information technology. *College and Research Libraries*, *61*(5), 421-431.

Wiberley, S. (1991). Habits of humanists: Scholarly behavior and new information technologies. *Library Hi Tech* 9(1), 17–21.

Wiberley, S. & Jones, W. G. (1994). Humanists revisited: A longitudinal look at the adoption of information technology. *College and Research Libraries* 55, 499–509.

Wiberley, S. & Jones, W. G. (2000). Time and technology: A decade-long look at humanists' use of electronic information technology. *College and Research Libraries* 61, 421–31.

Wyatt, S. (2003). Non-users also matter: The construction of users and non-users of the Internet. In N. Oudshoorn & T. Pinch (Eds.), *How Users Matter: The Co-construction of Users and Technology* (pp. 67-79). Cambridge, MA: The MIT Press.

Notes

¹A socio-technical interaction network (STIN) is a construct of people in relationships with each other and with their technologies. A key element of the STIN framework is that it doesn't assume that either the people or the technologies play a larger role in determining the nature of the relationships. For more on the STIN framework, see Meyer (2006).

²Retrieved 28 October, 2010 from http://www.oldbaileyonline.org/

³Retrieved 28 October, 2010 from http://www.oldbaileyonline.org/static/Publishinghistory.jsp

⁴Retrieved 1 December, 2010 from http://www.oldbaileyonline.org/static/Project.jsp

⁵http://www.londonlives.org/

⁶http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx

⁷http://www.bl.uk/reshelp/findhelprestype/news/newspdigproj/burney/index.html

8http://blpc.bl.uk/

⁹http://www.thetimes.co.uk

10http://lwlimages.library.yale.edu/walpoleweb/

11http://www.nationalarchives.gov.uk/

12http://library.wellcome.ac.uk/

¹³http://www.exlibrisgroup.com/category/MetaLibOverview

14http://scholar.google.com

15http://books.google.com/

¹⁶An online reference management tool available at http://www.zotero.org/

¹⁷A portal offering free tools for textual analysis, available at http://portal.tapor.ca/portal/portal

¹⁸The next generation textual analysis portal provided by the TAPoR team, available at http://hermeneuti.ca/voyeur

¹⁹http://www.oldbaileyonline.org/static/Legal-info.jsp

²⁰This relatively high level of blogging and online publishing indicates that our sample was somewhat skewed towards scholars engaged with digital approaches to scholarship and should not necessarily be taken to indicate a broader trend based on these limited data.

²¹DIAMM: http://www.diamm.ac.uk

²²OCR: Optical character recognition, an automated method of recognizing words and characters in a scanned document.

23http://www.diamm.ac.uk

²⁴http://twitter.com/

²⁵http://www.facebook.com

²⁶http://www.docs.google.com

²⁷https://www.yousendit.com/

²⁸http://www.skype.com

²⁹http://www.omifacsimiles.com/brochures/chantilly.html

³⁰http://en.wikipedia.org/wiki/Music_OCR

31http://www.rae.ac.uk/

³²As described in a later section, Corpus Linguistics is

the study of language in the form of texts, drawing upon collections of text known as *corpora* for linguistic analysis, as opposed to studying language in a spoken form.

33http://copac.ac.uk/

34http://www.bodleian.ox.ac.uk/bodley/eresources

35http://www.lib.cam.ac.uk/

36http://www.natcorp.ox.ac.uk/

37http://lion.chadwyck.co.uk/

³⁸An online journal database containing over a thousand titles, available to scholars primarily through university libraries' subscriptions, http://www.jstor.org/

³⁹An online database containing periodicals, dissertation abstracts, and other primary sources, available to scholars primarily by library subscription, http://www.proquest.co.uk

40http://newspapers.bl.uk/blcs/

⁴¹Microsoft Photo Viewer is pre-installed on Windows PCs. A description is available at http://en.wikipedia.org/wiki/Windows_Photo_Viewer

42http://www.titania.bham.ac.uk/

43http://www.tlg.uci.edu/

44http://www.claymath.org/library/

45http://www.gutenberg.org

46http://www.archive.org/

47 http://plato.stanford.edu/

48http://philpapers.org/

49http://zetoc.mimas.ac.uk/

50http://www.scottishcorpus.ac.uk/

51 http://www.lexically.net/wordsmith/

52http://www.sketchengine.co.uk/

53http://ucrel.lancs.ac.uk/wmatrix/

54http://www.lexisnexis.com/uk/legal/

55http://www.eprints.org

56http://www.gla.ac.uk/enroller

⁵⁷Note: Individual projects participating in this research collaboratory include: *Cultures of Knowledge: An Intellectual Geography of the Seventeenth-Century Republic of Letters* (University of Oxford); *Circulation of Knowledge* (national project in theNetherlands, that includes the Descartes Center, University of Utrecht, Huygens Institute, National Library, University of Amsterdam, DANS, and the Virtual Knowledge Studio); *Mapping the Republic of Letters* (Stanford University). For simplicity, in this report, the collaboration is referred to as *The Digital Republic of Letters*

⁵⁸United States: National Science Foundation, National Endowment for the Humanities, Social Science Research Council and Mellon Foundation; United Kingdom: Joint Information Systems Committee; Netherlands: Netherlands Organization for Scientific Research and Netherlands Common Language Resources and Technology Infrastructure

