How do Researchers in the Humanities Use Information Resources?

Ellen Collins

Research Information Network, UK,
ellen.collins@rin.ac.uk

Michael Jubb

Research Information Network, UK,
Michael.jubb@rin.ac.uk

Abstract

Information is at the heart of research. Every stage of the research cycle involves discovering, accessing, generating, manipulating, interpreting or presenting information, in order to advance knowledge. Researchers operate within a complex information environment, with needs that they themselves perhaps do not always fully understand, and are dependent upon services which are fast-changing and affected by much wider social trends. This paper examines how researchers in the humanities currently access, use and share information, paying particular attention to the influence of new technologies in changing information — and consequently wider research — practices. It finds that researchers are adopting new technologies where this helps them to work more effectively, but that traditional practices still dominate in some areas. There is mixed evidence of new research questions emerging as a result of engagement with new technologies.

Key Words: information resources; humanities; new technologies; research cycle

Background

Several factors influence how researchers engage with information throughout the research cycle. These might be grouped into two broad categories.

First, the supply factors — in other words, what information is made available to researchers, and how; and how are they able to reuse and disseminate that information? Second, the demand factors, and specifically the role of researchers' learned behaviours and accepted norms in determining how and where information should be found, analysed and shared. As we shall see, new technologies are particularly important influences upon both supply and demand factors, shaping the nature of information itself and the routes by which researchers access, analyse and disseminate it.

In terms of the peer-reviewed, published body of literature, the move to an online environment has been fairly well documented. Data for the United Kingdom show that higher education library expenditure on electronic-only serials has climbed from 9.6% of all serial expenditure in 2001 to 48.6% of such expenditure in 2009 (RIN, 2010). As of 2008, an average of 99.6% of all large publishers' titles were available online; even small publishers offered a very creditable 88% of their content in electronic format (Cox and Cox, 2010). UK researchers have seized upon this increase in provision, with article downloads more than doubling in the three academic years between 2003/04 and 2006/07 (Nicholas *et al.*, 2009). As a result of this increase in provision and usage, Nicholas *et al.* (2011) suggest that researchers' habitual preference to 'power browse', or seek the key messages in articles rather than read the whole thing in a linear fashion, is being supported. In addition, the way that researchers locate such information is changing — the same research showed a strong preference for 'gateway providers' such as Google, PubMed or Web of Science, as opposed to publisher or library websites (Nicholas *et al.*, 2011).

Research does not turn exclusively upon re-reading what others have written, of course, and the way that researchers handle data has also been affected by changing technologies. As Jim Gray recognised, technological advances have altered the way that researchers capture, curate, analyse and visualise data at every scale (Hey *et al.*, 2009). This is particularly noticeable in the sciences, with cyberinfrastructure or e-science programmes in several countries dedicated to seizing the possibilities offered by large-scale computing. However, it is also beginning to apply in other fields, with disciplines such as digital humanities establishing themselves within universities, although they have not necessarily received the same level of support for developing their infrastructures (Borgman, 2009).

When it comes to sharing their research findings, academics are again presented with a widening range of options thanks to changes in technology.

The democratising influence of the web has been felt in academia, as elsewhere, and there has been considerable interest in the role that social media can play in research communication. Researchers have indicated that such tools are particularly useful at the dissemination stage of research (Nicholas *et al.*, 2010), but habitual usage is generally relatively low (Nicholas *et al.*, 2010; Proctor *et al.*, 2010). However, almost half of researchers make their work in progress available to their own network, and 37% make their data available online within their private network (Proctor *et al.*, 2010). Publishing developments such as PLoS One, an open access online-only journal which publishes all articles judged to be technically sound, have also made it easier for researchers to share their findings, although the evidence suggests that publication in such journals is not yet esteemed as highly as more traditional resources (Proctor *et al.*, 2010).

This illustrates the importance of the second set of factors which influence researcher behaviour — researchers' own habits. Although these habits are personal, there are discernable trends. Age and seniority of researchers can affect decisions to engage with information resources; for example, older, more senior researchers are more likely to use web 2.0 technologies (Proctor *et al.*, 2010). Like other studies, this confounds the notion that the 'Google generation' of young researchers are somehow more at home with new technologies (Carpenter *et al.*, 2010). Similar distinctions have been found in browsing habits, with more junior researchers less likely to browse. The authors speculate that this may be because they have less time to devote to maintaining an awareness of current research. They also highlight the importance of networks of collaborators and colleagues in raising awareness of significant contributions to the literature, and stress that these are built up over the course of a career, once again putting more established researchers in a different positions to those at the start of their working lives (Bennett and Buhler, 2010).

