

# The Impacts of Digital Collections: Early English Books Online and House of Commons Parliamentary Papers

## Highlights from a study by Oxford Internet Institute

EEBO houses over 132,000 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.

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.

In 2015, Jisc, the UK higher education, further education and skills sectors' not-for-profit organisation for digital services and solutions' and ProQuest, a company well-known for preserving and enabling access to rich and varied information, commissioned a study that focused on the impacts to research and teaching of two collections – Early English Books Online (EEBO) and House of Commons Parliamentary Papers (HCPP). EEBO and HCPP have been available in the UK for over a decade, and are well-known in the research community, and appeal to users from multiple disciplines.

The research uses the principles of the TIDSR (Toolkit for the Impact of Digitised Scholarly Resources) toolkit (<http://bit.ly/TIDSR>) created by Oxford Internet Institute to better understand the quantitative and qualitative empirical evidence for the impacts that digital collections can be shown to have. Analysis was conducted into the long term usage trends of these collections, the bibliometric data from Scopus, Google Scholar and other sources of citations, as well as conducting interviews and surveys with academics, librarians and students in higher education institutions in the UK.

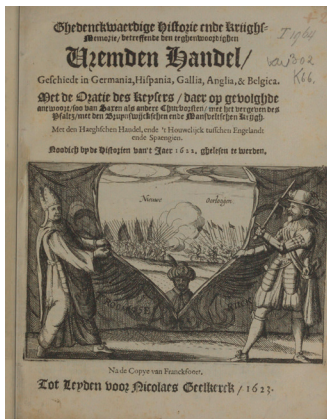
*This summary highlights 10 key findings. The full study is available:*

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

# 10 Key Findings

"I couldn't do my research without EEBO. In fact I wonder how much of my research has been shaped by having EEBO"

(UK Research Student)



"It's a great resource for regional and local history... we have students scattered all around the four nations... it is a pretty valuable bank of stuff that students can enjoy."

(Open University Faculty Member)

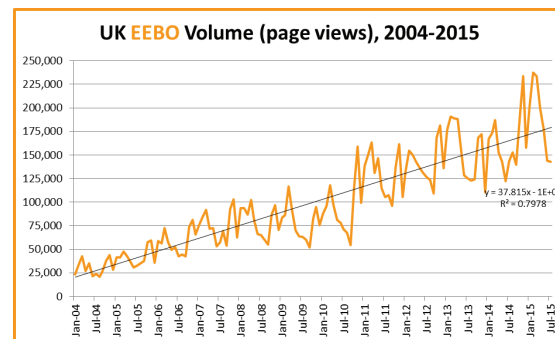
## One: The use of digital resources is changing, but in an incremental way

The impact of digital collections is not a 'big bang' moment that immediately changes everything about humanities research. One reason digital resources have been so successful is they are familiar to researchers. For scholars trained to work with aged documents open on a library table, seeing high-quality scans is reassuring, and possibilities for doing research with traveling to far-flung libraries is compelling. Digitization allows the scholar to roam in and out of the archive and examine any text computationally or serendipitously. However, even these digital resources still take time to become embedded in various academic practices, and this is still an ongoing process.

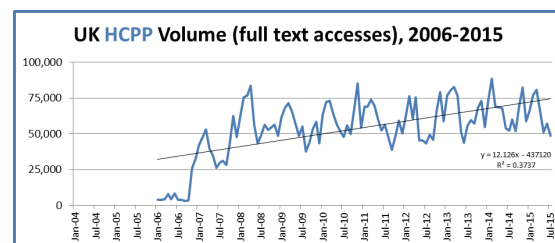
## Two: EEBO and HCCP usage has increased steadily over the past decade

The House of Commons Parliamentary Papers found its audience relatively quickly, and then saw its usage remain stable.

Early English Books Online, on the other hand, appears to still be finding new audiences. While the growth may have slowed a bit in recent years, there is room for growth before it stabilizes. In September, 2016, new content was added to this valuable resource!



EEBO usage has been increasing at a relatively linear pace. Whereas usage doubled from 50k page views per month to 100k per month from 2005-2008, it stayed there until 2011. The next doubling to 200k page views per month took a further 4 years, underscoring the fact that digital collections take time to become embedded in research.



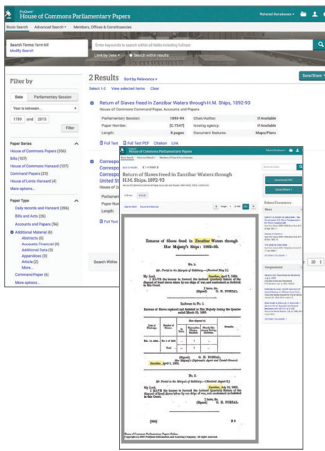
HCCP usage also increases in a linear, but less marked fashion. Monthly full-text accesses increased rapidly to 50k by 2007, but have fluctuated between approximately 50-75k full-text accesses per month ever since.

## Three: While top universities are most likely to use EEBO and HCCP, less research-intensive institutions also benefit from both collections

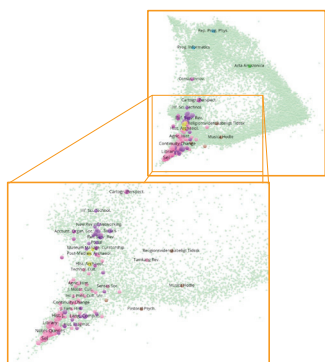
We found a cluster of courses on literature and book history offered at a non-Russell Group, non-research-intensive university, an opportunity for EEBO that would not have been possible based on the institutions' limited special collections. The EEBO-based courses let students discover rare books, leading to interest in book history and its subsequent disciplines.