59http://plato.stanford.edu/entries/grotius/

60http://en.wikipedia.org/wiki/Christiaan_Huygens

61 http://plato.stanford.edu/entries/descartes-works/

62http://ella.slis.indiana.edu/~katy/ivlab/index.html

63 http://www.huygensinstituut.knaw.nl/

64http://gallica.bnf.fr/

⁶⁵http://www.utwente.nl/ub/help/glossary/glossary.doc/dutch_union_catalogue.html

⁶⁶http://www.ub.uu.se/en/Collections/Early-imprints/ Special-collections/Waller-Collection/

67http://www.digento.de/titel/101053.html

68http://artfl-project.uchicago.edu/

69http://nl.wikipedia.org/wiki/Henk_Nellen

⁷⁰https://republicofletters.stanford.edu/

⁷¹The panel was titled "Building the Humanities Lab: Scholarly Practices in Virtual Research Environments." A full description is available at http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-611.html

72https://republicofletters.stanford.edu/

73http://www.dropbox.com/

⁷⁴DOI is a digital object identifier (see http://www.doi.org/).

Full acknowledgements

In addition, other project members provided editorial comments and support during the research process. The full list of project members is as follows:

Oxford Internet Institute: Eric T. Meyer (PI), Monica Bulger, Grace de la Flor, Kathryn Eccles, Christine Madsen, Michael Thelwall, Ralph Schroeder, Adham Tamer, Arthur Bullard

University College London: Claire Warwick (Co-PI), Melissa Terras

eHumanities Group, Royal Netherlands Academy of Arts and Sciences: Sally Wyatt (Co-PI), Smiljana Antonijevic, Anne Beaulieu

Oxford e-Research Centre: Marina Jirotka (Co-PI), Annamaria Carusi

The project's advisory board was also key to helping develop the methods and offered important comments on the project's findings: David Robey (Chair), William Dutton, Dan Shapiro, Caroline Warman, and Paul Wouters.

We are also deeply grateful to all the participants in the project, who generously agreed to help us understand their information practices.

Hugh Adlington, English Department, University of Birmingham

Colin Allen, Department of History and Philosophy of Science, Indiana University Bloomington

Wendy Anderson, School of English & Scottish Language & Literature, University of Glasgow

Dawn Archer, School of Journalism, Media & Communication, University of Central Lancashire

Margaret Bent, Faculty of Music, University of Oxford

Erik-Jan Bos, Descartes Centre for the History and Philosophy of the Sciences and Humanities, University of Utrecht

James Brown, Faculty of History, University of Oxford

Henry Clarke, Philosophy Department, UCL

Nicole Coleman, Stanford Humanities Center, Stanford University

Amber Corfield-Moore, Philosophy Department, UCL

Julia Craig-McFeely, Faculty of Music, University of Oxford

Adam Crymble, History Department, King's College London

Michael Cuthbert, Music Program, Massachusetts Institute of Technology

Dan Edelstein, Department of French and Italian, Stanford University

Nadine Elzein, Philosophy Department, UCL

Amy Erickson, Department of Geography, University of Cambridge

Costas Gabrielatos, Department of Linguistics and English Language, Lancaster University

Susan Gane, Department of History, University of Birkbeck

Sebastian Gardner, Philosophy Department, UCL

Guido Gerritsen, Huygens Institute, Royal Dutch Academy of Sciences

Marcus Giaquinto, Philosophy Department, UCL

Drew Gray, Department of History, University of Northampton

Dave Gunning, English Department, University of Birmingham

Kat Gupta, School of English Studies, University of Nottingham

Benjamin Heller, Worcester College, University of Oxford

Tim Hitchcock, History Department, University of Hertfordshire

Howard Hotson, Faculty of History, University of Oxford

Neil Jefferies, Bodleian Libraries, University of Oxford

Elizabeth Eva Leach, Faculty of Music, University of Oxford

Fiona Leigh, Philosophy Department, UCL

John Levin, Centre for Computing in the Humanities, King's College London

Jonny McIntosh, Philosophy Department, UCL

Michaela Mahlberg, School of English Studies, University of Nottingham

Oliver Mason, English Department, University of Birmingham

Tony McEnery, Department of Linguistics and English Language, Lancaster University

Wijnand Mijnhardt, Descartes Centre for the History and Philosophy of the Sciences and Humanities, University of Utrecht

Jim Mussell, English Department, University of Birmingham

Henk Nellen, Huygens Institute, Royal Dutch Academy of Sciences

Christopher Peacocke, Philosophy Department, UCL
Sarah Richmond, Philosophy Department, UCL
Geoffrey M. Rockwell, Department of Philosophy, University of Alberta
Pat Rogers, Department of English, University of South Florida
Thomas Schmidt-Beste, School of Music, Bangor University
Alison Sealey, English Department, University of Birmingham
Susan Sommers, History Department, Saint Vincent College
Jason Stoessel, School of Arts, University of New England, New South Wales, Australia
Janice Turner, History Department, University of Hertfordshire
Charles van den Heuvel, Huygens Institute, Royal Dutch Academy of Sciences
Lee Walters, Philosophy Department, UCL
Bencie Woll, Deafness Cognition and Language Research Centre, UCL

The Research Information Network has been established by the higher education funding councils, the research councils, and the national libraries in the UK. We investigate how efficient and effective the information services provided for the UK research community are, how they are changing, and how they might be improved for the future. We help to ensure that researchers in the UK benefit from world-leading information services, so that they can sustain their position as among the most successful and productive researchers in the world.

The Research Information Network 96 Euston Road London NW1 2DB contact@rin.ac.uk

www.rin.ac.uk