Discipline is widely recognised as another crucial influence on information behaviour (Tenopir *et al.*, 2004; Nicholas *et al.*, 2009). Even within a discipline, there can be identifiable sub-groupings: Tenopir *et al.* (2005) found that US medical faculty with MD degrees are less likely to read content online than medical faculty with PhDs. Nicholas *et al.* (2009) found that researchers in the same subject but different institutions use different strategies, facilities and functions when reading electronic journals. And at an even more local level, the habits of colleagues within a research group or collaboration can have an

important effect upon information behaviours at every stage of the research cycle (RIN and NESTA, 2010).

Few researchers have attempted the herculean task of understanding all academics' information behaviours at every stage of the research cycle. A range of studies have looked at use of specific information tools or services — for example, web 2.0 (Nicholas *et al.*, 2010; Proctor *et al.*, 2010) or e-journals (Nicholas *et al.*, 2010). Others have adopted a narrower focus, looking at how researchers within a discipline or institution use certain tools or services (Tenopir *et al.*, 2005; Nicholas *et al.*, 2009). But we still lack a detailed understanding of how researchers within a specific discipline create, use and manage information as an integral part of an entire research project, and of how such behaviours are changing.

To address this gap, the Research Information Network commissioned a series of in-depth case studies, looking at researcher behaviour in different disciplines. The first set, on researchers in the life sciences, was published in 2009 (Williams *et al.*, 2009), and the third, on the physical sciences, will be published later this year. This paper presents the results from the second study, looking at researchers in the humanities (Bulger *et al.*, 2011).

The traditional notion of the humanities researcher as a lone scholar in a dusty archive has long been redundant. Recent work found that 65% of humanities researchers had collaborated beyond their own department in the previous five years (Meyer *et al.*, 2009). Other studies suggest that researchers are engaging with new technologies in the humanities, and are aware of how this may shape their research, in terms of both process and outputs (Warwick *et al.*, 2008). This study attempts to understand how these developments in practice are changing the way researchers handle information within their research, in six different cases. It also looks at how the changing nature of information handling, and in particular the technological developments of recent years, has changed the nature of research itself.

Method

The cases were selected to offer a diversity of practice, including traditional methods alongside more advanced digital resource use. To ensure that the full range of humanities practices were represented, the researchers recruited

participants through different kinds of cases, looking at users of a specific resource or database and researchers within a traditional department, within a field and within a collaborative project. The cases were as follows:

- Resource: Old Bailey Online, a database containing the criminal proceedings of London's central criminal court from 1674–1913.
- Resource: Digital Archive of Medieval Music (DIAMM), a database containing medieval polyphonic music manuscripts dating from around 800–1550.
- Department: University of Birmingham English Department, which has 45 members of academic staff and combines the study of language and literature.
- Department: UCL Philosophy Department, ranked first in the UK Research Assessment Exercise (RAE) in 2008 and with staff covering a range of research areas including philosophy of mind, history of mathematics, ethics and metaphysics.
- Field: Corpus Linguistics, a 'born-digital' field where researchers study a 'corpus' of texts from a variety of sources including newspapers, literature, spontaneous talk or broadcast.
- Collaborative Project: The Digital Republic of Letters, an international humanities collaboration developing an online research collaboratory of seventeenth-century correspondence.

Although the research team conducted a series of interviews and focus groups within each case to gather a range of experiences, the overall findings cannot be taken as representative of the humanities as a whole. There is probably a disproportionate representation of technologically-enabled researchers, as at least three of the cases involve some kind of digital resource. But, taken together, the cases show the diversity of practice and illustrate how researchers are engaging with new technologies, so that support services including funders, publishers and librarians can better meet their needs.

Findings

Search and retrieval strategies differed considerably between fields, depending upon the kinds of resources that are available. Historians using Old Bailey Online, for example, tended to use the database as their first port of call when seeking information, before turning to Google or other more generic services.

Similarly, philosophers took advantage of the well-established Phil Papers pre-prints archive to find the most up-to-date publications. But there is still a strong awareness of resources provided by the library, which was specifically mentioned by users of Old Bailey Online, and by researchers in the philosophy and English departments. Members of the Digital Republic of Letters project mentioned the importance of direct contact with their colleagues in identifying relevant literature and data.

