Ten highlights from The Impacts of Digital Collections: Early English Books Online and House of Commons Parliamentary Papers by Eric T. Meyer and Kathryn Eccles

Introduction

In 2015, in cooperation with ProQuest, Jisc commissioned this study of the Impacts of Digital Collections focused on two particular collections: Early English Books Online (EEBO) and House of Commons Parliamentary Papers (HCPP). These two collections are just a fraction of the number of collections that Jisc has purchased on behalf of its member institutions. While an understanding of these two collections is not necessarily generalizable to all digital collections (or even all Jisc-provided collections), they were selected because they are both relatively mature in the sense of having been available to users for over a decade, were thought to be well known in the research community, and also appeal to users from multiple disciplines.

Our team has undertaken related studies of approximately 20 different digital collections over the last decade, and EEBO and HCPP compare well both quantitatively and qualitatively to other digital collections. Taking into account the fact that the earlier studies only reflect a portion of the time covered by the current study, EEBO and HCPP appear to be in the upper third of resources we have looked at in terms of usage and impact. They seem to fall into the same general category as resources like British History Online and Old Bailey Proceedings Online in their overall visibility and measurable academic impacts. These impacts go beyond simple numbers: we have shown clearly in our previous studies that smaller niche resources like the Digital Image Archive of Medieval Music (DIAMM) or Histpop can also demonstrate their importance, and that impacts must be understood not just to be related to size but also must consider their influence within specific areas

of research and teaching. HCPP and (particularly) EEBO can also be shown to be playing an influential role within certain disciplines.

EEBO houses over 125,000 digital versions of some of the earliest printed material in English, including books, play scripts, sermons, public and legal documents, religious material and some of the earliest gems of English Literature. It is accessible via subscription to the Jisc Historical Texts platform or via subscription on the ProQuest platform. The related EEBO-TCP (Text Creation Partnership) made 25,000 of the texts (but not images) of these items freely available in 2015, and 28,000 more are available to EEBO-TCP partners and subscribers. EEBO is available to institutions through Jisc as part of the Jisc eCollections as well as the from ProQuest.

The House of Commons Parliamentary Papers is a digital resource which holds the sessional papers covering the 18th, 19th and 20th Century, as well as documents dating even further back – from the mid to late 17th century, detailing parliamentary activities of the time. HCPP is available to institutions through Jisc as part of the ProQuest Archives 2014-2017 or via subscription from ProQuest.

This short summary is intended to highlight **ten key findings** from the study. The full report of this study is also available:

Meyer, E.T. & Eccles, K. (2016). The Impacts of Digital Collections: Early English Books Online & House of Commons Parliamentary Papers. London: Jisc. Available online: **microsites.oii.ox.ac.uk/tidsr/case-study/2016-idc**
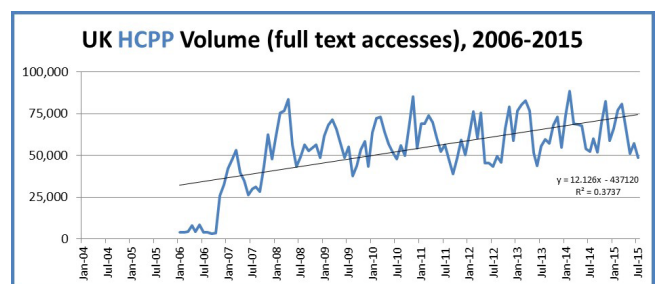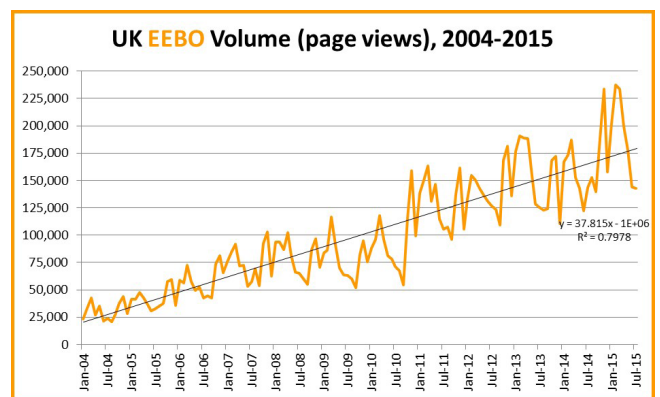
Key findings

One: The context of the use of digital resources is changing, but these changes are incremental and have a long development cycle prior to the realisation of impact.

