The Impacts of Digital Collections

Early English Books Online &

House of Commons Parliamentary Papers

Eric T. Meyer & Kathryn Eccles
Oxford University Consulting
Oxford Internet Institute, University of Oxford
01 March 2016

OXFORD
UNIVERSITY
CONSULTING

**The Impacts of Digital Collections
Early English Books Online & House of Commons Parliamentary Papers**

Final Report on Usage and Impact

This report was funded by JISC and ProQuest, and is an output of Oxford University Consulting (<http://isis-innovation.com/>) and the Oxford Internet Institute (<http://www.oii.ox.ac.uk>), both at the University of Oxford.

All images by the authors unless otherwise indicated.

Questions or queries about this report may be directed to:

Professor Eric T. Meyer
Oxford Internet Institute, University of Oxford
1 St Giles, Oxford, OX1 3JS, United Kingdom
Tel: +44 (0) 1865 287210
Email: eric.meyer@oii.ox.ac.uk

Please cite as:

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>

Table of contents

Acronyms & Abbreviations	4
Introduction	5
<i>EEBO: Early English Books Online.....</i>	5
<i>HCPP: House of Commons Parliamentary Papers</i>	5
Research design & methods.....	6
<i>Usage Statistics</i>	6
<i>Bibliometrics.....</i>	6
<i>Survey.....</i>	7
<i>Users.....</i>	7
Usage Data	8
Overall usage of the ProQuest and Jisc platforms by UK institutions.....	8
Geographic usage of the ProQuest platforms by UK institutions	10
Detailed usage of the ProQuest platforms by types of UK institutions	11
Bibliometrics	14
Total Publications based on EEBO and HCPP	14
EEBO Publication Overview.....	16
HCPP Publication Overview.....	20
Analysis of Dissertations and Theses	22
Qualitative Impacts	25
Interviews and Opinion-Gathering.....	25
<i>Participants</i>	25
Key themes.....	25
<i>Early adopters</i>	25
<i>Centrality, access, funding</i>	26
<i>Ease of access/Keyword searching.....</i>	27
<i>Materiality.....</i>	27
<i>Serendipity and Sharing</i>	28
<i>Teaching context.....</i>	29
<i>HCPP at the Open University.....</i>	30
<i>Digital citation.....</i>	31
<i>Digital Humanities approaches to EEBO and HCPP</i>	32
Conclusions	33
Survey of Researchers	34
Use and Awareness of Selected Digital Resources	34
Other important resources	35
Important methods for discovering primary sources	37
EEBO and HCPP in detail	39
<i>Attitudes toward EEBO (sorted by mean overall importance).....</i>	39
<i>Attitudes toward HCPP (sorted by mean overall importance).....</i>	40
<i>Method for Accessing EEBO</i>	41
<i>Method for Accessing HCPP.....</i>	41
<i>Reasons for using EEBO.....</i>	42
<i>Reasons for using HCPP.....</i>	42
Citing Digital Sources.....	42
Improvements	44
Demographics	47
Key Findings & Conclusions	49
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.	49

Two: The usage of both Early English Books Online and House of Commons Parliamentary Papers has been increasing steadily over the past decade.....	49
Three: While researchers at top universities are most likely to use EEBO and HCPP, less research-intensive HE institutions also benefit from both collections.	49
Four: Researchers rely heavily on specific digital collections that they return to regularly, which is resulting in incremental changes in scholarly behaviour.	50
Five: Resource use in the humanities is extremely diverse, and this makes providing access to needed resources and tools particularly challenging.	51
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.	51
Seven: The number and range of disciplines that refer to EEBO and HCPP is much more diverse than expected.....	51
Eight: Researchers are more concerned with the content and functionality of the digital collections than in who provides the access.....	52
Nine: The UK is unusual for providing national-level access across institutions through Jisc's national purchasing.....	52
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.....	52
Conclusion: Digital collections have become fundamental to modern scholarship.	53
References Cited	54
Appendix 1: Interview Guides	55
EEBO Interview Guide	55
HCPP Interview Guide	55

Acknowledgements

The authors wish to thank Jisc and ProQuest for funding this research, and also to both Jisc and ProQuest for providing data. We are also indebted to the participants in our interviews and surveys who have generously provided their time and thoughts.

Acronyms & Abbreviations

CSV	Comma-separated values, a text-based file format for data exchange
EEBO	Early English Books Online
FE	Further Education, which focus on qualifications for careers or to enter university
HCPP	House of Commons Parliamentary Papers
HE	Higher Education, mainly universities in the United Kingdom
OII	Oxford Internet Institute, University of Oxford
REF	Research Excellence Framework
TIDSR	Toolkit for the Impact of Digitised Scholarly Resources (http://bit.ly/tidsr)

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.

These projects have all used the TIDSR (Toolkit for the Impact of Digitised Scholarly Resources) toolkit (<http://bit.ly/TIDSR>) to better understand the quantitative and qualitative empirical evidence for the impacts that digital collections can be shown to have. The toolkit was initially funded by Jisc in 2008 as part of a project to measure the impact of what was called Phase 1 digitisation, five digital collections that had been digitised with support from Jisc. It has since been used by many additional projects¹ and has been widely recognised as best practice in numerous publications, as well as by the AHRC².

EEBO: Early English Books Online

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.³ 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.

The Early English Books Online was studied extensively in 2012-2013 as part of a project⁴ funded by Jisc and carried out by the University of Oxford Bodleian Libraries in partnership with the Oxford Internet Institute. The report from that project (Siefring & Meyer, 2013) was the starting point from which this project updated the evidence of EEBO impacts. That project focused in particular on EEBO-TCP (Text Creation Partnership), whereas this study expanded the focus to look at EEBO more generally.

HCPP: House of Commons Parliamentary Papers

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.⁵ HCPP is available to institutions through Jisc as part of the ProQuest Archives 2014-2017 or via subscription from ProQuest.

¹ <http://microsites.oii.ox.ac.uk/tidsr/selected-references-toolkit>

² Call document: [Digital Equipment and Database Enhancement for Impact \(DEDEFI\) Scheme](#) (AHRC, September 2009)

³ <http://www.proquest.com/products-services/databases/eebo.html>

⁴ <http://microsites.oii.ox.ac.uk/tidsr/case-study/464/eebo-tcp>

⁵ <http://www.proquest.com/products-services/House-of-Commons-Parliamentary-Papers.html>

The House of Commons Parliamentary Papers has not previously been studied *in toto*, but a portion of what is now the HCPP (specifically the 18th Century Official Parliamentary Publications Portal, part of BOPCRIS, the British Official Publications Collaborative Research Information Service) was included as part of the original Jisc-funded TIDSR usage and impact study (Meyer, Eccles, Thelwall, & Madsen, 2009). Because those data are both older and focused on only one part of the HCPP before it was moved to ProQuest, the data reported here are wholly new findings.

Research design & methods

Usage Statistics

Usage statistics for both EEBO and HCPP were provided by ProQuest and Jisc for the period from 2004-2015. These data included results aggregated at the monthly level per UK institution that has a subscription to the collections. The statistics include numbers of web sessions, searches, 'hits', document views, PDFs access, full text views, searches and other data. Some of these data are inconsistent across the time period due to changes in access modes and policies, so we have relied most heavily in this report on the most consistent fields: [Document/Page Image Views] (ProQuest) and [Page Views] (Jisc) in EEBO and [Full Texts Accessed] in HCPP. These are slightly different in practice, but provide comparable data in terms of trends. The absolute numbers, however, cannot be directly compared for any of the variables available.

The data were provided by ProQuest in Excel format, which were then imported into an Access database where they were cleaned and combined with publicly available data⁶ from the 2014 Research Excellence Framework (REF) exercise in the UK, and geo-coded using data obtained from Google Maps. As institutional names were not consistent among the various sources, a table was used to link the different data sources.

Additional data were provided by Jisc from the Historical Texts platform, dating from June 2014 until the end of 2015. These data were also in Excel format, and were imported into the same database and linked with the other data described above.

Bibliometrics

Bibliometric data from 4 separate sources were extracted and combined for analysis. Scopus,⁷ Google Scholar,⁸ JSTOR,⁹ and ProQuest Dissertations and Theses Global¹⁰ were queried to extract publications that somewhere mentioned EEBO or HCPP. Details of the queries used are included in the footnotes with the data below. This is not an exhaustive approach, since we know from previous work (Siefring & Meyer, 2013) that at least half of all authors using resources such as EEBO do not include any indication that they have used the digital collection in their work, and simply cite the original source as if they consulted the paper version. Unfortunately, we have no reliable way to discover these uses automatically.

Scopus data were downloaded by exporting the results to CSV text files containing authors, publication details, citation information, and author address information. The data were then imported into Access for cleaning and analysis. The author details and address information in particular is not in a queryable format, so we used bespoke code to automatically extract this information from each record.

JSTOR data were downloaded from JSTOR Data for Research in tab-delimited files, which were then imported into Access for cleaning and analysis.

Google Scholar data were downloaded using *Harzing's Publish or Perish* tool, which allows one to download approximately 1000 Google Scholar results based on the number of citations. Further results are not returned by the Google API, and we did not circumvent this limitation.

⁶ <http://www.ref.ac.uk/>

⁷ <http://www.scopus.com/>

⁸ <https://scholar.google.co.uk/>

⁹ <http://www.jstor.org> and <http://dfr.jstor.org>

¹⁰ <http://search.proquest.com/pqdtglobal/>

The Scopus, JSTOR, and Google Scholar data were combined by title (after considerable manually-assisted cleaning based on source and title). The title data are the only items suitable for finding matches between the datasets, but the titles are particularly problematic for a study of this sort as many of the publications use non-ASCII characters in their titles, which the three systems interpret differently and export in non-compatible formats.

ProQuest Dissertation and Theses Global data were exported using two separate mechanisms via ProQuest directly in Excel format and also via RefWorks to a separate Excel file, as different fields are available via these two mechanisms. The data were then recombined and cleaned in Access locally.

Survey

We have previously surveyed users of the 18th C. Parliamentary Papers (as part of the original TIDSR study) and users of EEBO (as part of the EEBO-TCP study) using a bank of questions that have been tested both in these studies but also in additional studies funded by the ESRC and by the Research Information Network (RIN). For this project, we used updated questions based on those from these previous studies to include specific questions about EEBO and HCPP, but also about the general value of digital resources and their impacts on research and teaching (as with previous surveys). The online survey was done with professional software from Qualtrics.

Users

Throughout the project, we conducted sixteen interviews with users of both the EEBO and HCPP platforms (as indicated above). We contacted relevant stakeholders discovered during our research, using snowball sampling to find additional interviewees with relevant information, and were provided by Jisc with contact details for an additional set of interviewees.

Usage Data

The usage data in this report are based on UK usage statistics provided by ProQuest for the period January 2004 – July 2015 (EEBO) and January 2006 – July 2015 (HCPP) and by Jisc for the period June 2014 – July 2015. For ease of understanding, EEBO data are reported in **orange** throughout this report, while HCPP data are shown in **blue**.

Overall usage of the ProQuest and Jisc platforms by UK institutions

The volume statistics¹¹ show a steady increase in usage from 2004-2015. Usage data from 2004 (for EEBO) and 2006 (for HCPP) from ProQuest and Jisc show an upward trend. The EEBO volume includes data from the Jisc Historical Texts page view volume, which starts in June 2014.

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 3 years (2005-2008),

¹¹ The volume is measured slightly differently on EEBO and HCPP due to differences in how the data are collected and recorded. For EEBO, the volume is measured as the total of the field [Document/Page Image Views] for data from ProQuest, and [Page Views] for data from Jisc Historic Texts (starting in June 2014). In HCPP, the volume is measured by the related, but slightly different, field [Full Texts Accessed]. As a result, the absolute volumes cannot be directly compared. Other available fields, however, such as Searches, proved to have noise in the data that made them unsuitable for our purposes.

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 4 years.

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 50-75k full-text accesses per month ever since.

Neither of these is 'better' than the other, rather this demonstrates different patterns of use for digital collections. The House of Commons Parliamentary Papers collection appears to have found its audience relatively quickly, and its usage remains 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 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.

The rises and falls in the data are unsurprising, and show the sort of annual cycle one would expect to see in primarily academic resources, with increased volume during academic terms and reduced volume during holidays, as shown in the following charts.

Here we can see that annual patterns of use (combining data from the same month each year across the entire dataset) follow the academic year. The lowest levels of use for both resources are in August and September, steadily increasing to a peak in November. This is followed by a short dip during the December holidays, increasing to another peak in February and March. As we will see below, academic use makes up the bulk of the uses of both EEBO and HCPP, although it is worth noting that even though the absolute volume for EEBO is the higher of the two collections (again, reiterating that the data are not completely comparable), the HCPP data shows a less steep difference between the peaks and troughs in usage. This suggests that there are uses of HCPP that are slightly less tied to the academic year than those of EEBO, which as we will see is borne out in more detailed analysis of the data below.