Researchers' approach to managing the information that they find is, in their own words, 'haphazard'. Researchers using Old Bailey Online created mini-databases of useful content for each of their projects, either on their hard drives or using online tools. Participants in the Digital Republic of Letters project cut and pasted important sections of their primary texts into Word documents, to be reordered and dropped into publications at a later stage. Many participants wanted better organisational tools linked to the data sources themselves, and where these were available they were well-used — DIAMM, for example, offers annotation and bookmarking facilities, which were popular with the musicologists.

As the literature suggests, collaboration is relatively common within the humanities, although not necessarily formally organised. Again, there are noticeable differences between subjects, with philosophers saying that collaborating to collect and analyse data was not encouraged in their field. Other disciplines were more enthusiastic, using the web to share their data reasonably openly among a community of interested peers. DIAMM was initially established as a repository for images of delicate or rare medieval manuscripts, and is therefore equipped with a facility for researchers to upload their own transcriptions. Many do so. In other fields, the infrastructure is not provided and researchers must find their own way to share. For example, corpus linguists use personal or discipline-specific websites to share their corpora and computer code that they have developed for analysis.

Perhaps because of the informal nature of collaborations, the way that humanities researchers communicate with each other also tends to be relatively informal. Most researchers, when asked about project communications, mentioned email, telephone and face-to-face meetings as their main channels. Even the most complex collaboration within the study — the Digital Republic of Letters — relied primarily upon these methods, with researchers flying internationally to share their expertise where necessary. Some researchers

had begun to engage with web 2.0 tools for sharing work in progress, including Dropbox (Old Bailey Online), Skype (DIAMM) or wikis and Google Docs (English). But on the whole, traditional methods prevailed.

As the literature suggests, social media are viewed in a more positive light as a way to communicate research findings. There was some evidence of researchers blogging and tweeting, particularly younger researchers in philosophy and corpus linguistics. Younger philosophers mentioned that they used blogs to stay up to date with the newest contributions in their field. Philosophers also post papers on university websites prior to publication and receiving feedback, both solicited and unsolicited. However, this was not a new practice made possible by the web, but rather a continuation of a long tradition of pre-publication discussion, which previously happened at conferences, on email lists or face-to-face. This is just one example of researchers adopting new technologies where they enhance existing practices.

This pragmatic approach meant that in areas where new technologies do not meet every existing need there was a continuing demand for print; this was true of all the cases. Sometimes this resulted in an interesting hybrid approach. The philosophers, for example, used electronic versions of key texts to search for important words or passages, but used the print editions of the same texts to cross-reference page numbers. Academics in the English and philosophy departments made extensive use of Google Books to skim chapters or sections and decide whether it was worth acquiring the full text, either by buying the book or obtaining it via the library.

In other instances, the continuing demand for print seems to be based more on habit than objective judgements about usefulness. Most DIAMM users mentioned 'a notion in the field that you can always get more out of seeing the original than seeing a digital image of it', and a consequent pressure to use hard-copy manuscripts in archives, even when they themselves saw the benefits of using electronic versions. Similarly, some DIAMM users and researchers in the philosophy department felt that it was more prestigious to publish in print journals than in electronic-only ones, although they did not necessarily agree that this ought to be the case. This may be linked to the habit, which researchers in English, philosophy and the Digital Republic of Letters project admitted, of citing print versions of sources, even when the majority of the research has been conducted using an electronic resource.

One important question for the RIN project, and one which has been asked repeatedly in relation to e-science, is whether the new tools and technologies had fundamentally changed the nature of research. The answer, as with more scientific subjects, is complex. Researchers in most of the cases agreed that electronic resources *enable* new kinds of research, but do not necessarily *inspire* it. Old Bailey Online, for example, allowed researchers easily to search for every occurrence of an individual's name. But most interviewees suggested that this is something they had always wanted to do but which had previously been impossible, not least because nobody would fund the necessary months of work in the archives. This was a theme in the responses of other participants, particularly the corpus linguists (who also deal with occurrences of words or small phrases within large bodies of text). Incidentally, the technology sometimes brought its own problems: Old Bailey Online users sometimes had to search for a dozen different spellings or versions of a name to find every appearance of an individual within the records, since a computer cannot recognise the relationship between, for example, 'Dunckerley' and 'Tunikerly' in the same way as a human.

