

Sustaining the EEBO-TCP Corpus in Transition

Report on the TIDSR Benchmarking Study

Judith Siefring, Bodleian Libraries, University of Oxford
Eric T. Meyer, Oxford Internet Institute, University of Oxford
01 March 2013



JISC

Sustaining the EEBO-TCP Corpus in Transition: Report on the TIDSR Benchmarking Study

This report was funded by JISC, and is an output of the Bodleian Libraries (<http://www.bodleian.ox.ac.uk/>) 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:

Judith Siefring
EEBO-TCP
Bodleian Digital Library Systems and Services (BDLSS)
Osney One Building, Osney Mead, Oxford, OX2 0EW, United Kingdom
Tel: +44 (0) 1865 280042
Email: judith.siefring@bodleian.ox.ac.uk

Dr. 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 this report as:

Siefring, J. & Meyer, E.T. (2013). *Sustaining the EEBO-TCP Corpus in Transition: Report on the TIDSR Benchmarking Study*. London: JISC.

Available online: <http://ssrn.com/abstract=2236202>



Table of contents

Acknowledgements	4
Acronyms & Abbreviations	4
Executive Summary	5
Introduction	6
Context of the Study	6
Research design & methods.....	6
Quantitative Impacts	7
Survey of Researchers	7
Analytics	27
Bibliometrics.....	29
Web 2.0 Impacts	32
<i>Twitter</i>	32
<i>Google Blog Search</i>	34
Qualitative Impacts	39
Focus Groups.....	39
<i>SECT Workshop</i>	39
<i>Digital Humanities Summer School focus group</i>	39
<i>Digital Citation Focus Group</i>	39
Interviews and Opinion-Gathering.....	42
<i>Librarians</i>	42
<i>Encoding Experts</i>	43
<i>Projects</i>	43
<i>EEBO-TCP Editors</i>	44
User Feedback.....	47
<i>Conference</i>	47
Conclusions	49
Appendix	50
List of projects based on or related to EEBO-TCP	50

Acknowledgements

The authors wish to thank all the participants in this research, some of whom are named in the report and some of whom remain anonymous, but all of whom have helped us to better understand how Early English Books Online and the Text Creation Partnership are having impacts. Particular thanks go to Jonathan Blaney, Simon Charles, Amanda Flynn, Colm MacCrossan, Michael Popham, Rebecca Welzenbach and Pip Willcox.

Finally, thanks to JISC for funding our research.

Acronyms & Abbreviations

BHO: British History Online

CERL: Consortium of European Research Libraries

EBBA: English Broadside Ballad Archive

ECCO: Eighteenth-Century Collections Online

EEBO: Early English Books Online

EEBO-TCP: Early English Books Online Text Creation Partnership

ESTC: English Short Title Catalogue

ODNB: Oxford Dictionary of National Biography

OED: Oxford English Dictionary

OII: Oxford Internet Institute, University of Oxford

SECT: Sustaining the EEBO-TCP Corpus in Transition

Executive Summary

Between March and November 2012, the SECT project, funded by JISC's Digital Preservation and Curation programme, carried out a benchmarking study of the use and impact of the EEBO-TCP corpus using the Oxford Internet Institute's Toolkit for the Impact of Digital Scholarly Resources (TIDSR). In summary, the main findings of the study were:

Usage statistics from ProQuest (the point of access for most users) show a steady increase in EEBO usage from 2004-2011. Bibliometric analysis indicates that EEBO-TCP is having an impact both in published scholarship and in new scholarship being produced as part of masters and doctoral work.

EEBO's reputation is very high amongst the user community – it is considered reliable, easy to use and easy to find. Users consider EEBO to be important for their own research and teaching, but particularly for research. They strongly believe in its importance to their field or discipline, and value its contribution to new research possibilities.

The study identified particular areas where work could be done to improve the long-term sustainability and usefulness of the resource.

- Usage statistics show that there are considerable bumps when new collections are announced, but these seem to be often followed by rather steep declines. The reasons for these declines should be looked at in more detail.
- The TCP's visibility and profile should be significantly developed in order to create a stronger brand and distinguish EEBO-TCP from EEBO. This will have particular importance in relation to the lifting of restrictions on Phase One TCP texts in 2015.
- Steps could be taken to raise the project's visibility online and through social media.
- EEBO-TCP should make more effort to target users and potential users working outside the disciplines of traditional History and English Language and Literature.
- The TCP should look in detail at metadata and documentation in order to provide more (and more useful) information or to clarify/improve the existing data.
- The most useful (and sustainable) resources will be the ones which work in tandem with each other and which link with each other in useful ways. EEBO-TCP should look to develop such links.
- The completeness of the corpus and the possibility of future corrections being made to the data are areas of particular importance in the user community. An exploration of potential funding models to allow the addition of more texts and the correction of mistakes in the data should be carried out.
- Absence of citation to digital resources is a significant problem in digital humanities – EEBO-TCP should take steps to make citation easier and to publicize the importance of citing online material.
- Careful planning for the release of the Phase One TCP texts into the public domain in 2015 needs to be carried out as soon as possible.
- The TCP must consider how to sustain the expertise and knowledge that has been developed by project staff over the course of the project. Opportunities should be found to properly exploit the significant knowledge gained over many years.