Geographic usage of the ProQuest platforms by UK institutions

By geo-coding the data to match the usage data to UK institutions, we were able to show the spread of use across the UK. The results¹² are mapped below. Several things are apparent in these two maps. First, use of both resources is spread across the entire UK, although the intensity of use (as shown by the size of the bubbles) tends to be more heavily concentrated in London and the southeast. Of course, many of the UK's largest higher education (HE) institutions are located in these areas, so one would expect that there would be a lot of use of research and educational resources in these areas. A second thing to note is the relatively broader spread of HCPP use in the UK; this is partly due to more use of HCPP by further education (FE) institutions, as compared to EEBO (more details below).

¹² Data visualized using CartoDB (<https://cartodb.com>).

Detailed usage of the ProQuest platforms by types of UK institutions

This broad-scale data doesn't give much information about who is actually using these two resources. For that, one must further examine the data based on the types of institutions using the collections, as shown in the following tables.

EEBO							
	N _{tot}	FTE	% FTE	N _{eebo}	Volume	Usage % (REF)	Usage % (Overall)
Overall Total	214			103	13,409,918		100%
Universities	155			87	12,660,772		94%
English REF	89	1813.1	100%	74	12,575,811	100%	94%
Top third	30	1045.5	58%	28	10,113,015	80%	75%
Middle third	30	534.0	29%	27	1,732,779	14%	13%
Bottom third	29	233.7	13%	19	730,017	6%	5%
History REF	83	1626.0	100%	70	12,448,748	100%	93%
Top third	28	1037.2	64%	27	9,907,182	80%	74%
Middle third	28	440.3	27%	24	1,881,321	15%	14%
Bottom third	27	148.5	9%	19	660,245	5%	5%
FE Colleges	41			6	14,953		<1%
Museums/Libraries	7			6	239,125		2%
Other Institutions	10			4	495,068		4%

Source: Data provided by ProQuest for EEBO (2004-2015). REF data from public sources. FTE stands for Full Time Equivalent (e.g. 1 full-time position), and shows the total number of FTE faculty submitted to REF panels by each university. The % FTE column shows what proportion of all REF faculty were submitted by institutions that were subsequently ranked in the top, middle, and bottom third of the REF results. N_{tot} shows the total number of institutions that fall into each category, while N_{eebo} shows the number that are also present in the EEBO usage data. Volume shows the overall volume (measure by page views) of each type of institution. Usage % (REF) shows the relative proportion of usage as measured by volume that originates from each of the types of REF institution. Usage % (Overall) shows the relative proportion of usage as measured by volume that originates from each type of institution overall.

HCPP							
	N _{tot}	FTE	% FTE	N _{hcpp}	Volume	Usage % (REF)	Usage % (Overall)
Overall Total	214			167	6,141,990		100%
Universities	155			122	5,381,595		88%
English REF	89	1952.9	100%	87	4,984,644	100%	81%
Top third	30	1072.2	55%	30	2,694,609	54%	44%
Middle third	30	568.9	29%	29	1,667,942	33%	27%
Bottom third	29	311.8	16%	28	622,093	12%	10%
History REF	83	1781.4	100%	82	5,115,269	100%	83%
Top third	28	1081.2	61%	28	2,995,959	59%	49%
Middle third	28	478.5	27%	28	1,571,596	31%	26%
Bottom third	27	221.7	12%	26	547,714	11%	9%
FE Colleges	41			36	7,295		<1%
Museums/Libraries	7			1	26,827		<1%
Other Institutions	10			7	726,171		12%

Source: Data provided by ProQuest for HCPP (2006-2015). REF data from public sources. FTE stands for Full Time Equivalent (e.g. 1 full-time position), and shows the total number of FTE faculty submitted to REF panels by each university. The % FTE column shows what proportion of all REF faculty were submitted by institutions that were subsequently ranked in the top, middle, and bottom third of the REF results. N_{tot} shows the total number of institutions that fall into each category, while N_{hcpp} shows the number that are also present in the HCPP usage data. Volume shows the overall volume (measured by full text accesses) of each type of institution. Usage % (REF) shows the relative proportion of usage as measured by volume that originates from each of the types of REF institution. Usage % (Overall) shows the relative proportion of usage as measured by volume that originates from each type of institution overall.

In the more detailed data shown above, several interesting elements are worth pointing out. First, the bulk of UK volume for both EEBO and HCPP are made up of traffic from universities (94% and 88% respectively).

For EEBO, researchers from 87 different universities viewed over 12.6 million documents/page images, and the vast bulk of this traffic was from universities that submitted faculty to the 2014 Research Excellence Framework (REF)¹³ in English¹⁴ or History.¹⁵ In addition, there is an imbalance in EEBO use toward institutions in the top-tier of institutions. Looking at institutions with English Language and Literature faculty, those ranked in the top third in terms of research power (Research Fortnight, 2014) account for 58% of the faculty (in the % FTE column) but 80% of the EEBO usage volume. Likewise, institutions with History faculty in the top third (which is not independent of the English faculty data since many institutions are top-ranked in both fields) account for 64% of faculty but 80% of EEBO usage. In other words, the top institutions as measured by REF outputs use the EEBO collection even more than their greater numbers would suggest on their own. Conversely, those institutions in the bottom third account for a smaller proportion of faculty (13% of English and 9% of History), but also an even smaller portion of EEBO use (6% and 5% respectively). This suggests that usage of digital collections and the presence of leading researchers is correlated.

Turning to the HCPP data, we see less of an imbalance between the size of the faculty and the use of HCPP. Here, usage percentages tend to track relatively closely with the percent of faculty in English and History at leading institutions: top-tier History departments, for instance, comprise 61% of submitted REF faculty, and account for 59% of HCPP usage volume.

Further Education (FE) colleges make up a relatively small proportion of volume for both EEBO and HCPP, and HCPP is used (relatively lightly) by many more FE colleges (36, compared to only 6 using EEBO). This would indicate that FE colleges are interested in the HCPP content; whether the actual use of HCPP could be increased through an education campaign is beyond the scope of this project, but worth speculating about.

Museums and libraries that are not specific to a particular institution, on the other hand, are much more likely to use EEBO than HCPP. The six museums and libraries in the data are much more likely to use EEBO; one (institution number 8 in the next table) is a particularly heavy user. One striking difference in the HCPP data is the relatively heavy use (12%, n=726,171 full texts accessed) by non-REF institutions. The reason for this becomes clearer in the next table.

In the table below, we look at the EEBO and HCPP volume data by top institutions. The table above lists all the subscribing UK institutions that account for at least 1% of the total volume on each platform. The names of individual institutions have been removed for commercial reasons.

In the table, institutions are numbered according to their position in the table of EEBO volume. So, institution 1 (a Higher Education institution in Jisc Band A, which is the most expensive subscription band) accounts for 16.1% of EEBO usage (thus is the heaviest user), and the same institution accounts for 5.6% of HCPP usage (the second heaviest user). The two halves of the table can be compared to understand the different patterns of use.

For instance, the heaviest HCPP-using institution (number 35, accounting for 7.1% of HCPP use) accounts for less than 1% of EEBO use. In fact, nearly a third of the top HCPP-using institutions (10 out of 33) do not fall into the top 30 EEBO using institutions that account for 1% or more of overall volume.

In the Early English Books Online data, one of the most striking results is the huge volume originating from a single university. However, this finding should not be overstated, as a number of library staff at that institution were employed to work on a large project related to EEBO.

¹³ The Research Excellence Framework, or REF, is a "system for assessing the quality of research in UK higher education institutions" (<http://www.ref.ac.uk/>). A total of 154 universities submitted work in 2014 based on the publications of over 52,000 academic staff, which was rated as world-leading (4*), internationally excellent (3*), recognized internationally (2*) or recognized nationally (1*).

¹⁴ As a measure of the size and quality of English faculty at REF institutions, we used the data from Panel 29, English Language and Literature. A total of 89 institutions made submissions to this panel.

¹⁵ As a measure of the size and quality of History faculty at REF institutions, we used the data from Panel 30, History. A total of 83 institutions made submissions to this panel.

Rank	EEBO					HCPP				
	Inst	Jisc Band	Type	Volume	% Overall	Inst	Jisc Band	Type	Volume	% Overall
1	1	A	HE	2,154,506	16.1%	35	A	HE	437,982	7.1%
2	2	D	HE	742,269	5.5%	1	A	HE	346,753	5.6%
3	3	A	HE	735,203	5.5%	8 •	E	HE	315,221	5.1%
4	4	C	HE	601,356	4.5%	3	A	HE	259,371	4.2%
5	5	D	HE	486,912	3.6%	17	C	HE	214,294	3.5%
6	6	A	HE	451,460	3.4%	104 •	C	Aff	173,874	2.8%
7	7	B	HE	440,241	3.3%	24	B	HE	171,843	2.8%
8	8 •	E	HE	438,125	3.3%	2	D	HE	163,215	2.7%
9	9	A	HE	401,557	3.0%	10	A	HE	145,698	2.4%
10	10	A	HE	397,817	3.0%	5	D	HE	143,771	2.3%
11	11	A	HE	392,323	2.9%	11	A	HE	140,597	2.3%
12	12	A	HE	364,663	2.7%	25	B	HE	132,121	2.2%
13	13	A	HE	324,458	2.4%	105	E	HE	127,321	2.1%
14	14	C	HE	323,415	2.4%	26	D	HE	123,547	2.0%
15	15	D	HE	277,755	2.1%	4	C	HE	115,066	1.9%
16	16	A	HE	272,801	2.0%	12	A	HE	113,354	1.8%
17	17	C	HE	232,048	1.7%	33	D	HE	101,259	1.6%
18	18	E	HE	229,955	1.7%	36	D	HE	99,551	1.6%
19	19	B	HE	209,641	1.6%	9	A	HE	93,864	1.5%
20	20	B	HE	208,906	1.6%	6	A	HE	93,794	1.5%
21	21 •	F	HE	207,070	1.5%	13	A	HE	88,096	1.4%
22	22	F	HE	206,584	1.5%	18	E	HE	79,758	1.3%
23	23	D	HE	185,625	1.4%	71	C	HE	76,373	1.2%
24	24	B	HE	176,194	1.3%	30	B	HE	70,015	1.1%
25	25	B	HE	174,197	1.3%	14	C	HE	69,540	1.1%
26	26	D	HE	160,437	1.2%	55	E	HE	67,841	1.1%
27	27	E	HE	146,923	1.1%	16	A	HE	66,988	1.1%
28	28	D	HE	142,786	1.1%	53	C	HE	63,841	1.0%
29	29	E	HE	130,672	1.0%	40	E	HE	63,394	1.0%
30	30	B	HE	127,448	1.0%	20	B	HE	61,588	1.0%
31						58	G	HE	60,778	1.0%
32						34	B	HE	60,325	1.0%
33						23	D	HE	58,934	1.0%

Source: Data provided by ProQuest for EEBO (2004-2015) and HCPP (2006-2015). The percentage of overall volume is calculated using the total volume from the previous table. Only institutions that account for at least 1% of total volume are shown. Institution names are not included for privacy and commercial reasons. The Jisc Banding is the basis for subscription fees for Jisc-provided online resources in our Catalogue; all institutional members (Further and Higher Education institutions, Specialist Colleges and Research Councils in the UK) are ranked into one of ten Jisc Bands appropriate to each sector.¹⁶

• = Non-REF

The other universities in the top-five are also unsurprising, as all were in the top-10 in the REF Power Ranking (Research Fortnight, 2014). The other most active institutions were also REF universities, with the exception of a library which serves more than one REF university and a major research organization.

For the House of Commons Parliamentary Papers data, we see less dominance of a single university, and also the presence of more non-REF institutions. There is a major library in the seventh spot. At the top of the list in terms of volume, however, we have a university which is not a heavy user of EEBO and further which is considerably smaller in terms of research staff (14.3 submitted to the history REF panel, compared to 130 submitted by the top EEBO-using institution) and considerably lower power (~40th) and quality (~30th) rankings. However, this university's use of HCPP highlights the fact that HCPP isn't used just for research by faculty but also that done by students in the course of their studies. The students at this one university at least are making heavy use of this online resource; additional investigation would be required to understand their use of other resources.

¹⁶ See <https://www.jisc-collections.ac.uk/Help-and-information/JISC-Banding/> for more details. While new bands were introduced in 2013, the shift to new bands will only be complete in 2017, so we have used the old banding here.

Bibliometrics

Bibliometric analysis of sources citing EEBO & HCPP were performed using data from Scopus, Google Scholar, ProQuest, and JSTOR. These results show the extent to which EEBO and HCPP are mentioned and cited in the literature, although they do not necessarily find all uses of the resources that did not cite or mention them by name. As a result, these findings need to be understood NOT to represent “publications using EEBO or HCPP”, which would be a larger but largely undiscoverable set of publications. This is due to the fact that humanities scholars who are the primary users of these resources tend not to cite materials in such a way as to be able to detect their digital origins (Blaney, 2014).

Thus, these results, which focus on publications that either mentioned EEBO or HCPP or include a citation that does so, are biased towards publications discussing the resources themselves, or that specifically acknowledged them somewhere in the text, notes, or references. The numbers should be seen as the most conservative estimates of the citation impact of the resources, and the actual uses of materials in the collections may be much higher. Even with these limitations, however, we can nevertheless glean interesting insights into the use of both resources.