Researchers also suggested that electronic resources helped to put their research questions in a wider context. For users of DIAMM, this benefit was experienced on two fronts. First, the high-quality images allow more detailed examination of delicate manuscripts, permitting researchers to look very closely at the tiny, tell-tale habits which distinguish the anonymous scribes from each other. Second, once these connections have been made, the DIAMM database can record them, making links between different records which allows subsequent users to place their own scholarship in its place within the discipline. DIAMM users, like Old Bailey Online users, felt that linked digital archives would further enhance this facility.

Researchers attached to the Digital Republic of Letters project were the only ones within the study to suggest that their research questions had changed fundamentally as a result of the new technologies. The connections that they have built between various archives and the visualisation techniques that they have employed to 'see' these connections, have inspired new questions about relationships and knowledge exchange in the seventeenth century. It is perhaps relevant that this project involved the greatest degree of interaction and parity between technologists and humanists; both worked together to build the technology and therefore were able to create something that would really advance both disciplines. In this way, they may have

avoided the limitations outlined by one of the developers working on Old Bailey Online, who felt that researchers were limiting themselves to doing research that the tool supported, rather than changing the tools to ask new kinds of question.

Conclusions

The humanists in this study were, overall, pragmatic users of new technologies. When they felt that a new tool would enhance their research practices, they would experiment with it, and often subsequently adopt it. A note of caution ought to be sounded about the enthusiasm for data sourcing and management tools such as Old Bailey Online and DIAMM, since participants were recruited due to their use of these services and thus constitute a purposely biased sample. The same is true, to a lesser extent, of the corpus linguists, whose discipline is effectively 'born digital' — only possible through technology. Nonetheless, the use of these tools is echoed by the philosophers' use of websites to share their unpublished papers. In all cases, existing research practices are made easier, quicker or more effective thanks to new technologies.

However, it was clear that researchers made clear assessments of the value of new technology and, where they did not consider it to be useful, would not engage with it for its own sake. This was particularly apparent when it came to social media such as Twitter and blogs. While researchers recognised the value of engaging with a wider audience through such means, they also felt that it could become a distraction from the main business of research. In other instances, new technologies were not seen as sufficiently good replacements for older practices. For example, despite some awareness of tools such as Google Docs and Dropbox, most researchers continued to use email lists to share and edit work in progress with their collaborators.

In other instances, rejection of new technologies was less easily explained, and there is clearly some residual habitual attachment to print. The experiences of DIAMM users illustrates this: they were under pressure from their colleagues to continue using original manuscripts even though digital versions provide new research opportunities. Some researchers' desire to publish in print as opposed to online-only journals may also reflect this problem, although it is also true that print journals tend to have higher impact factors, so it may be a rational decision on the part of researchers.

Evidence about the impact of these changing technologies on research questions and the nature of humanities research, is mixed. Researchers on the Digital Republic of Letters project were certain that the technologies they developed had changed the nature of the research questions they asked. In other cases, notably Old Bailey Online, the developers expressed some frustration at researchers' passive use of the tools; they saw potential for changing the nature of research questions, but felt that the researchers themselves were only interested in doing the same type of work, but more efficiently.

These findings should be of interest to librarians on a number of fronts. Researchers in most cases were aware of the library and the services it offers, and named it as one of the resources they would turn to in seeking information. Most tended to use Google as a first point of call, but later moved onto more specialised resources, such as those provided by the library. Linking the library catalogue to Google searches may help increase the visibility, and thus use, of resources.

The hybrid print-electronic use observed in several of the cases is also interesting. The use of Google Books to skim publications for relevant content before making decisions about purchasing or requesting books via the library is potentially important. Encouraging this practice could help ensure that new acquisitions are actually needed, and not requested by researchers on a speculative basis. The ongoing interest in both print and electronic versions of content, and the tendency to cite the former even where the latter has been the primary source of information, suggest that a push towards e-books will need to be accompanied by considerable advocacy and support for researchers.

Finally, several researchers mentioned their concern that digitised versions of source material are often partial and unsustainable. Many are created through project funding, leaving them vulnerable in the longer term. It is rare for an entire collection to be digitised (due to financial constraints), and equally rare for data to be linked across archives or platforms. The report on which this paper is based recommended that library and information services 'could help by providing a structure that maintains digital tools and archives, and promotes development of innovative resources' (Bulger *et al.*, 2011, p. 77). In the current straitened economic climate of most university libraries, this may seem like a low and somewhat risky priority. However, the enthusiasm with which some of these tools are being used suggests that they could be a valuable asset for library services in the longer term.