The most important point to take from this study is that the impact of digital collections is not a 'big bang' moment that immediately and fundamentally changes everything about humanities research. One of the reasons that digital resources have been so successful in the humanities is that the digital sources that have emerged over the last decades immediately felt familiar to the researchers who have grown to rely on them. For scholars trained to work with primary documents open on a library table, seeing high quality scans of the same documents on a screen in one's own office is certainly different in many ways, but it is also reassuringly familiar: the look (if not the feel) of the page is still there, and new possibilities for looking at primary sources originally located thousands of miles apart on a single screen without needing to travel is compelling. The digitisation of materials such as EEBO unlocks astonishing special collections, allowing the scholar to roam in and out of the archive at will, and giving them the freedom to examine any text they like and to explore computationally or serendipitously. Had the first digital scans produced nothing but x-ray images or databases of spectrometer readings for statistical analysis, it is safe to speculate that they would not have seen the widespread adoption reported in this study and elsewhere. However, even though these digital resources are clearly important to scholars, they still take time to become embedded in various academic practices, and this is still an ongoing process.

Two: The usage of both Early English Books Online and House of Commons Parliamentary Papers has been increasing steadily over the past decade.

Usage data from 2004 (for EEBO) and 2006 (for HCPP) from ProQuest and Jisc show an upward trend. EEBO usage has been increasing steadily at a relatively linear pace. Because the increase is linear, the time for volume to double takes somewhat longer over the time period: whereas usage doubled from 50k page views per month to 100k per month in about three years (2005-2008), it stayed at approximately that level until 2011, when usage began to increase again and the next doubling to 200k page views per month took a further four years, underscoring the fact that digital collections take time to become embedded in research.

UK EEBO Volume (page views), 2004-2015

$y = 37.815x - 1E+06$
$R^2 = 0.7978$

UK HCPP Volume (full text accesses), 2006-2015

$y = 12.126x - 437120$
$R^2 = 0.3737$

Ten highlights from The Impacts of Digital Collections: Early English Books Online
& House of Commons Parliamentary Papers by Eric T. Meyer and Kathryn Eccles

3

HCPP usage also increases in a linear, but less marked fashion. Monthly full-text accesses increased rapidly to 50k by 2007, but have fluctuated between approximately 50-75k full-text accesses per month ever since.

Neither one of these two is 'better' than the other, but nicely demonstrate different patterns of use for digital collections. The House of Commons Parliamentary Papers collection appears to have found its audience relatively quickly, and then saw its usage remain relatively stable (taking into account the monthly fluctuations that are tied to the academic calendar).

Early English Books Online, on the other hand, appears to still be finding (and is being found by) new audiences. While the growth may be slowing a little in recent years, it appears that there is still more room for growth before it stabilises.

Three: While researchers at top universities are most likely to use EEBO and HCPP, less research-intensive HE institutions also benefit from both collections.

During the study, we found a cluster of courses on literature and book history being offered at a non-Russell Group, non-research-intensive university. The kind of courses offered through access to EEBO would simply not have been possible based on the institutions' limited special collections. The EEBO-based courses offered students the opportunity to come into contact with rare books and to think about the materiality of these books, leading to interest in and awareness of book history research, the potential to work with special collections, book conservation and literary heritage.

In one of the cases in the study where we examined a particular university, the embedding of HCPP within the Open University's courses resulted in heavier than expected usage statistics. Interviews with members of library and academic staff at the OU provided a rich

context to these statistics. Courses are designed around the digital content, and this content is particularly suitable for the widely dispersed OU student body. As many of the OU students are part-time, and more assessment-driven, the style of engagement with HCPP is carefully tailored to deliver not only access to the information contained within the collection, but essential digital literacy skills to understand and extract that information. The HCPP collection is integral to the OU's offering in History and (to a lesser extent) Law.

For both resources, however, researchers at top ranking institutions use them more heavily than the rest of the sector. This was particularly notable for EEBO: faculty at institutions in the top third of the REF (Research Excellence Framework 2014) rankings comprise 58% of all faculty in the English REF, but those same institutions account for 80% of the usage of EEBO. For HCPP, the difference between size of faculty and resource usage is less marked: while faculty at institutions in the top third of the History REF rankings include 61% of all History REF faculty, HCCP usage at those same institutions accounts for 59% of all HCPP use at REF universities. In other words, uptake of these resources is strongest in universities that performed well in the REF.

Four: Researchers rely heavily on specific digital collections that they return to regularly, which is resulting in incremental changes in scholarly behaviour.

In the survey of researchers done for this study, 97% of respondents (n=238) report that specific databases or collections are an important method for discovering primary sources. In fact, this was the option most cited as important, and also was ranked most highly of all the other important methods (including search engines such as Google) for discovering new primary sources. Many of the participants interviewed qualitatively for this study recalled the appearance of EEBO and HCPP on the

research scene, were early adopters of these resources, and recalled the transformative impact of these digital collections on their research.