At one Open University, the use HCCP in History and Law was heavier than expected, even among the part-time students.

However, researchers at top-ranking institutions still use the resources more heavily. For EEBO, faculty 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. The difference is less marked for HCCP; while faculty in the top-third of the History REF rankings include 61% of all History REF faculty, HCCP usage accounts for 59% of all HCCP. In other words, uptake of these resources is strongest in universities that performed well in the REF.



“I just use EEBO all of the time.”  
(UK Researcher)



Source: Scopus data visualized with VosViewer, using overlay map data as described in Leydesdorff, Moya-Anegón, and Guerrero-Bote (2015).

#### Four: Researchers return to digital collections regularly, resulting in incremental changes in scholarly behaviour

In this study's survey, **97% of respondents (n=238) reported as “important” that specific databases or collections are a method for discovering primary sources, ranking them highly over other methods like Google.** Many participants were early adopters of EEBO and HCPP, and recalled the transformative impact of these digital collections on their research.

Over time, the part digital resources play in allowing serendipitous research has become clearer. 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.

#### Five: Resource use in the humanities is diverse, which makes providing access a challenge

While researchers pointed to a few collections with very high regular use, there is also diversity of resource usage. In fact, **survey respondents listed 136 digital collections.**

One difference between the humanities/social sciences and other disciplines is the huge diversity not only of research topics, but also of the tools needed. In previous research, we showed that some fields like nuclear physics rely on a well-defined set of journals and have little need to search widely beyond those sources. In the humanities, there is no similar one-stop location. This makes providing infrastructure that can support this diversity more of a challenge.

#### Six: Citation evidence shows growing mentions of EEBO or HCPP, and these publications in turn are reasonably well-cited

Source	EEBO	HCPP
Google Scholar	5,950 <sup>18</sup>	1,780 <sup>19</sup>
Google Scholar cleaned & analyzed <sup>20</sup>	799	506
JSTOR	296 <sup>21</sup>	73 <sup>22</sup>
Scopus <sup>23</sup>	467 <sup>24</sup>	191 <sup>25</sup>
ProQuest Dissertations & Theses	1,002 <sup>26</sup>	147 <sup>27</sup>

Using conservative estimates of the publication impact of EEBO and HCPP over that last 15 years, we saw growth in publications referencing these two resources – including mentions of individual books, articles, and dissertations and theses.

But recent data shows a decline in author citations. It is likely that publications are relying on EEBO and HCPP, but that this use is undetectable using search methods. Only a minority of authors who used these resources included any indication in their citations. This is an issue we noted previously (Meyer, 2011; Meyer, et al., 2009; Siefing & Meyer, 2013) and it is clear there is still room for additional training as well as system design that can nudge scholars toward citing digital resources, such as support for automatic citation and human-readable URLs (Meyer, 2011, pp. 41, 56).

Also, while citations to these works dropped off in recent years, humanities disciplines tend to have a longer time-scale before publications reach their peak citations, often taking 6-8 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 more diverse than expected

Several pieces of evidence point to the wide diversity of disciplines using EEBO and HCPP. **EEBO has been mentioned in 773 publication outlets. 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. HCPP was mentioned in 508 publication outlets.** For HCPP, journals focusing on historical and legal specializations 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 (literature, history, theatre, music, religion, and political science) and HCPP (history, law, economics, anthropology, and women's studies).



"I use primary resources a lot in teaching and digital resources are a really good way of getting those, and I have lots of activities where students have to go and find a particular thing from a digital resource."

(UK History Faculty Member)

"I do work from home a lot. So it does mean instant access to things that I'd perhaps have to wait to do, which in some ways I think is a good thing."

(UK Researcher)

#### **Eight: Researchers are more concerned with the digital content and functionality than in who provides the access**

While researchers are passionate about gaining access to vital content, they are less concerned with how that access is gained. Certainly users will have (often strong) opinions about one interface over another, but they have little interest in whether access was provided by their library, a national broker such as Jisc, or through some other mechanism.

At some level, libraries and organizations like Jisc became victims of their own success by providing seamless access. Whereas users once struggled to remember passwords and subscription details, most users today – on their campus network or remotely – access primary collections and journals with a single click and no further authentication. Often, it is only when traveling from their home institution or taking a job elsewhere that people become aware that the access they expect is no longer available.

#### **Nine: The UK has seen the benefits of investing at a national-level and 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. Among many reasons is the expectation at any leading knowledge institution that their scholars will be able to access the research materials and publications 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 that scholars rely on these digital collections. Since 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 evident, although there is still plenty of room for growth, particularly in teaching**

Increasingly active communities of scholars are doing digital scholarship with EEBO/HCPP. The EEBO-TCP partnership, for example, 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.

## Conclusion: Digital collections are fundamental to modern scholarship



Even though digital collections have the advantage of demanding relatively little of researchers, the growth in usage of these collections takes time, and the measures of impact such as citations 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.

The consistent story across a whole series of studies done by our team and others 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.



**Learn more... explore more... discover more - visit [www.proquest.com](http://www.proquest.com)**