Acknowledgements

References

Bennett, D. and A. Buhler (2010): 'Browsing of E-Journals by Engineering Faculty'. *Issues in Science and Technology Librarianship*, Spring 2010. [Accessed 21 June 2011 from http://www.istl.org/10-spring/refereed2.html]

Borgman, C. (2009): 'The Digital Future is Now: A Call to Action for the Humanities'. *Digital Humanities Quarterly*, 3(4) [Accessed 21 June 2011 from http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html]

Bulger, M., E. Meyer, G. de la Flor, M. Terras, S. Wyatt, M. Jirotka, K. Eccles and C. Madsen (2011): '*Reinventing research? Information practices in the humanities*'. Research Information Network.

Carpenter, J., L. Wetheridge, N. Smith, M. Goodman and O. Strujivé (2010): *Researchers of Tomorrow; A three year (BL/JISC) study tracking the research behaviour of 'Generation Y' doctoral students. Annual Report*. Education for Change. [Accessed 21 June 2011 from http://explorationforchange.net/attachments/056_RoT%20Year%201%20report%20final%20100622.pdf]

Cox, J. and L. Cox (2010): *E-only scholarly journals: overcoming the barriers*. Research Information Network.

Hey, T., S. Tansley and K. Tolle (2009): 'Jim Gray on eScience: A Transformed Scientific Method'. In T. Hey, S. Tansley and K. Tolle (eds) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. [Accessed 21 June 2011 from http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf]

Meyer, E., K. Eccles, M. Thelwall and C. Madsen (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 Resources (TIDSR)*. [Accessed 21 June 2011 from http://microsites.oii.ox.ac.uk/tidsr/system/files/TIDSR_FinalReport_20July2009.pdf]

Nicholas, D., I. Rowlands, P. Huntington, D. Clark and H. Jamali (2009): *E-journals: their use, value and impact*. Research Information Network.

Nicholas, D., I. Rowlands and D. Wamae (2010): 'Charleston Observatory 2010: Are social media impacting on research?'. At *XXX Annual Charleston Conference*, Thursday

4 November 2010 [Accessed 21 June 2011 from http://www.ucl.ac.uk/infostudies/research/ciber/Charleston-2010.pdf].

Nicholas, D., I. Rowlands, P. Williams, D. Brown and D. Clark (2011): *E-journals: their use value and impact final report*. Research Information Network.

Proctor, R., R. Williams, J. Stewart, M. Poschen, A. Voss, H. Snee, M. Asgari-Traghi, M. Chen, R. Combs, R. Elvey, B. Harries, A. Pechurina and H. Brown (2010): *If you build it, will they come? How researchers perceive and use web 2.0*. Research Information Network.

Research Information Network (2010): *Trends in the finances of UK higher education libraries, 1999–2009.*

Research Information Network and NESTA (2010): *Open to All? Case studies of openness in research.*

Tenopir, C., D. King and A. Bush (2004): 'Medical Faculty's Use of Print and Electronic Journals: Changes Over Time and Comparison with Other Scientists'. *Journal of the Medical Library Association*, 92(2), 233–241. [Accessed 21 June 2011 from http://scholar.cci.utk.edu/carol-tenopir/files/tenopir_jmla_article_042203.pdf].

Tenopir, C., D. King, P. Boyce, M. Grayson and K. Paulson (2005): 'Relying on Electronic Journals: Reading Patterns of Astronomers'. *Journal of the American Society for Information Science and Technology,* 56(8), 786–802. [Accessed 21 June 2011 from http://scholar.cci.utk.edu/carol-tenopir/files/tenopir_jasist_article_042503_preprint.pdf].

Warwick, C., M. Terras, P. Huntington and N. Pappa (2008): 'If You Build It Will They Come? The LARIAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data'. *Literary and Linguistic Computing*, 23(1), 85–102.

Williams, R., G. Pryor, A. Bruce, S. Macdonald, W. Marsden, J. Calvert, M. Dozier and C. Neilson (2009): *Patterns of information use and exchange: case studies of researchers in the life sciences*. Research Information Network.