This is particularly evident with EEBO, which inspires passionate responses from its users about both its strengths and its weaknesses. One EEBO user, for instance, commented "I just use EEBO all of the time... It's fantastic, it's marvelous!", while a question in our survey asking for potential improvements to EEBO yielded 87 often very detailed replies (compared to seven replies about improvements to HCPP).

One difference between EEBO and HCPP in this respect is that HCPP appears to be used differently as a central resource. Our survey, usage data and interviews all suggested that HCPP is widely used as a secondary research and teaching tool, but these types of usage look quite different to the usage of EEBO as a primary research collection.

One type of behaviour that participants in previous studies by our team worried about losing was the opportunity for serendipity, but over time the part digital resources play in allowing serendipitous research has become clearer. According to one person interviewed in this study, "EEBO is fantastic in terms of serendipity... in a way that special collections aren't." The ability to dive in, to experiment with keyword searching and not know exactly what that search will retrieve, is seen to be a powerful experimental tool both for researchers and student: "There's things I find on EEBO and I go, 'Oh that's new or I've never seen that before' and then 20 minutes later I'm following something else and seeing what I can find."

Five: Resource use in the humanities is extremely diverse, and this makes providing access to needed resources and tools particularly challenging.

In our survey of researchers, while there were a few collections with very high regular use (such as JSTOR used by 93% of respondents, or Google Books used by 91% of respondents), there is also incredible diversity of resource usage. Given the opportunity to list additional important resources, respondents listed a total of 136 different digital collections.

One of the differences between the humanities (and also the social sciences) and some other areas of research, especially certain scientific disciplines, is that there is a huge diversity not only of research topics, but also of the resources and tools needed to work on those topics. In previous research, we showed that some fields like nuclear physics rely on a very well-defined set of journals and have little need to search widely beyond those sources. One example is the arXiv.org pre-print server: many physics and mathematics researchers visit arXiv.org daily, and know that all important new papers will appear there first. In the humanities, there is no similar one-stop location where the most important materials will first appear. This makes providing infrastructure that can support this diversity more of a challenge.

Six: The citation evidence that is available shows a growing literature that mentions using EEBO or HCPP, and these publications in turn are reasonably well-cited.

Using techniques that produce a conservative estimate of the publication impact of EEBO and HCPP, we largely see a growth in publications overtly referencing these two resources over the last 15 years. This is true of books, articles, and dissertations and theses.

Ten highlights from The Impacts of Digital Collections: Early English Books Online
& House of Commons Parliamentary Papers by Eric T. Meyer and Kathryn Eccles

5

Recent data shows an apparent decline in citations, and while this could be because fewer papers are being written based on EEBO/HCPP, this is not consistent with the other data in this report. It is more likely that publications are relying on these digital collections, but that this use is undetectable using search methods. As we have shown in the survey data in this study, only a minority of authors who used EEBO or HCPP included any indication of their use of these collections in their citations. This is an issue we have noted previously (Meyer, 2011; Meyer, et al., 2009; Siefring & Meyer, 2013) and it is clear there is still room for additional training as well as system design that can nudge scholars in ways that will increase their likelihood of citing digital resources in transparent ways, such as additional support for automatic citation and human-readable URIs (Meyer, 2011, pp. 41, 56).

Also, while citations to these works drop off in recent years, this is to be expected when examining research in the humanities, as humanities disciplines tend to have a longer time-scale before publications reach their peak citations, often taking six - eight years before receiving half of the citations they will eventually receive.

Seven: The number and range of disciplines that refer to EEBO and HCPP is much more diverse than expected.

Several pieces of evidence point to the wide diversity of disciplines that are using EEBO and HCPP. EEBO has been mentioned in 773 different publication outlets, 564 of which only had one article mentioning EEBO. The greatest numbers of articles are in journals related to English literature, philology, libraries, and history, but the work is spread out rather than concentrated. HCCP was mentioned in 508 different publication outlets, 435 of which had only a single article mentioning HCPP. For HCPP, journals focusing on various historical and legal specialisations represent the most common sources, but many other areas including geography, sociology, and area studies are represented. This is also reflected in the subject classifications for dissertations and thesis that reference EEBO and HCPP which include (for EEBO) literature, history, theatre, music, religion, and political science and (for HCPP) history, law, economics, anthropology, and women's studies.

Eight: Researchers are more concerned with the content and functionality of the digital collections than in who provides the access.