Total Publications based on EEBO and HCPP

Publications related to Early English Books Online and House of Commons Parliamentary Papers¹⁷

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

It is apparent in the data shown above that both resources are having an impact in the published literature. One thing to keep in mind is that all the databases use different search mechanisms and index different bodies of literature. For instance, both JSTOR and Scopus allow searching all fields in the database, but have a narrower range of materials available to search (i.e. only those publications included in each index, which are selected using fairly stringent criteria of impact and scholarly importance). Google Scholar, on the other hand, searches a much wider selection of publications, including not just journal publications, but also things such as reports,

¹⁷ All searches carried out on various days in August 2015. Search counts are before cleaning and combining data. See Figure 1 for post-cleaned data numbers used for analysis.

¹⁸ Google Scholar EEBO search terms: "eebo-tcp" OR "eebo tcp" OR eebo OR "early english books online"

¹⁹ Google Scholar HCPP search terms: "House of Commons Papers" OR "House of Commons Parliamentary Papers" OR "parlipapers.chadwyck.co.uk"

²⁰ Google Scholar's API limits downloading to the first 900-1000 results, so only approximately the 1000 most cited were available for more detailed analysis. Google data extracted using Harzing's Publish or Perish (<http://www.harzing.com/pop.htm>).

²¹ JSTOR EEBO search term: eebo-tcp OR "eebo tcp" OR eebo OR "early english books online" in full-text, including all content

²² JSTOR HCPP search term: "House of Commons Papers" OR "House of Commons Parliamentary Papers" OR "parlipapers.chadwyck.co.uk"

²³ Web of Science is normally also used in bibliometric analysis. However, due to limits of its coverage, the search terms of interest here yielded very small samples (EEBO n=25, HCPP n=10), so these data were not included in the analysis.

²⁴ Scopus EEBO search term: ALL("eebo-tcp" OR "eebo tcp" OR eebo OR "early english books online" OR "quod.lib.umich.edu/e/eebogroup" OR "www.textcreationpartnership.org/tcp-eebo" OR

"www.bodleian.ox.ac.uk/eebotcp/" OR "eebo.chadwyck.com" OR "data.historicaltexts.jisc.ac.uk/view?pubId=eebo")

²⁵ Scopus HCPP search term: ALL("House of Commons Papers" OR "House of Commons Parliamentary Papers" OR "parlipapers.chadwyck.co.uk")

²⁶ ProQuest Dissertations & Theses Global EEBO search term: "eebo-tcp" OR "eebo tcp" OR eebo OR "early english books online", Years 2000-2014

²⁷ ProQuest Dissertations & Theses Global HCPP search term: "House of Commons Papers" OR "House of Commons Parliamentary Papers" OR "parlipapers.chadwyck.co.uk", Years 2000-2014

unpublished documents hosted on academic servers, and presentations.²⁸ One difficulty with the Google Scholar data, however, is that the data are not fully accessible for analysis; the Google Scholar API limits downloading to the first 900-1000 responses, which means that a full analysis of the Google Scholar data is not possible without circumventing those limits (which we did not do). This is why the row “Google Scholar cleaned & analysed” is less than 1000, once duplicates and erroneous documents were removed from the downloaded data. We chose to sort by citation frequency, so these are the most cited papers from Google Scholar.

The ProQuest Dissertations & Theses database indexes doctoral and (to a lesser extent) masters theses globally. The coverage, however, is uneven, as some countries have a nearly universal expectation that completed doctoral work will be deposited with ProQuest, whereas in other countries few or no universities use ProQuest to archive such materials. In the United Kingdom, relatively few universities use ProQuest; more use institutional archives such as the Oxford Research Archive,²⁹ DSpace@Cambridge,³⁰ Edinburgh Research Archive,³¹ and many others, while some make use of the EThOS service provided by the British Library.³² Unfortunately, searching all the individual university archives systematically is not feasible, and the British Library yielded too few results to be meaningful.

Given all those caveats, we can see some interesting results in our data.

EEBO and HCPP Publications by Source³³

The Venn diagrams above show a visual representation of the number of publications in our overall analysis set for each collection which originated in each of the main sources of data. While the relative sizes of each are somewhat interesting, even more interesting is the very low degree of overlap among the three. We did extensive and detailed checking of the results, and the differences appear mainly due to a combination of the sources indexed by each service and the search algorithms used by the service. When cleaning these data, it was apparent that there were many sources only indexed by one or two of the services. However, there were also a

²⁸ Google Scholar also indexes a small number of documents (n=44) directly from EEBO. These were excluded from the analysis.

²⁹ Oxford eTheses on ORA: http://www.bodleian.ox.ac.uk/ora/oxford_etheses

³⁰ DSpace@Cambridge: <https://www.repository.cam.ac.uk/>

³¹ Edinburgh Research Archive: <https://www.era.lib.ed.ac.uk/>

³² EThOS at the British Library: <http://ethos.bl.uk/>

³³ Venn diagram excludes ProQuest Dissertations & Theses, which have little to no overlap with the other sources

number of instances where the same journals were indexed, but the different search algorithms returned slightly different publications. So for instance, Google Scholar and Scopus together might have returned six articles from journal X, with both returning 3 of them but Scopus including 1 article that Google Scholar didn't find, and Google Scholar finding 2 that Scopus didn't find. Note in particular that only a tiny fraction ($n=14$ for EEBO and $n=4$ for HCPP) of publications were returned by all three search services.

This is interesting in terms of scholarly practice, since it reinforces the need to use multiple sources when searching for relevant literature. Only using JSTOR, which is popular among humanities scholars, yields the smallest number of results, but many of them are only available via JSTOR. Google Scholar, on the other hand, which is popular among research students due to its ease of use and ubiquity, yields many more results, but they tend to be less closely focused on the topic. In addition, the ability to more narrowly control one's search results is less sophisticated in Google Scholar compared to JSTOR or particularly Scopus, which has far more ability to precisely define the scope of one's search.

EEBO Publication Overview

Looking at the results of the Early English Books Online (EEBO) search over time, we see the expected sort of profile for a digital resource.

EEBO-related Publications and Citations of those Publications

All the sources allow one to readily count the publications per year, and these data show a steady growth in publications over the last decade, which indicates that the online collections are having a positive impact on scholarship. One question suggested by the data is whether the apparent drop in publications after 2012 represents a trend or not. The data from 2015 should be discounted, as it was based on partial data largely from the first half of the year, and also given the time between publication and those publications being added to the databases we queried (which would also affect the 2014 data to a lesser extent). The fewer publications retrieved with our methods from 2013 and 2014 also seem somewhat related to our method for working with Google Scholar and JSTOR data, as the declines in those sources account for the fall; the same time period using just Scopus data (which is the most reliable) shows increases through 2013, with a slight decrease in 2014 back to 2012 levels.

Nevertheless, one must consider several possibilities: first, that fewer papers are being written based on EEBO. This is the simplest explanation, but is not consistent with the other data in this

report. The second possibility is that papers are being written relying on EEBO, but that this use is undetectable using search methods (i.e. there is no mention made specifically of EEBO, nor a citation to any URL or other clear indication of reliance on EEBO). Based on our work with humanities scholars in this study and elsewhere, this non-citation of collections definitely happens, but whether it has happened more in the last couple of years than previously is not evident. We must conclude that while this trend is worth watching in the coming years, it does not yet merit serious concern.

In addition, we see that citations³⁴ to the articles that mention EEBO are also growing, which indicates a growing secondary impact on scholarship. The apparent decline in citations over the last five years should not be taken to be a particularly worrying trend, as the humanities tend to have a much longer time-scale before publications reach their peak citations than more rapidly moving fields such as medicine or physics, for instance. The actual half-life (i.e. amount of time before a publication has received ½ of all the citations it will ever receive) varies widely by discipline. One study by Davis (2013) estimated a citation half-life in the humanities between 4-5 years (compared to as short as 24 months in health sciences), while Tang (2008) calculated a citation half-life in history of 7.13 years. Wiberley (2003), looking at literary studies and art scholarship citation patterns, found that over half of citations in the sources he studied were to works published more than 20 years before their own publications.

While there is a somewhat anomalous peak in 2006 due to two highly cited (n=174 and 143) papers, the overall trend up through publications from 2009 or so is upwards. One would expect that papers published in the last five years will start to be more highly cited in the coming years, however being certain of that would require a follow-up study in several years.

It should also be noted that the humanities disciplines that are the primary users of EEBO (and HCPP) tend to be cited less frequently than the sciences or social sciences, where one would expect a higher number of citations for a set of publications this size. We will look at some more detailed citation data shortly.

EEBO Author Location, Scopus publications³⁵

Country of Author	N	% of total authors
United States	189	42.4%
United Kingdom	101	22.6%
Canada	43	9.6%
Australia	14	3.1%
Netherlands	9	2.0%
Spain	8	1.8%
Finland	7	1.6%
Total	446	100.0%

The Scopus data also lets us see the country of the authors in the publications, shown above. We can see that the United States and the United Kingdom are by far the most common locations of authors mentioning their use of Early English Books Online, followed by two other English-speaking countries, Canada and Australia. Note that this pattern of publication is unsurprising given general patterns of publication in English-language journals, where the US and UK tend to dominate across most disciplines.

If we look at the journals of the publications, we see a range of journals one might expect, focusing on English literature, language, and period studies.

³⁴ Citation data from Scopus and Google Scholar, using the maximum reported citations from each source. Citation data not available from JSTOR.

³⁵ Only countries with more than 1% of publications shown. 21 additional countries are also represented in the data.

EEBO Publication Sources

Publication Source	N Pubs	Total Max Times Cited	Average Citations per Pub
SEL - Studies in English Literature	80	8	0.1
Books (from Google Books)	30	343	11.4
Literature Compass	19	38	2.0
Studies in Philology	13	9	0.7
Modern Philology	12	15	1.3
Notes and Queries	12	0	--
The Review of English Studies	11	13	1.2
Shakespeare Quarterly	10	73	7.3
Renaissance Studies	10	14	1.4
Against the Grain	9	23	2.6
PMLA	9	22	2.4
Huntington Library Quarterly	9	7	0.8
Literary and Linguistic Computing	8	87	10.9
Shakespeare	8	24	3.0
Journal for Early Modern Cultural Studies	8	5	0.6
Early American Literature	7	4	0.6
ELH - English Literary History	7	2	0.3
The Serials Librarian	6	39	6.5
English Literary Renaissance	6	27	4.5
Language and Computers	6	20	3.3
Early Theatre	6	8	1.3
Renaissance Quarterly	6	2	0.3
Journal of Medieval and Early Modern Studies	6	1	0.2
ANQ - Quarterly Journal of Short Articles Notes and Reviews	6	0	--
The Historical Journal	6	0	--
College & Research Libraries	5	225	45.0
Journal of Library Administration	5	113	22.6
Historical Journal	5	18	3.6
Library	5	15	3.0
Milton Quarterly	5	9	1.8
Past and Present	5	9	1.8
Review of English Studies	5	5	1.0
Early Music	5	3	0.6
Texas Studies in Literature and Language	5	0	--
Sources with 4 publications	20	344	17.2
Sources with 3 publications	31	331	10.7
Sources with 2 publications	124	583	4.7
Sources with 1 publication	564	1885	3.3
Total³⁶	1094	4324	8.98

The most frequent publication in the data set is *SEL – Studies in English Literature*, with 80 publications mentioning EEBO in some way. It is difficult to know the readership of *SEL* without additional data, but the citation figure suggest that papers published in *SEL* do not tend to get cited. Only 8 citations to these 80 papers shows up in these data, which is consistent with *SEL*'s reported citations per document of 0.15.³⁷

The second most common source is Google Books, which reflects the tendency of humanities scholars to publish in books. The time it takes to research, write, and publish a book is also one

³⁶ Totals are slightly less than the total number of publications overall due to missing data.

³⁷ <http://www.scimagojr.com/journalsearch.php?q=5800211823&tip=sid>

of the reasons for the relatively long delay in growth of citations. Books in this sample, however, have one of the highest impacts, with an average of 11.4 citations per book. Other sources that have been cited particularly frequently include *Literary and Linguistic Computing* (10.9 citations per publication) and *Journal of Library Administration* (22.6 citations per publication). Again, however, we should stress that disciplinary differences are at play here: the latter two journals are in fields with higher citations numbers generally, so can't be compared directly to the sources that publish more typical humanities oriented content.

The general use of EEBO based on this sample, however, is rooted in humanities disciplines. In the following visualisation, we can see the publications in this EEBO sample overlaid on a standardized map of knowledge (Leydesdorff, Rafols, & Chen, 2013). The underlying map (the grey dots) was constructed by Leydesdorff and colleagues using all the journals in Scopus to calculate the frequency with which journals cited each other. Thus, two journals that cite each other frequently are located close to each other on the base map, and two journals that never or rarely cite each other are located far apart. The resulting map then can be used to visualise which areas of knowledge are represented by any given Scopus set of journals.

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

The bottom left-hand portion of the map is where the humanities including English, history, and the more humanities-focused social sciences reside. We can see that our EEBO sample is clustered quite heavily into that bottom left-hand portion. However, even though much of the use is in one area, EEBO is not exclusively used in (for instance) English language and literature publications, rather its influence spreads more broadly into a number of areas.

HCPP Publication Overview

Turning our attention now to similar data for the House of Commons Parliamentary Papers (HCPP) publications, we see a similar but slightly different pattern of publications in all areas.

HCPP-related Publications and Citations of those Publications

With the HCPP queries, we were able to retrieve fewer publications than with the EEBO search. While this is consistent with the apparently somewhat lower usage statistics for HCPP than EEBO, we suspect based on some test searches that it is also somewhat related to an even more entrenched preference for citing materials in the HCPP collection as if one consulted the paper version, and to include no mention of the reliance on HCPP either in the citations or in the text or acknowledgements of the paper. This makes discovering uses of HCPP via automated means particularly difficult, and is one of the reasons why we also report below on qualitative interviews focused on uses of HCPP, to better understand the types of uses being made of HCPP.

As with the EEBO data, we see a steady climb in publications over the last decade (although less steep than that shown in the EEBO data), and an overall increase in citations to these publications (although again, with a less marked increase apparent). We again see the effect of relatively long delays between publication and subsequent citation by other authors, with relatively few papers written in the past 2-3 years having reached higher levels of citations. However, the drop off in citations in later years is much less marked than in the EEBO data, which again can help us understand disciplinary differences. As we will see momentarily, there are more publications mentioning HCPP that fall within the social sciences compared to the EEBO publications, and the social sciences tend to be cited more than humanities publications.

Looking at the location of authors in the Scopus data, we can see a marked difference compared to EEBO: over half (52.8%) of HCPP authors in the data are based in the United Kingdom. For EEBO, about a quarter were based in the U.K., with over 40% in the U.S. With the HCPP data, we see many fewer American publications (11.9%), and a small number of publications from Commonwealth countries such as Australia (7.3%), Canada (6%) and New Zealand (1.8%).

HCPP Author Location, Scopus publications³⁸

Country of Author	N	% of total authors
United Kingdom	115	52.8%
United States	26	11.9%
Australia	16	7.3%
Canada	13	6.0%
Ireland	5	2.3%
New Zealand	4	1.8%
France	3	1.4%
Germany	3	1.4%
Total	446	100.0%

Turning to the detailed sources in publications, we see the results in the following table:

HCPP Publication Sources

Publication Source	N Pubs	Total Max Times Cited	Average Citations per Pub
Books (from Google Books)	43	1841	42.8
Journal of Legal History	7	13	1.9
Law and History Review	6	3	0.5
Victorian Literature and Culture	4	9	2.3
The International History Review	4	5	1.3
Journal of Social History	4	5	1.3
War in History	4	3	0.8
International Journal of Middle East Studies	4	2	0.5
Journal of British Studies	4	0	--
The Journal of Economic History	3	45	15.0
Geoforum	3	27	9.0
Journal of Southern African Studies	3	12	4.0
Journal of Economic History	3	9	3.0
Past and Present	3	8	2.7
Journal of Historical Sociology	3	6	2.0
The Historical Journal	3	5	1.7
Australian Historical Studies	3	3	1.0
Slavery & Abolition	3	3	1.0
Journal of Victorian Culture	3	3	1.0
The Journal of Imperial and Commonwealth History	3	3	1.0
Irish Studies Review	3	1	0.3
Explorations in Economic History	3	0	--
Sources with 2 publications	51	241	4.7
Sources with 1 publication	435	1191	2.7
Total³⁹	605	3438	3.7

While books were near the top of the table in the EEBO data, here books are the most common source of mentions of the HCPP by far, with 43 publications in the data, which in turn have been cited a strong 42.8 times per publication. As mentioned above, the journals on this list represent a somewhat broader range of publications than with EEBO, with many falling outside the humanities and more in the social sciences, or in journals that are a blend of social science and humanities (e.g. *Journal of Legal History*, and *Law and History Review*).

³⁸ Only countries with more than 1% of publications shown. 12 additional countries are also represented in the data.

³⁹ Totals are slightly less than the total number of publications overall due to missing data.

This is also reflected in the overlay map below, where we can see that the HCPP journals are spread more widely into the social sciences and even the sciences.

Source: Scopus data visualized with VosViewer, using overlay map data as described in Leydesdorff, et al. (2015).

This is consistent with what we have been seeing throughout the previous sections: that HCPP appears to be used less than EEBO overall, but that HCPP is used across a broader cross-section of disciplines and institutions, and is of particular import to British scholars. EEBO, on the other hand, appeals to a more focused community, but that community is spread more widely across the English-speaking world.

Analysis of Dissertations and Theses

One final area of exploration is how both EEBO and HCPP are used and mentioned in dissertations and theses. The ProQuest Dissertations and Theses Global database includes doctoral and masters level academic outputs. As mentioned in the beginning of this section, there is a bias in these data insofar as institutions in North America are more likely than British universities to require that students submit their completed post-graduate work to ProQuest for archiving; many British universities have their own institutional archives that are the preferred depository for theses. Also, the ProQuest database is more likely to contain doctoral theses,

again because more institutions require the deposit of doctoral work than do so for masters work (as shown in the donut charts embedded below). Even though relatively little of the data reported here is from the U.K. (EEBO n = 9 ; HCPP n = 12), the overall trends can be instructive to examine, as can the research areas of the theses.

In the two charts above, we see data that is consistent with all the other data reported above: both EEBO and HCPP show a marked upward trend in related publications over the last decade, the absolute number of results obtained by our EEBO search is higher than that obtained by our HCPP search, and the increase is steeper in the case of EEBO compared to HCPP.

Finally, we turn our attention to the subject classifications in the theses in our sample. Each thesis can have multiple subject classifications; in practice, most had from 2-4 classifications attached, which were extracted from the classification data reported by ProQuest.

The EEBO theses are, as one might expect, heavily weighted towards British and Irish literature, with 590 instances of theses classified in that category. European history is the next most frequent (n=196), followed by close to one hundred each on the topics of theatre, religious history, and theatre history.

HCPP theses, again consistent with all the other data reported above, are less dominated by a single classification, with European history in the top spot with 52 instances, followed by a smaller number of theses on modern history, history, law, and American history. We also see a pattern similar to that shown in the overlay maps above: topics extending well into the social sciences including political science, South Asian studies, economics, and women's studies, for instance.

Top 30 Subject Classifications in Theses

Rank	EEBO		HCPP	
	Classification	Count	Classification	Count
1	British and Irish literature	590	European history	52
2	European history	196	Modern history	17
3	Theater	103	History	14
4	Religious history	98	Law	14
5	Theater History	93	American history	13
6	American history	78	Canadian history	11
7	Women's studies	73	Middle Eastern history	11
8	Science history	55	Economic history	10
9	Comparative literature	50	African history	10
10	American literature	47	Latin American history	9
11	Modern literature	47	Political science	9
12	Medieval literature	44	World History	8
13	Music	36	British and Irish literature	8
14	Religion	36	Religious history	8
15	Gender studies	35	History	8
16	Philosophy	33	South Asian Studies	7
17	Theology	32	Economics	7
18	Rhetoric	31	Science history	6
19	Art history	31	International law	6
20	Romance literature	30	Caribbean Studies	6
21	Political science	24	Black history	5
22	Literature	23	Cultural anthropology	5
23	Biographies	22	Women's studies	5
24	American studies	20	Art history	5
25	Medieval history	20	International relations	4
26	Middle Ages	16	International Relations	4
27	Literature	16	Recreation	4
28	History	15	Military history	4
29	Law	15	Religion	4
30	Modern history	15	Native American studies	4
	Others	504	Others	106

Putting aside data anomalies, these data generally demonstrate the importance of EEBO and HCPP to new scholarship being produced as part of masters and doctoral work. This already successful area is one that should be encouraged, since today's post-graduate students are the faculty and researchers of tomorrow.

Qualitative Impacts

The quantitative results in the preceding section give one view of the impact of EEBO and HCPP, but to put these raw numbers in context, we also gathered qualitative information using interviews.

Interviews and Opinion-Gathering

Participants

In order to develop the findings of the quantitative study described above, interviews were conducted with key users of EEBO and HCPP. An interview guide (Appendix 1) was used with each of the participants, with variations to allow for more detailed analysis of certain areas and to distinguish between different types of usage. Eight interviews were commissioned for each resource, six academic users and two librarians or information professionals. Interviews were sought with users from different types of institutions, with a balance of Russell Group (RG) and non-Russell Group institutions, and we attempted to locate a variety of different post-holders in order to elicit understandings about the function of digital resources for academics at different career stages.

List of Interviewees

EEBO			HCPP		
Institution	Subject	Role	Institution	Subject	Role
Non-RG	English Literature/ Book History	Professor	Non-RG	Modern British Social/ Cultural History	Reader
Non-RG	English Literature	Senior Lecturer	Non-RG	Modern British History	Research Assistant
RG	English Literature	Part-time Lecturer	Non-RG	Modern British History	Senior Lecturer
RG	Italian	Senior Lecturer	Non-RG	History/Digital History	Senior Lecturer
RG	Digital Editing/ Digital Projects	Digital Editor	RG	Informatics	Research Fellow
RG	Research Services	Librarian/PhD student	Non-RG	Resource Management	Librarian
Non-RG	English Literature	Lecturer (Early Career)	RG	Library/Special Collections	Official Papers Librarian
Non-RG	English and Computational Linguistics	PhD student	RG	Modern British and Irish History	PhD student

Interviews were recorded and transcribed, and a draft of the report shared with all participants.

Key themes

Despite a small sample of interviews, a number of overlapping themes emerged during this phase of the research, as described below. As many of these themes were shared between the two digital resources, the following section is ordered thematically, and the two resources are largely discussed together.

Early adopters

Both EEBO and HCPP were in the vanguard of digitized resources for humanities scholarship and teaching, and both are well embedded in their target communities, as evidenced above in the quantitative analysis. This is borne out in several interesting ways in the qualitative work, where evidence emerged that information about both had spread quickly, largely through word of mouth, through colleagues and other scholarly networks⁴⁰. Many of the participants

⁴⁰ Despite the growing importance of digital tools to Humanities scholars, a 2010 study of Humanities Information Practices (Bulger, et al., 2011) showed that 95% of Humanities scholars considered peers and experts to be a key source of information about the discipline.

interviewed for this study recalled the appearance of EEBO and HCPP on the research scene, were early adopters of these resources, and recalled the impact of these digital collections on their research.

Some of those interviewed were using EEBO in the 'dial-up era' and recalled warnings about the fact that 'large text files might crash your browser' (IDC04). Access to these documents in digital format was completely different and highly prized:

'I can remember because I did an edition of a 1630s play, and I remember a friend of mine that was at the college going and sneakily doing a photocopy of the only printed version where later on I could just download it.' (IDC09)

Early adopters were also very astute about what the digitized resource could offer that the original documents, or the microfilms, could not: 'I think I twigged quite quickly what I could do with it' (IDC04). Others recall colleagues who were taking advantage of the resource before institutional subscriptions were introduced: 'I have colleagues who were frantically downloading as many page images as they could get because there was a, 'Gosh, wow, we have to use this,' even before it was institutionally subscribed'. (IDC08)

Users of HCPP also reported that the resource had quickly acquired a positive reputation:

'I think when I was doing my PhD and started off as a researcher I became aware of digital resources generally, and this is one of the larger and more useful ones which I think most people are aware that it's there and it's something that everyone can draw on really.' (IDC06)

Many of the academics interviewed remembered the last project they had done before the digital resource was released, and the impact of access to the digital resource on their next project: 'When I was doing my PhD it wasn't available, [HCPP] wasn't digitised, so I used to use Parliamentary papers in hard copy in stacks in Bodleian, and then immediately, as soon it came along, I began using it'. (IDC02) Those who could recall the early days of access to EEBO suggested that the rate of release of materials into the EEBO database resulted in difficulty being able to locate a watershed moment where their access to EEBO transformed their research. Participants described a slow and steady change, resulting in it becoming 'second nature' to use EEBO. (IDC08) This regular release of data (which continues, the latest updates to the collection are reported on the EEBO website to have been released in August 2015) leads to on-going serendipitous discoveries by some users and a feeling that new discoveries are always possible.

Centrality, access, funding

The extent to which EEBO, in particular, sits at the very heart of many scholars' research practices was consistently described. Users described EEBO as absolutely central to their research, even when their institutions held fine special collections containing many rare and early books. This was particularly true of research students, who appeared to use the digital resource to do their preparatory or initial work, and who were then more confident in seeking out additional material in libraries and archives. Participants stated 'I couldn't do my research without EEBO. In fact I wonder how much of my research has been shaped by having EEBO to be honest' (IDC12) Another EEBO user commented 'I just use EEBO all of the time... It's fantastic, it's marvellous!' (IDC08).

Users appeared to be very aware of the different versions of EEBO, and that the provision of different levels of product were related to difficult financial decisions made by their institutions. This awareness had often come from moving institutions, and finding that the new version of or interface for EEBO they were using was different. One early career academic had recently taken up a new position to find themselves hampered by the limitations of JISC Historical Texts, having previously had access to the EEBO-TCP content: 'I do find the search facility not great, not anywhere near as good as EEBO'. (IDC13) Another user found JISC Historical Texts 'a struggle', and 'frustrating, because to start with it only offers you to browse for a title... and I'm always looking for the advanced search button, and that's not immediately obvious'. (IDC11)

JISC Historical Texts was deemed good for teaching, however, its different functionality providing an accessible and useful introduction to EEBO for students: 'I mean Jisc is very, very useful, I use it in teaching loads... their first assignment is to go and look at something on Jisc and

talk about its paratext so it's incredibly useful.' This particular user had developed strategies both for their own research and for their students, which relied on the sensitivity of the search on the English Short Title Catalogue to find an accurate catalogue reference that could then be used to unlock the texts through JISC Historical Texts: 'frankly the ESTC search facility is very good; it just adds another stage into it I guess'. (IDC13)

Whilst most users of EEBO had a good sense of the different versions available, and the work of the EEBO-TCP project to enhance the collection, academics didn't always have access to decision-making about the funding of such digital collections, which could point to a lack of linking between university departments, university libraries and special collections, or a lack of understanding about the richness and the frequency with which these digital collections are engaged.