A theme running throughout our research is that while researchers are passionately interested in gaining access to the content they require for their studies, they are less concerned with how that access is gained as long as it works well. Certainly users will have (often strong) opinions about the functionality of specific interfaces over others, but they have little awareness or interest in whether that access was provided by their local library, by a national broker such as Jisc, or through some other mechanism.

At some level, libraries and organizations like Jisc have become victims of their own success over the last 20 years at providing access to materials seamlessly for researchers at their institutions. Whereas users once struggled to gain access to electronic collections and needed to remember multiple passwords and subscription details, most users today who are working on their campus network (or remotely with locally managed passwords such as those provided by Eduroam or Shibboleth) find that they can access many primary collections as well as journals with a single click and no further authentication. As a result, they often don't even realize how this access was provided, and don't necessarily view themselves as 'library users' while accessing the digital collections. Often, it is only when traveling from their home institution or taking a job at a different institution that people become aware that the access they expect is no longer available, and thus become aware that access differs across institutions.

Nine: The UK is unusual for providing national-level access across institutions through Jisc's national purchasing.

Studies too numerous to cite have shown that the UK punches above its weight in the academic world. By almost any measure of research and teaching, the academic influence of the UK globally is disproportionate to our relatively small size. The reasons for this are numerous, but one of the baseline expectations at any leading knowledge institution is that their scholars will be able to access the research materials and publications they need to advance their scholarship.

It is impossible to show any causal link between Jisc's policies of providing national-level access across institutions via the policy of making national purchases of key digital collections and the prominence of the UK in the academic world, but it is clear from this and previous studies that scholars rely on these digital collections. Since individual institutions do not need to individually negotiate access, this would appear to have some democratizing effects, as researchers and students clearly benefit from access to resources that might otherwise not be provided by their local institutions.

Ten: Shifts to humanities data science and data-driven research are of growing interest to scholars, although there is still plenty of room for growth in this focus on digital humanities, particularly in teaching.

One new area of growth that was highlighted, particularly in the interviews, is in the area of digital humanities. There are increasingly active communities of scholars doing quite different types of digital scholarship with these resources, showing the potential that can be unlocked by accessible forms of data, such as the data created by the EBBO-TCP partnership. EEBO-TCP has created standardized, accurate XML/SGML encoded electronic text editions of early print books, and in 2015, released 25,000 texts into the public domain allowing users to copy, post, publish,

distribute, and otherwise share the data. EEBO-TCP has since been used in 'hackathons' to encourage creative re-use of the data, and to invite ideas for creating tools to access the collection for research or teaching, while HCPP was used in the large collaborative project Trading Consequences. In this case, the XML was exported directly from ProQuest to enable researchers in informatics to work with historians to identify commodities and places in huge datasets. We would suggest that similar trends could emerge in relation to other digital resources if they make themselves more open to data sharing and creative reuse.

Currently, these computational approaches are relatively few, but the availability of large digital collections opens up the possibility of large scale analysis to become a more important part of the overall humanities scholarship landscape.

Ten highlights from The Impacts of Digital Collections: Early English Books Online
& House of Commons Parliamentary Papers by Eric T. Meyer and Kathryn Eccles

7

Conclusion: Digital collections have become fundamental to modern scholarship.

Even though digital collections have the advantage of demanding relatively little of researchers in terms of fundamentally new scholarly behaviour (for even when such new uses are possible, they are not required), the growth in usage of these collections takes time, and the measures of impact such as citations traceable to the digital versions of materials take even longer because of slowly changing citation practices and the relatively long time between publication of new outputs and citations to those publications. Looking at the broader picture of digital collections more generally, it is clear that patience is necessary.

EEBO and HCPP have been shown to be relatively mature digital collections, both of which are demonstrably embedded in the day to day practices of researchers and scholars across a range of disciplines. This does not rule out the possibility of either collection finding new audiences and having new uses emerge. However, were these digital collections to be lost, it is safe to say that their loss would be noticed, and that such a loss would have a profoundly negative impact on knowledge creation in the UK.

The consistent story across a whole series of studies done by our team and others over the last decade is that digital collections have become a fundamentally important part of modern scholarship that would be immeasurably set back if the infrastructure to allow researchers continued access and to support new ways of using the information embedded in primary sources were allowed to decay.

Resources

EEBO:
Jisc eCollections:
http://bit.ly/Jisc_ecollections_1824

http://bit.ly/Jisc_ecollections_1819

Early English Books Online – Text Creation Partnership
http://bit.ly/Jisc_collection_1709

HCPP:
ProQuest Archives
http://bit.ly/Jisc_collections_2063