Ease of access/Keyword searching

Ease of access and the efficiency of keyword searching remains a huge factor in scholars' use of these digitised materials. One EEBO user gave an excellent example of the difference between searching through original special collections for a reference, compared to the affordances of the digital database:

'You can't be sitting in a library and whittling through the folio of Donne's sermons looking for the passage where Donne compares God to millions of elephants in size. With the full text creation partnership versions, it's instantly there. It's instantly finding your way to research material and teaching material that you can use.' (IDC08)

This casual use of EEBO was reported by several participants, one of whom discussed the fact that this type of research work, checking references and quotations for accuracy, would previously have not been considered: '20 years ago you would have thought that would be far too much effort because you could just rely on the secondary source you were working from'. (IDC04) EEBO is thus enabling the use of primary resources in these new contexts, creating accuracy and increased contact with the digital collection.

Users of HCPP reported similar gratitude for the timesaving that the digital resource afforded: '[It's] the usual story of saving a lot of time being able to access it whenever rather than having to go into the library'. The same participant reported the efficiency of keyword searching: 'Again, the keyword searching that's better than going through an index... if you're looking for a particular word or topic or person.' (IDC02) In particular, the instant access to digital materials ensured that research could often be more efficiently slotted in around myriad other institutional and other responsibilities: '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.' (IDC02)

Several of those interviewed for this project worked on a part-time basis, some combining research with other paid work, while others had developed flexible work models in order to balance family commitments. The ease of access and efficiency of the digital collections was particularly valuable to these interviewees, allowing for some research work to be amalgamated into these flexible arrangements. One HCPP user reported that the growth of digital research materials online had changed her life ('I can work from home!') and her approach to research:

'The Internet has transformed my life. I'm often out for meetings but pretty much all my research, unless it's archival – and I still like the archive, don't get me wrong, but I like nothing better than starting with Googling.' (IDC07)

This researcher had, by 'Googling' been able to find references to the Parliamentary Papers in Google Books, which were then further researched and discovered through HCPP.

Materiality

The original materials underpinning the EEBO collection lend themselves much more to the study of the material, than those forming the basis of HCPP, resulting in very different responses to the question of how participants use the originals in the era of digital access. Researchers of the material culture of the book can gain considerable understanding and additional research

data from looking at the originals. Digitized content has transformed researchers' behaviour around the original materials in a number of interesting ways. One participant described observing something anomalous in a printout from an EEBO page, and following this up in a research project, the focus of which was entirely on aspects of the materiality of the printed book: '[I] found some interesting examples [in EEBO], some of which have to be seen in the original to get the sense of their real impact'. (IDC04) One academic who described using text-mining tools to unlock the transcribed text, described how this led to the original books:

'What you lose in transcribed texts is all the layout and typography and the different fonts and that kind of thing. So although that's not my primary interest, I would... I always like to sort of, reconnect with the materiality of the original.' (IDC11)

The extent to which EEBO facilitates discovery through open access to collections and full-text searchability has had a particular impact on several of the scholars interviewed for this project:

'One of the things that EEBO does... is that it gives you in a way that it's very difficult, particularly in these days of rare book collections being behind cages and in storage, the breadth and diversity of the printing record'. (IDC04)

Several participants suggested that their research practices had changed as a result of better, searchable full-text access to EEBO, particularly in relation to the way in which they used the original materials. One of the scholars interviewed has a particular research focus on the history of the printed book, and tends to use EEBO alongside the originals on a regular basis, stating 'I feel like I know when to use EEBO and when not to':

'For [one author] I will use digital copies as a check but the primary work I will do will be with the physical copy because I need to. To the point that I might get quite annoyed with librarians to allow me to see the original copy, because there's a microfilm, on the grounds that I need to see specific copies and I need to be able to examine those copies and describe them.' (IDC04)

Another participant reported a similar experience of balancing *digital primary research*, so to speak, with traditional special collections work: 'I'm troubling the libraries less often, and I'm using original materials less, and using, in a curious way, *more* primary materials via EEBO.' (IDC08)

Serendipity and Sharing

Many of those interviewed discussed the fact that digital resources made it much easier to share research materials and findings with colleagues, leading to increased opportunities to discover new research pathways and materials. One participant found that, particularly in the early days, news of what was in the collection would spread quickly: 'The people who were researching the 17th century at the same time would tell each other, and we were all using it'. (IDC09) Another participant recalled being sent in a completely fresh research direction as a result of a tip-off from a colleague in another department about something unusual they had spotted in EEBO:

'A colleague at work said, 'Oh have you ever seen this before?' and he showed me an EEBO printout. I went and looked on EEBO and had never seen anything like that before, and then from that... I went and looked at the original, and then made that into a research project looking for unusual examples of printing.' (IDC04)

The same participant underlined the impact of EEBO on serendipitous research: '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 students: '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.' (IDC04)

This type of experimental searching has become a useful teaching tool:

'That's one of the ways I introduced [EEBO] to my students. I say, 'Search for your home town, search for your name, search for your favourite hobby.' Because then they get a real sense of it – because they don't know what they're going to get, and they could get anything, and I can get anything. I'll see things that I've never seen

before. But they get a sense of the real breadth of material, and some of them take that first encounter and then turn that into a research essay.' (IDC04)

It isn't just students, but seasoned scholars who are finding their way to new research opportunities through this kind of serendipitous searching. One participant described looking for some contextual references for an article, and stumbling across 'just marvellous' anecdotes and evidence: 'You find your way to most extraordinary things.' The same user described the process of discovery: 'you hear the name of a writer who's unfamiliar... and you look at what else is there and suddenly you understand what the writer is' (IDC08). Another user, primarily interviewed for their use of HCPP, told me about having used EEBO for a commissioned research project for a heritage organisation:

'The woman I was looking at was an 18th Century writer and it saved me a trip to [The British Library]. Straight in, I just needed a nice quote really. Plus for myself, I wanted to see what the children's books she wrote were like. So that was great... It was only one afternoon, but I was surprised. You know, rather than getting frustrated and giving up, it was good.' (IDC07)

Teaching context

Each and every participant with teaching responsibilities reported using EEBO and HCPP with their students, both at undergraduate and postgraduate levels. One of the observations of the TIDSR report was that digitised resources were encouraging the use of (digital) primary sources in the undergraduate seminar room, and this seems now to be very widespread (Meyer, et al., 2009). Participants described using EEBO to 'enthuse undergraduates and MA students' and to 'try to intrigue them and make them think that they can *find* something in a primary text' (IDC08). This then becomes an important pathway between the digital surrogate and the primary text. As the students become more familiar with the sort of materials they can find in EEBO, they build confidence and skills in using primary texts:

'To enthuse the students with [EEBO] and the books... we went to the library and spent an hour and a half looking at 20 early printed books from the collection laid out... and they just looked through the books and thought, 'This is the real stuff'. I think that is complementary. (IDC08)

Part of that inspiration is located within the idea that a digital collection may not just allow young researchers to access such rare and early materials, but that the affordances of the digital technology may encourage them to fashion and shape a new kind of scholarship:

'When I talk to the MA students about research methods I'm forever saying to them, 'Be a modern scholar.' There have been wonderful, wonderful scholars in the past, and more or less everything that can be said about canonical text has been said. Be a modern scholar, use these resources to find the archival text and the rest of it, and also using full text searches, suddenly one feels like a fake version of a great scholar. How does he know this stuff or has does she know this stuff? Well I don't quite know it but EEBO finds it for me. I'm telling people be like that and exploit these tools, and then we can create a new scholarship of this age.' (IDC08)

Students often find the real value of EEBO in their final year, while researching their dissertations. Indeed, EEBO was incorporated into teaching in one institution in a way that positively encouraged student engagement with (digital) original materials:

'We had to come up with, all of us who were teaching it, a couple of texts that were only available, obviously they're available in hard copies somewhere, but had no modern editions that students that draw on. So they had to go to EEBO to use it, and do something quite original from it... I think they got quite a lot out of it.' (IDC09)

Access to the digital resources contained within EEBO appears to have encouraged students to pursue doctoral research. One participant described having taught with EEBO at a university where the special collections were negligible, and having been able to inspire students to pursue careers doing research work on critical editions, working with special collections and in book conservation: 'those are students where you think, 'Wow' you have these extraordinary skills and if you had not taken this module you wouldn't necessarily be on the career path you're on

now'. The same academic pointed out that the majority of their students would go on to be 'regular citizens' but that the course they had taken, made possible through EEBO, would nonetheless have a significant impact: 'I said, 'As a result you now understand the printed heritage of Britain', and when it comes to things like special collections and going to exhibitions, your local library and all of those kinds of things, they value them'. (IDC04)

HCPP, of course, is quite a different resource, providing instead of rare books, a core understanding of the rich historical context of Britain, the former Colonies and the wider world, and this can also have a profound effect on students. Participants described using HCPP with undergraduate students as part of their core training around digital research skills and to encourage students to unlock relevant issues and themes through searching and retrieving information from digital archives:

'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 or use a digital resource before a seminar, and then we come back and it generates a discussion.' (IDC06)

The same participant described the impact of access to the digital resource as having transformed the kind of research output students' produce in their second and final years: 'Digital resources... [have] really greatly expanded the range of what they can do... for a dissertation or an independent essay'. The subject matter covered by the HCPP resource can be particularly useful and satisfying to undergraduate historians: 'I think a lot of students come to history with a very firm idea that it's all about politics. I don't agree with that idea, but actually that's what fires up their imagination.' (IDC06)

HCPP at the Open University

Quantitative analysis of the usage data detailed above showed that the Open University is a particularly heavy user of HCPP, and interviews with academics and librarians from OU provided a rich context to these statistics. The type of distance learning pioneered and practiced by the OU depends heavily on digitised resources, with course materials created in partnership between librarians and academic course leaders providing digital literacy training alongside access to resources such as HCPP. Academics described the usefulness of HCPP in OU teaching, 'because 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'. (IDC05) Students are typically closely guided in their use of HCPP, reflecting the particular pedagogical context of the OU:

[The course material] contains quite a big section there on how to use the Parliamentary Papers in terms of searching and browsing, but a lot of students go for the search, as they do! ...So it is a bit of traditional teaching combined with a bit of 'go out and find things for yourself'. (IDC05)

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. An OU Librarian interviewed for this study reflected on this combined skills delivery, and how well suited the HCPP collection is for supporting this:

'I sat with an academic in the history department and we drafted something together, and I went off and got it all on paper and she fed back to me, and then we produced a more finished item. So we both input ideas, her professional knowledge and my professional knowledge of the sorts of things that students struggle at that level, the type of skills, my knowledge of how they approach things, and we've put together activities in that way'. (IDC10)

Students using HCPP, both in the OU and elsewhere, are more likely to be encouraged to use the traditional, systematic approach to finding materials within the digital archive. Several participants mentioned the importance of teaching students the limits of HCPP: 'we try and teach them that the OCRs not very good, so they might miss quite a few key words just because the computer's translated them incorrectly'. (IDC02) Poor OCR comes up again, in relation to

caution urged on students using the resource: 'I know that they scanned a lot of documents from microfilm and there are all sorts of problems with OCR on those documents'. (IDC05) A non-OU specialist librarian responsible for Parliamentary Papers also commented on the occasionally poor OCR of the HCPP, suggesting that this is one reason to encourage students to consult the paper archives as well as the digital. (IDC14) This was not a recommendation delivered by the OU, however, where librarians couldn't recall receiving a single query about the paper archive.

Occasionally poor OCR and the slightly idiosyncratic nature of the content within HCPP has led many scholars to encourage traditional, painstaking research skills in their students:

'I think the way that Parliamentary Papers is set up, it's almost set up to force you to do things in a very traditional way in terms of going through the material. You really need to know what Select Committee or Command Paper you're looking for. Even to use the search I think you've got to have a really good understanding of what actually a parliamentary paper is.' (IDC05)

The OU's embedding of these 'huge primary sources that everyone's aware of' (IDC10) into their teaching programme has produced a template for delivering distance learning through digital collections from which other campus-based universities could learn. One potential downside of their dependence on these digital collections, however, is that platform or content changes can wreak havoc. An OU Librarian interviewed shortly before the House of Commons Parliamentary Papers site was refreshed confided 'if the search screens change. That's a major nuisance. For instance, HCPP's planning a change... and I'm dreading it because it works really well at the moment' (IDC10). Identifying these potential conflicts is clearly important for both content providers and institutions.

Digital citation

Citation of digitised materials has long been a source of uncertainty among scholars, particularly in relation to collections such as EEBO and HCPP, which are essentially digital surrogates of original materials. In the TIDSR report in 2009, we found that 'of the scholars who published results based on materials in the five collections, over one-third only cited the physical item represented in the digitised collection and made no reference to the digitised collection at all' (Meyer et al., 2009). In the EEBO-TCP benchmarking study of 2013, similar uncertainty existed around citation practices, leading to a string of recommendations for scholars to improve guidance in this area (Siefring & Meyer, 2013).

Participants in this study reported considerable, continued uncertainty in this area. When asked 'Do you have a good understanding of how to cite digital resources correctly?' the majority of participants said that they didn't, and that either proper guidance didn't exist or was conflicting. In many ways, some felt that things had become less clear over the years since digital resources had arrived on the scene:

'It's a bit of a tricky one actually. Previously when digital resources first turned up it was common practice to ask students to give a textual description of it and the URL and the date at which they accessed it. A couple of years ago I think that was standard, but I think that has started to fall out of fashion.' (IDC06)

Many reported using a 'fudge' of their own design, of which the below is a common example:

'Certainly in the past I have tended to make footnotes and make citations as if I'd had the original book in my hand, and then probably within the text that I'm writing, I might mention the EEBO page images of this book. So that my reader understands that I've been using EEBO for this.' (IDC08)

This problem was common across both EEBO and HCPP. One user of HCPP reported their own uncertainty about how to cite the resource properly, and the lack of certainty with which they felt they could offer a solution to students asking for guidance, despite very clearly feeling that it was an important factor to get right:

'So citing research is incredibly important, we try to teach it properly but you just get so massively confused with what digital content you do use traditional citation

for and what digital content needs to be cited that it's come from a digital [resource].' (IDC05)

When asked about their own habits, one EEBO user reported: 'I've changed my view on that... So initially my view was that you didn't need to cite EEBO because... as I said, 'EEBO is a library'. However, this user gave a very interesting view on why scholars may not be citing EEBO, and these reasons were rather different:

'I know there's been debate about this amongst early modern scholars about whether we should be more open about EEBO, and I think there is a little bit of embarrassment about it because people don't see it as proper library research that if you publish something and it just says, 'EEBO copy' after everything it would look a little – Really? Really? I think it's unfounded.' (IDC04)

The lack of citation not only provides difficulty for resource providers or academics looking for evidence of the impact of these resources, it also prevents an open scholarly dialogue about the changing research habits and practices created by these digital collections, as one academic pointed out: 'I've seen EEBO misused in that respect, and that's where I think people should be more upfront about their use because it shows the limits of what's possible to deduce from the evidence', adding 'given how important the resource is that strikes me as something of an oversight in terms of research diligence or diligence on the part of the researchers themselves' (IDC04).

Participants were very keen on resources providing their own recommended citation, but were quick to point out that there is very little consistency across resources. The lack of standards here were considered very problematic, and led to uneven practices among participants. Participants also reported that they had sometimes felt pressure from publishers to cite in a traditional way, removing long URLs: '[Publishers] really don't like a lot of those URLs in their footnotes. They hate them and they tell you to take them out' (IDC05). One participant reported that they were permitted URLs in the bibliography but not in the footnotes, where details of the location of archival materials are traditionally noted.

Digital Humanities approaches to EEBO and HCPP

So far, this section has largely focused how EEBO and HCPP are delivering traditional research, scholarship and learning. As the results detailed in our quantitative study showed, 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. EEBO has been much more open to scholars thanks to the EEBO-TCP partnership which 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. The marked-up texts and associated files were previously only available to users within the EEBO-TCP partnership, but on release to the public, users were free to copy, post, publish, distribute, and otherwise share. This openness has led to the development of a number of significant tools for analysing and viewing the EEBO-TCP collection, such as Lancaster University's Corpus Query Processor⁴¹, the searchable EEBO-TCP database provided by Brigham Young University⁴², and the EEBO N-gram viewer at Washington University in St Louis⁴³. These tools allow computational linguists to expose connections between texts and writers, and to produce highly innovative Digital Humanities research and teaching.

One of the researchers we interviewed for this project was conducting a PhD using these new techniques, and had developed a highly complex workflow between the different datasets, allowing for different kinds of data cleaning and treatment. This student had done significant work to scope out the various EEBO-TCP datasets and the affordances of these, to pursue a research project related to Shakespeare's language. The workflow could often end with the JISC

⁴¹ Corpus Query Processor, Lancaster University, <https://cqpweb.lancs.ac.uk> consulted 2nd February 2016

⁴² Early English Books Online Corpus, Brigham Young University, <http://corpus.byu.edu/eebo/> consulted 2nd February 2016

⁴³ EEBO N-gram Viewer, Early Modern Print, Washington University in St Louis, <http://earlyprint.wustl.edu/tooleebospellingbrowser.html>, consulted 2nd February 2016

Historical Texts images of the relevant publication, linking highly complex analytical processes to more traditional digital textual scholarship with the EEBO collection. This sort of Humanities Big Data approach is becoming increasingly attractive to scholars interested in experimenting with new techniques, and the open access nature of EEBO-TCP has encouraged a proliferation of experimental techniques and datasets giving the emerging scholars powerful and creative tools with which to build their own research careers. This sort of work also reflects the shift from the lone Humanities scholar, to a more collaborative, interdisciplinary approach comprising teams of researchers and technical experts.

The release of the EEBO-TCP data has had additional impacts on the growth of this kind of scholarly approach. The Partnership's Oxford team has maintained an active Twitter presence (@OxfordEEBOTCP) which two of our interviewees described as essential to the flow of information not just about the EEBO-TCP project, but about the potential for examining the collection through digital means. One participant described being persuaded by the relevance of Twitter for Academia through the Oxford EEBO-TCP account, saying 'it's because they're very present, so seeing what's going on there [is important]' combined with academic conferences and meetings, 'you end up having these conversations about the resources online, which is really great; it's all quite ad hoc'. (IDC13) When the EEBO-TCP data was released to the public early in 2015, the Oxford team organised a one-day 'hackathon', to encourage creative re-use of the data, and a more extended 'hackfest' to invite ideas for creating tools to access the collection for research or teaching. Only one of the EEBO interviewees had not heard of these events, or of work associated with them, and several expressed a desire to see such events repeated in the future. The EEBO-TCP project has delivered significant, innovative approaches to scholarship building on the EEBO dataset.

Despite not having similar data available to the public, HCPP has also had an impact on the world of Digital Humanities, when it 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 across huge datasets. This analytical approach is similar to the PhD research on EEBO described above, but with the effort spread across a team of researchers, each with their own specialist skills. The scale of what could be achieved was significant:

'Well, the big positive thing of the project was that beforehand people had studied commodities in isolation or just 2 or 3 together and they would have to go through documents, thousands of documents to draw one map and this tool, this interface now lets you look at so many different types of commodities'. (IDC03)

HCPP does not allow researchers to routinely export the XML, but if it chose to release some of this data in the future, it may see the kind of impact delivered by the release of EEBO-TCP data to the public, through widespread increased Digital Humanities inquiry and collaboration.

Conclusions

It is perhaps no surprise that the users interviewed here were overwhelmingly positive about EEBO and HCPP, since they were selected for their experience with and creativity in using the resources. The strength of their conviction, however, points to the extent to which these two resources have transformed research practices and created new research pathways for established academics, as well as providing such an inviting route to discovery for students and early career researchers, including new areas of research such as Digital Humanities.

Survey of Researchers

A survey of researchers was carried out in November and December 2015. The survey was distributed via email to a list of over 500 researchers who had participated in previous studies related to digital resource use, to a selection of email lists, and also advertised on the Jisc Historical Texts website. The survey received 243 valid responses. The survey was designed to only present relevant questions to each respondent based on their previous answers, so some of the data below has fewer than 243 responses, as noted in each table or chart.

One of the first questions in the survey was a screening question we have used in various forms in previous studies (Bulger, et al., 2011; Meyer, et al., 2009; Siefring & Meyer, 2013) that presents respondents with a variety of digital collections and asks which resources they use regularly or on occasion, or if they have heard of it without using it, or finally if they have not even heard of it. By embedding the two collections we are most interested in (EEBO and HCPP) in the list (and neither collection was mentioned by name in the email invitation or the welcome screen), we can get some sense of how these resources compare to others. It is important to note, however, that this is not a random sample of the population, and the distribution methods relied to a certain extent on purposive methods (including posting an invitation on the EEBO home page). Thus the results should be interpreted with care: higher use and awareness of any given resource should be seen as indicative, rather than a statistically significant relationship.

Use and Awareness of Selected Digital Resources

N=243

In the table above, we can see that 75% of respondents reported using EEBO either regularly (n=130) or on occasion (n=54). The number of people who reported using EEBO regularly was nearly equal to those reporting using Google Books, a finding which is very consistent with our earlier study of EEBO (Siefring & Meyer, 2013) where 69% of respondents to a similar survey used EEBO regularly and 67% used Google Books regularly.

The House of Commons Parliamentary Papers collection, which we are also interested in here, had less usage and awareness, with 39% (n=94) using it regularly or on occasion. This is consistent, however, with the data in the first sections of this report, which showed overall lower usage of HCPP than EEBO, which indicates that the surveyed population does not seem to

be inordinately skewed from the user population. Most respondents had heard of HCPP, however, even if they didn't use it personally; only 20% (n=49) were unaware of HCPP entirely. This pattern of high awareness and low regular use tends to suggest that HCPP has a somewhat concentrated audience of users.

Other important resources

Respondents were asked (via a free-text field) to identify other resources that they used regularly and considered important. The following lists all those identified in alphabetical order, with numbers indicating the number of responses for those with more than one response. You can see that the resources span a huge range (and indeed, one respondent simply put "huge numbers, too many to list"). Respondents who entered any resources (n=98) typically listed between 1-5 resources, with the majority who chose to answer this listing 1 or 2.

Other important resources identified (listed alphabetically, n=136)

- 18th Century Newspapers
- 19th Century British Newspapers
- Academic Search Premier
- Accessible Archives
- 2 Ancestry.com
- Archive.org
- Austlit
- Band camp
- 3 Bibliography of British and Irish History
- Biblioteca Cervantes
- BL 19th Century Newspapers
- BLMO: British Literary Manuscripts Online
- Bodleian luna
- British History Online
- 6 British Library Digitised Manuscripts
- British Museum website
- British Newspaper Archive
- 7 Broadside Ballad Archive
- 2 Burney Collection Newspapers
- BYU Corpora
- Calendar of State Papers Online
- Canadians.org
- 2 Cecil Papers
- Chronicling America
- 2 COPAC
- Cornell Witchcraft Collection
- 2 CQPweb
- DB.history.go.kr
- DBNL
- 2 DBpia
- 2 DEEP: A Database of Early English Playbooks
- Dictionary of the Scots Language
- Digital collections of rare books libraries, most often Beinecke and Folger
- Digital Quaker Collection
- Digital Renaissance Editions
- Discogs
- 2 Early Modern Letters Online
- Early Stuart Libels
- Ebrary
- EBSCO OneSearch
- e-codices
- Edina Digimap, for historical OS maps
- Electronic Enlightenment
- Embase
- e-rara.ch
- Ethnologue
- eTHoS
- Evans digital
- FamilySearch.org

- 2 Folger Digital Texts
Foreign Libraries with digital images of liturgical MSS
- 3 Gale NewsVault
Google Image
Groves dictionary of music and musicians including Oxford dictionary of music
Hartlib Papers
Harvard Library
HeinOnline
Historical Abstracts
Historical Thesaurus of English
House papers at Yale
ISTC (Incunabula Short Title Catalogue)
Iter
- 3 Japanese National Diet Library Digital Collection
JISC Media Hub
Kindai Digital Library
Kiss
- 2 Korea University Japanese Archive search
Korean History Database
Korean History Online
KRpia
KVK (Karlsruhe Virtual Catalog)
- 2 Lexicons of Early Modern English
- 2 Library catalogs including my university's
Library catalogue of Buffalo University
Library catalogues of archival holdings (often incomplete)
Library catalogues such as the Vatican Libraries or the Marciana's
Library of Congress (USA)
Library sites involving urban history, such as the Mitchell in Glasgow and the Free Library in Philadelphia.
- 2 London Lives
Luna commons
- 2 Making of America (MOA)
Making of the Modern World (the Goldsmiths-Kress Collection)
- 2 MEMSO
MGH Online
Middle English Dictionary
Monumenta Germaniae Historica
Munich Digitization Center (MDZ)
Museum collections websites: Victoria & Albert Museum, British Museum, Royal Collection, New York Public Library
National Archives - Discovery search engine
National Central Library (Korea)
National Library of Ireland online manuscripts
- 2 National Register of Archives
Naver News Library
NetLibrary
Newberry Library
- 4 Newspaper Archive
Newspapers.Com
- 2 Nexis
- 4 ODNB
- 6 OED
- 6 Old Bailey Online
Ovid
- 5 Oxford Dictionary of National Biography
- 3 Oxford English Dictionary
- 2 Perdita Manuscripts
Perseus digital library
Peter Beal's catalog of manuscripts
Post-Reformation Digital Library
- 6 Project Muse
PubMed
Pubmed
purl.pt
Questia

	Record label websites and catalogues
	RED
	Renascence Editions/Luminarium
	Rock's Backpages
	Royal Society website
	Schoenberg Database of Manuscripts, Digital Scriptorium
	SCRAN for Scottish historical images
	ShelleyGodwinArchive.org
	Soundcloud
	Specialist music websites
6	State Papers Online
	The Clergy Database
2	Times Digital Archive
	Trove
	Union First Line Index of English Verse (hosted by Folger Shakespeare Library)
2	Universal Short Title Catalogue (USTC)
2	VD16
	VD17
3	Web of Science
	Wiki Source
	Wikimedia Commons
2	WorldCat
	Zhongguo jiben guji ku

One respondent also noted “As someone who is not a member of an academic institution I've found in the past that I often can't access resources I'd like to – e.g. EEBO – so I had to ask a friend to search for me for particular pamphlets I needed from EEBO.”

We also asked the people answering the survey “How important are the following to you for finding primary sources that you think are or may be available in digital form? By primary sources, we are referring to things like original documents (such as historical books, diaries, speeches, laws, manuscripts, letters, interviews, news items, official records) and creative works (such as poetry, drama, novels, music, art).”

Important methods for discovering primary sources

N=238

Specific databases or collections (which are obviously the focus of this report) are seen as important to our respondents. 97% (n=231) of respondents indicated that specific databases or collections such as those discussed in the previous section were important for finding primary sources that they know or suspect might be available in digital form.⁴⁴ References in journal articles (92% important) and references in books (90%) were also considered important, underscoring the importance of referencing material in a way that allows the reader to discover that they can access primary sources digitally. Searching a library catalogue was important to 91% of respondents and next most important, again unsurprisingly, was Google or other search engines, which were considered important by 88% of respondents.

In the middle range, we see most of the face-to-face ways of finding primary sources: contacting known experts (63%), asking people I know (61%), attending conferences (59%), and consulting librarians or other information professionals (53%).

Finally, the least important methods to those who answered this survey are digital social media tools: social network sites such as Facebook or Twitter (28%) and email lists (34%) are at the bottom with Wikipedia (42%) and publisher information (23%).

Median rank-order importance of different methods for discovering primary sources

N=238

In addition to choosing whether a method was important for finding primary sources that are or may be available in digital form, respondents were asked to place those items that they considered important in rank order. This ranking ordering allows us to ask “For those items considered important, which are ranked as most important?” These results are shown above.

We can see that specific databases or collections stand out even more as the most important method for finding new primary sources, with an average importance ranking of 3.33 (where 1 is most important). The next group are similar to those reported as most important generally, but the order changes slightly: now Google and other search engines are in second spot (4.63), followed by JSTOR/other databases (4.79), library catalogues (4.84), and then references in books (5.04) and journal articles (5.14).

We also see a shift with regard to social network sites, which now appear in the middle of the pack (7.5) along with web links (7.44) and Wikipedia (7.79). How is this to be interpreted, in light of the previous table? The most obvious interpretation is that while relatively few people in

⁴⁴ Note that this question came before the questions about specific resources in the actual survey, to prevent priming respondents to give this answer.

this survey report using social digital methods for finding new primary sources, *for those who do use them* they are relatively important. This then begs the question: are the respondents who use these digital social tools qualitatively different from those who don't, or are they early adopters, and if others were to use these methods would find them similarly important? Answering this question is beyond the scope of this study, but it suggests some interesting follow-up work that could explore this question in more detail.

EEBO and HCPP in detail

Respondents who indicated that they used EEBO either regularly or on occasion were given a subset of questions relating specifically to EEBO. In total, 176 respondents answered these more detailed questions about EEBO.

Similarly, respondents who indicated that they used HCPP either regularly or on occasion were given a subset of questions relating specifically to HCPP. In total, 77 respondents answered these more detailed questions about HCPP.

Attitudes toward EEBO (sorted by mean overall importance)

N=176

Asking EEBO users the strength of their agreement on the above statements underlines previous work we have done with EEBO: that EEBO is considered important to research generally (87% agreement), to one's individual research (82%), and makes new research possible (92%).

There is slightly less strong agreement that EEBO is reliable (38% strong agreement, although still with about 91% agreement overall), is easy to find (40% strong agreement), and easy to use (28% strong agreement). Interestingly, however, even with this somewhat less strong agreement that EEBO is easy, there is relatively little evident feeling that more training or information is needed in how to use EEBO, with only 39% agreeing with that statement.

Attitudes toward HCPP (sorted by mean overall importance)

N=77

The attitude data about the House of Commons Parliamentary Papers is somewhat less clear-cut than we saw above with EEBO. Across all categories, there is much less strong agreement than we saw in the case of EEBO. As with EEBO, finding the resource important to one's field or discipline is important, but whereas 70% of EEBO users strongly agreed with this, only 28% of HCPP users indicated strong agreement (and 58% indicated some level of agreement). While researchers consider HCPP reliable (85% agreement), only 57% agree that it is easy to use, and a modest 43% agree that they have recommended it to others. These data certainly indicate that there is room for improvement with regard to the HCPP resource.

One of the questions that arose during our previous research on EEBO (Siefring & Meyer, 2013) was that a relatively high proportion of EEBO users (35%) didn't know how they accessed the resource, which is available via multiple interfaces. This confusion seems to have lessened somewhat, as seen in the following data, although it is possible that the generic answer 'at a library' accounts for some of the same people who answered the slightly differently worded question in the previous study as 'don't know'. One of the text responses to 'other', for instance, included:

I believe I access EEBO through my university library - but realise I'm not sure about that.

Others responded with names of their specific libraries, or mentioned their library authentication system (e.g. *via Athens*, or *with SSO* [Single Sign On]). A few indicated they access it via Chadwyck Healey, and several indicated that they access EEBO from several of the options listed in the chart, depending on their needs at the time.

Method for Accessing EEBO

N=176

These data highlight the importance of library subscriptions for continuing EEBO access.

Method for Accessing HCPP

N=74

Compared to EEBO, an even greater percentage of HCPP users (62%) report accessing the resource through their university library. There is relatively little awareness of ProQuest (11%) or Jisc (8%) as providers of access, even when the library access may be the result of agreements with these organizations.

Reasons for using EEBO

N=176

Reasons for using HCPP

N=74

Looking at the reasons EEBO and HCPP users give for using each resource, we see some similarities as well as some differences. HCPP is most often used by our respondents as a reference resource (81%), which is also the second most common use of EEBO (71%). While the next most common use of HCPP is consulting scans/images of documents (49%), this is by far the most common use of EEBO (84%).

At the other end of the scale, few respondents use EEBO (13%) or HCPP (4%) as a source of data for computational research.

Citing Digital Sources

An issue that always arises in trying to determine academic uses of digital resources is that those uses are difficult to track if citations to materials contained in a collection do not indicate in any way the origin of the material. Users of both resources were asked if they had ever written something using materials in the collection (EEBO or HCPP); 80% of EEBO respondents

(n=141) and 38% of HCPP respondents (n=31) indicated that they had written something based on these resources. They were then asked:

When you wrote the item(s) above, would you say that you indicated (either with a link or a citation) that the material had come from EEBO/HCPP?

N=140

N=31

The trend toward indicating that materials have come from a digital resource seems to be increasing slightly in recent years. When we asked the same question of EEBO users in 2013, 8% reporting citing the online version of EEBO only, and a further 45% cited the print version plus a URL. In these data, 66% of EEBO-using authors and 58% of HCPP-using authors somehow indicate that they have used the digital resource.

That being said, the resistance to the inclusion of digital resources that have been noted previously (Blaney, 2014) comes through very clearly in some of the comments on this question.

I usually consult an edition of the original text, having sourced it using EEBO.

Anyone wanting to follow my reference would know to look in EEBO for such material.

I also assumed that pretty much everyone knew that EEBO was the place to go for such resources. I wouldn't cite my university library in a footnote, even if I used a book that it owns.

I tend to acknowledge use of digital substitutes when relevant to my arguments; when not, I tend to acknowledge use of databases as I do libraries, in acknowledgements.

I used EEBO to find a source and cited the source but not how I got there. Your question now makes me feel worried actually that I have done this wrong as of course, it would advertise the collections brilliantly if the way of finding sources was flagged up more!

I try whenever possible to consult multiple editions of Renaissance books, including both EEBO versions and hard copies in libraries. I only cite EEBO when I am citing the results of a search or when there is a relevant difference between editions. Sorry--it's not my job to advertise your service for you.

I can't be bothered indicating which sources I consulted in print and which in electrons. For many texts, I've done both. What then?

Short pieces without bibliographies.

I usually tell others of EEBO as a valuable resource, but don't cite it formally.

I consider EEBO a surrogate rather than a real source. The source is the original publication.

The referencing can get very wordy-unwieldy

The essay in which I used EEBO texts was published a good 10 years ago; at that time it was

normal to use the same format as when using a printed source.

In conference papers which are presented orally I did not include citations period.

You don't really "cite" things in conference papers like that.

In an orally delivered conference paper, citations aren't seen by the audience.

Permanent URLs require logging in to see the source so are more frustrating than useful for non-subscribers, while subscribers need an institution-specific link.

Links to subscription-only resources are useless for those who don't have access.

Because it will not be accessible to a non-academic audience who don't have access to EEBO.

Publishers require the traditional method of reference to Parliamentary Papers. And anyway, the electronic resource is a mere copy of the actual volume I am referencing.

Gets unwieldy: I usually prefer to mention, upfront, that I used the database.

I don't think it is good practice BECAUSE URLs are unstable and because proxies have never been cited. I used to read microfilms of paper originals all the time and no-one suggested (rightly) that I should be referencing the microfilm - which would now be completely useless.

Without wading into this debate in this report, clearly there are strongly opinions on the part of the minority who choose not to include links or citations to digital resources.

Improvements

Users were also asked about potential improvements to EEBO and HCPP. As with other questions, there were more open-ended responses from EEBO users.

- The amount of space given over to the page image is too little in relation to the total screen. This is very very irritating when the page won't fit on the amount of screen allotted to it, forcing me to scroll down to read a single page.
- Better filters, so that more specific results can be brought up.
- The quality of some of the images is rather poor and when teaching students with sight impairments I have found it difficult to feel confident that they are not disadvantaged when using EEBO. The magnification is often very good, but it depends a great deal on the text. I agree it is important to have good quality copies of the originals, but perhaps there is also room for 'enhanced' originals?
- Commitment to ensure access particularly for scholars who cannot gain access through their institutions; clearly indicate level of access (e.g., inc/excl TCP) on each page, especially search results
- More search options: specifically, A-numbers from the TCP system would be amazing - but also wildcard searches and proximity searches
- Sort by DATE (chronologically) or AUTHOR (alphabetically)
- More full texts. Ideally colour scans of the documents, although I appreciate the enormity of this undertaking.
- Make it clearer/easier to download a whole text as pdf. This is something my students struggle with.
- List of duplicates/editions would be helpful, esp. for EEBO-TCP.
- Continue to expand TCP provision. Go over old images and improve them for quality and adjust transcriptions accordingly (crowd-sourced?).
- Not sure why you have this question given that it is now available through HT and hence there is a different platform for accessing it.
- More full-text records, better quality images.
- The JISC interface is less useful and user-friendly than ProQuest.
- Keep adding to it. More accuracy in transcriptions.
- User interface is extremely clunky for first-time users. No problem for people like me who have used it from the early days, but it urgently needs updating for younger students/researchers.
- Would benefit from a more streamlined way of downloading PDFs of documents, as well as easier ways of searching for/browsing illustrations (e.g. the ability to have a results page that shows thumbnails of illustrations rather than just title pages)
- It would be good to have searchable transcriptions of all first editions of texts; sometimes EEBO has transcribed a later edition but not the first.
- It cannot process some useful searches, e.g. (X OR Y) NEAR Z
- To have the entire EEBO corpus transcribed by TCP, rather than just a sub-set.
- More TCP! Better OCR!

- More texts
- Links to author biogs
- More provenance info
- Eliminate transcription errors (there are many)
- Justification/explanation of author/genre - many do not reflect the content of the actual text and there are lots of texts with an incorrect author listed
- Make it cheaper for small universities to afford
- Is it possible to digitally improve the quality of thumbnails?
- Get it to accept alternative forms of names
- More searchable full-text transcriptions would always be nice!
- Number individual page links using the document's pagination, not just the image numbers
- Make multiple-page PDF downloads easier
- Improve scans of hard-to-read pages
- Please allow this resource to be readily available for researchers and not be commercialized.
- Making the KWIC preview a little bit bigger
- The new JISC interface is poor, and doesn't find books that I know are there (having consulted them in the past)
- The Bibles are a mess. Some are listed under the translator, some under the publisher, some under a name picked out of the phone book; and since there are thousands of items that come up on a title or subject search for "Bible" or "Holy Scriptures," one can never be sure one has found all the EEBO editions of, let us say, the Geneva Bible, or the King James.
- New and better scans of documents
- Downloading either the transcription or the entire set of images is not intuitive. The search algorithms are also imperfect. It often takes several tries to find something I know is in the collection.
- It would be helpful if the search facility could find text in images as well as plain text.
- The message that pops up when you try to view a full-text transcription that's over about 500kb saying that "such a large download may crash your browser, especially over a modem" needs to go! It's not appropriate to the age of broadband and it has made my undergrad students give up using EEBO in the past as they find the "Warning" message off-putting.
- Several of the pages link to spam, or pictures of (what appears to be) a JPEG of a Shepherd's Crown.
- I have been experiencing significant problems with the website freezing, failing to load, etc, especially on Google Chrome
- Improve classifications, ie not all poetry is classed under the heading Poetry
- Difficult to navigate thumbnails, should be possible to type in page number in original manuscript and have EEBO direct me there.
- Search engine is not very responsive
- Right now, as I'm answering this, EEBO does not work properly.
- It's not working today or I can't make it work, which is why I'm doing this quiz.
- Speed.
- Variant spellings and variant forms searches miss a lot--those algorithms need improvement. Browse by title and by year and location would be good. Transcripts are pretty spotty, but I mostly don't use them except for searching.
- It would be helpful to be able to scroll continuously through an EEBO text rather than have to load page by page
- Bibliographical information should match that on the ESTC.
- Full text transcriptions has lines going through words, an inconvenience.
- Make the option to download full PDFs more obvious to casual user.
- Improve UI and reliability of servers.
- Provide more information on the range of results, other than no. of 'hits'. E.g. visualise the date ranges of results?
- Make items with marginalia or other notable paratextual evidence more easily searchable."
- It would be very helpful to be able to download entire documents not just page by page (or to show how to do it if it is already possible)
- Full-text transcriptions should be improved.
- Better images!
- Make it less 'clunky'.
- It would be brilliant if the images were colour or at least grayscale instead of black and white - but I know this would mean a huge amount of work. Hypertext cross-references to other databases (and if possible, research papers related to the item in question) would also be very helpful. Finally, a Google-type 'search by image' for the illustrations and woodblock initials would be just brilliant!
- Its a great search tool- the search engine is really fast...
- It is very 'clunky' to use
- An easier way of navigating through the scans for multi volume or large works would be useful, but I don't know how that could be achieved. The full-text facility is very helpful where it is available.
- The search facility is NOT good - it often takes several different searches to find a ms I know is there. Key word is more reliable than Title which is bizarre. The whole website is slow and clunky - and always returns documents in a small (unreadable) size, so one is constantly having to enlarge each page, which is slow and irritating. Searching for periodicals is straightforward enough, but again slow and clunky if you know exactly what you are looking for - still requiring several 'sifting' steps. It's also not entirely MAC-compatible, especially the 'marked list' facility, which is essential. This is a real problem. But it's a wonderful resource and essential for my research and teaching.

- Clearer indications about what each subscription includes; better meta-data.
- Collations and copy specific info are often inaccurate
- The ProQuest interface is old, crude and ugly.
- The cost for MARC records for this collection was astronomically prohibitive. My library has this collection but the usage is low because we only have short records for these titles. If we were able to purchase the MARC records at a reasonable price, these might become more useful/important to our collection.
- The TEI encoded version is essential. All the titles in EEBO should be encoded. The scans often leave much to be desired, having been made from microfilm. There should be a way of replacing them with better scans.
- The interface for moving to next page images can be slow. A format like Google Books is much better. It uses natural scrolling.
- The main problem for people at small school is access. I must drive to a (luckily, near-by) research university to use this resource.
- It would be beneficial to be able to move from the search result to the actual scans of the document without additional steps (eg: the thumbnails screen).
- It would also be very beneficial to have the downloading process (for scans of full documents) work more efficiently/quicker (again, for example, avoiding extra screens to have to respond to in the process of downloading).
- EEBO and most other databases need to be more easily accessed by independent researchers!
- I would use EEBO more often if my university subscribed
- Augment and sophisticate the full-text transcriptions and search functions. Make that second tranche of EEBO available to all subscribing institutions who had bought the first.
- Prioritize keying first editions.
- It would be helpful if the record for each item included some indication of the item's size and/or format. It's impossible to see from a scan whether a text is in quarto, octavo, etc., and that information is important both for my research and for my teaching.
- It took me forever to figure out that I can only download whole texts if I first mark the items and then go to my marked list. That is stupid. ECCO doesn't do that. That's why ECCO is easier to use. With ECCO, I can click an item and download it right from that screen.
- EEBO would greatly benefit from having a function which would allow users to report errors. Because of the absence of such an option, I have in the past tried to report errors through the 'help' option, but I have not received replies to my submissions and I have not had my suggestions acted on. I suggest something resembling the 'Found an error?' option which appears on the National Archives' 'Discovery' catalogue.
- I suggest introducing a means by which users can suggest works which they would like to have prioritised for transcription through the Text Creation Partnership.
- I also wonder about the possibility of providing shelf marks for the particular copies from which the image sets are taken (i.e. the copies which were microfilmed), although this might be a bit labour-intensive (and it might be felt that the sort of user who wants this information also tends to be fairly adept at finding it out independently).
- Please continue to work on improving image quality.
- EEBO could stand with some categorization to help narrow down the type of document a researcher is looking for (e.g., plays versus pamphlets, recipes, conduct books, etc.)
- In the case of the old EEBO none' in the case of the new one lots of improvements needed. For example entering a work by the STC number (which is THE one we all use) does not work. Other problem= why have all these in one catalogue from 1475 to 1900? Not pleased at all with this new thing
- the quality of images isn't always very good cf archive, biodiversity library etc.

- New interface imminent, so that may make life easier! Some of the images are a bit rubbish - poor scans from microfiche.
- The navigation and finding tools need to be improved. Sometimes I know there is material there that I cannot easily find
- Integration with rest of Proquest platform ; availability through EBSCO
- Being able to download an OCR version of a document would be incredibly helpful. As things stand it's often difficult if not impossible to find chance references from a downloaded document without reading the whole thing.
- The access to page images is very slow and clunky. A format like Google Books would be much better.
- Make it more widely available for limited use.
- There are a few too many steps to download documents. Also the search engine sometimes seems to return items in odd orders (and not always the same order each time with identical searches? though this is hard to tell because the order is so odd). But generally I find it a helpful and invaluable tool.

Demographics

Participants were asked a short series of basic demographic questions.

N=309

The plurality of respondents to the survey are faculty members (38.5%), with significant numbers of students (16.7% graduate / 9.9% undergraduate) and library staff (15.1%).

N=309

A majority of respondents are from England (54%), with a significant percentage from outside Europe (31%).

We also asked about the specific kinds of institutional affiliations. As the following table shows, a majority (52%) are affiliated with UK Higher Education (HE) institutions.

N=292

Respondents were asked “Please indicate which of the subject areas below you most closely identify with, in terms of your research, teaching, or learning. (Select all that apply.)” Their answers are in the following table:

Disciplines

N=228

Key Findings & Conclusions

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 3 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 4 years, underscoring the fact that digital collections take time to become embedded in research.

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, HCPP 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 7 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.

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 nudge⁴⁵ 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 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 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. HCPP 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,

⁴⁵ See Thaler and Sunstein (2008) for more information on the idea of using relatively simple ‘nudges’ to influence human choices and actions.

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 EBO-TCP partnership. EBO-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. EBO-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.

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.

References Cited

- Blaney, Jonathan. (2014). The Problem of Citation in the Digital Humanities. In *Proceedings of the Digital Humanities Congress 2012*, eds. Mills. Clare, Michael Pidd and Esther Ward. HRI Online Publications.
- Bulger, Monica, Meyer, Eric T., de la Flor, Grace, Terras, Melissa, Wyatt, Sally, Jirotko, Marina, et al. (2011). *Reinventing Research? Information practices in the humanities*. London: Research Information Network. Retrieved from <http://www.rin.ac.uk/humanities-case-studies>.
- Davis, Philip M. (2013). *Journal Usage Half-Life*. Ithaca, NY: Association of American Publishers. Retrieved from <http://www.publishers.org/attachments/docs/journalusagehalf-life.pdf>.
- Leydesdorff, Loet, Moya-Anegón, Félix de, and Guerrero-Bote, Vicente P. (2015). Journal Maps, Interactive Overlays, and the Measurement of Interdisciplinarity on the Basis of Scopus Data (1996-2012). *Journal of the Association for Information Science and Technology*, 66(5), 1001-1016.
- Leydesdorff, Loet, Rafols, Ismael, and Chen, Chaomei. (2013). Interactive Overlays of Journals and the Measurement of Interdisciplinarity on the basis of Aggregated Journal-Journal Citations. *Journal of the American Society of Information Science & Technology*, 64(12), 2573-2586.
- Meyer, Eric T. (2011). *Splashes and Ripples: Synthesizing the Evidence on the Impact of Digital Resources*. London: JISC. Retrieved from <http://ssrn.com/abstract=1846535>.
- Meyer, Eric T., Eccles, Kathryn, Thelwall, Michael, and Madsen, Christine (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)*. Oxford: Oxford Internet Institute. Retrieved from http://microsites.oii.ox.ac.uk/tidsr/system/files/TIDSR_FinalReport_20July2009.pdf.
- Research Fortnight. (2014). REF 2014: Results and preliminary analysis Retrieved from <http://exquisitelife.researchresearch.com/exquisite-life/2014/12/ref-2014-full-results-and-analysis.html>
- Siefring, Judith and Meyer, Eric T. (2013). *Sustaining the EEBO-TCP Corpus in Transition: Report on the TIDSR Benchmarking Study*. London: JISC. Retrieved from <http://ssrn.com/abstract=2236202>.
- Tang, Rong. (2008). Citation characteristics and intellectual acceptance of scholarly monographs. *College & Research Libraries*, 69(4), 356-369.
- Thaler, Richard H. and Sunstein, Cass R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press
- Wiberley, Stephen E, Jr. (2003). A methodological approach to developing bibliometric models of types of humanities scholarship. *The Library Quarterly*, 73(2), 121-159.

Appendix 1: Interview Guides

EEBO Interview Guide

Introduction

Can you describe your current role and what it involves?

Finding and using the resource

How did you find out about EEBO?

- Do you think EEBO is an easy resource to find?
- How do you navigate to EEBO?
- Which version of EEBO do you use (ProQuest/EEBO-TCP/JISC Historical Texts)? Any particular reason/Aware of differences?

How do you tend to use EEBO?

- Do you find it easy to use?
- What features do you particularly like/dislike?

Did you know about the original (analogue) resource? Have you used this?

- Have you used the original materials more/less/the same since you began using the digital version?

Have you passed on the details of this resource to anyone else?

- Colleagues/Students?

Research

What is your area of research?

Can you describe your current research topic?

How easy would it have been to conduct this research without EEBO?

- How else could you have accessed the information you need?
- Will you also use the analogue version in your research?

Have you used any other digital resources in your current research?

- If yes, which?
- If not, why not?

What do you think digital resources offer that's different to the original materials?

- Specifically, have they allowed you to do anything with your scholarship you could not have done otherwise?
- How important is the efficiency offered by digital tools such as EEBO?

Citation habits:

Have you completed a piece of work (published or otherwise) based on your use of EEBO?

How did you cite EEBO?

- Are there some digital resources that you routinely cite?
- Do you have a good understanding/adequate support to cite digital resources correctly?

Teaching

What courses do you regularly teach?

Do you use digital resources in your teaching?

- If yes, which?
- If not, why not?

How do you find suitable digital resources?

(How) do you think students benefit from using digital resources?

Do you think there are any disadvantages in using digital resources with students?

Feedback

Do you have any feedback based on your experience of the site?

Can you think of any improvements that you'd like to see implemented?

Any other comments?

HCPP Interview Guide

Questions for users

Introduction

Can you describe your current role and what it involves?

Finding and using the resource

How did you find out about HCPP?

- Do you think HCPP is an easy resource to find?
- How do you navigate to HCPP?

How do you tend to use HCPP?

- Do you find it easy to use?

- What features do you particularly like/dislike?

Did you know about the original (analogue) resource? Have you used this?

- Have you used the original more/less/the same since you began using the digital version?

Have you passed on the details of this resource to anyone else?

- Colleagues/Students?

Research

What is your area of research?

Can you describe your current research topic?

How easy would it have been to conduct this research without HCPP?

- How else could you have accessed the information you need?
- Will you also use the analogue version in your research?

Have you used any other digital resources in your current research?

- If yes, which?
- If not, why not?

What do you think digital resources offer that's different to the original materials?

- Specifically, have they allowed you to do anything with your scholarship you could not have done otherwise?
- How important is the efficiency offered by digital tools such as HCPP?

Citation habits:

Have you completed a piece of work (published or otherwise) based on your use of this resource?

How did you cite the resource?

- Are there some digital resources that you routinely cite?
- Do you have a good understanding/adequate support to cite digital resources correctly?

Teaching

What courses do you regularly teach?

Do you use digital resources in your teaching?

- If yes, which?
- If not, why not?

How do you find suitable digital resources?

(How) do you think students benefit from using digital resources?

Do you think there are any disadvantages in using digital resources with students?

Feedback

Do you have any feedback based on your experience of the site?

Can you think of any improvements that you'd like to see implemented?

HCPP only

Have you heard about the potential House of Lords content, and Parliamentary Hub?

Do you think this would be valuable to your research and/or teaching? In what ways?

Any other comments?

The Impacts of Digital Collections: Early English Books Online & House of Commons Parliamentary Papers
Eric T. Meyer & Kathryn Eccles
London: Jisc
2016