Introduction

The Early English Books Online Text Creation Partnership (EEBO-TCP) was established in 1999, as a collaborative project involving the University of Oxford, the University of Michigan, the commercial publisher ProQuest and the Council on Library and Information Resources (CLIR). The aim of the Text Creation Partnership was to create fully searchable XML-encoded transcriptions of the image sets of early printed books which form the basis for ProQuest's Early English Books Online, <http://eebo.chadwyck.com>. Phase I of the project, in which text production began, ran from 2001 to 2009, and created 25,363 searchable texts, which are available through the ProQuest interface, and also via a TCP interface, <http://eebo.odl.ox.ac.uk/e/eebo/> and <http://quod.lib.umich.edu/e/eebogroup/>. The full cost of Phase I production was around \$9 million. The project has now moved into Phase II, which aims to complete the corpus: one copy of every text printed in England or in English between 1473-1700. The completed resource will make available around 70,000 searchable electronic texts, and Phase II is projected to cost around \$10.08 million. The entire EEBO-TCP corpus will therefore represent an investment of c.\$20M, involving in excess of 90 person-years of effort applied over a 15 year period.

Context of the Study

The Bodleian Libraries and the Oxford Internet Institute sought and received funding from JISC under their Digital Preservation and Curation programme, managed by Neil Grindey, for the SECT: Sustaining the EEBO-TCP Corpus in Transition project. The first stage of the SECT project was to carry out a benchmarking study of the impact and use of EEBO-TCP, using the OII's Toolkit for the Impact of Digital Scholarly Resources (TIDSR), itself a JISC-funded initiative. The study concentrated primarily on the use and impact of EEBO-TCP in the UK.

This report outlines the results of the TIDSR study, which will be used as a basis for the creation of practical recommendations for improvements to EEBO-TCP, focussing on how best to secure the long-term sustainability of the corpus.

Research design & methods

A mixture of qualitative and quantitative methodologies offered through TIDSR were used to consider in detail the use and impact of EEBO-TCP. Quantitative methods of analytics, bibliometrics, Web 2.0 analysis, and an in-depth survey of researchers were used to build a detailed picture of the use and profile of the resource. This research was complemented by the qualitative data gathering through three focus groups, a conference, individual interviews, and email discussion. Updates on the project were made available via the project blog, www.bodleian.ox.ac.uk/eebotcp/SECT.

Quantitative Impacts

Survey of Researchers

An EEBO-TCP user survey was put up online in the summer of 2012. Participation was low during the summer vacation and the closing date of 1st December 2012 was chosen to take advantage of the return to study of staff and students for Michaelmas term. The survey was incentivized by offering entry into a prize draw. 220 people in total started the survey, 208 completed at least part of the survey, and 185 completed it in full. The survey sought to collect data on use of digital resources generally, of EEBO, and of EEBO-TCP. A summary and discussion of the results of the survey, following, will show the results of each question, and will suggest conclusions which can be drawn from the data.

Percentages of time on particular activities

Participants were asked to indicate what proportion of their time was spent in research and teaching, administration and other activities. This was a mechanism to enable us to only ask questions, for example, on teaching to those who have significant teaching responsibilities.

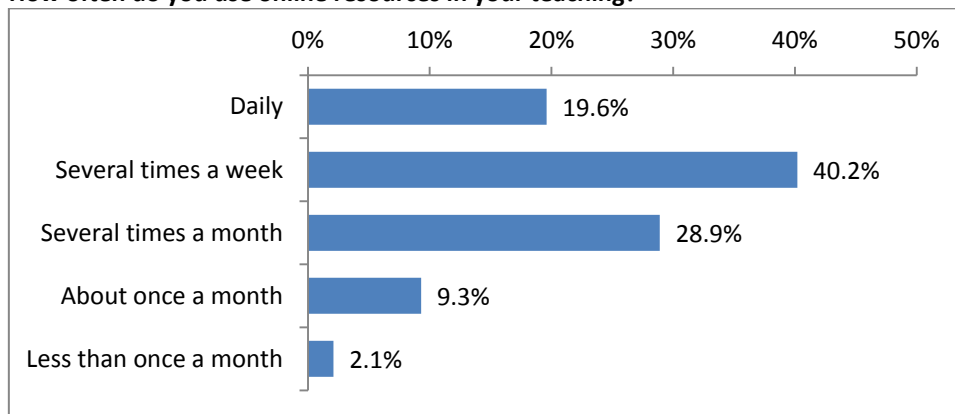
N=206	Mean percentage of time spent		Active*	
	% total	% active*	n	%
Research	52%	57%	190	91%
Teaching	18%	35%	99	48%
Administration	14%	37%	60	29%
Other activities	16%	25%	127	62%

* Active respondents are those reporting spending at least 20% of their time on the given activity.

The participants in this survey therefore spend the largest portion of their time engaged in research. Of the 91% of respondents who reported spending at least one-fifth of their time doing research, the average amount of time spent was 57%. Slightly less than half (48%) spent at least one-fifth of their time teaching, and of those, they reported spending 35% of their time on teaching activities.

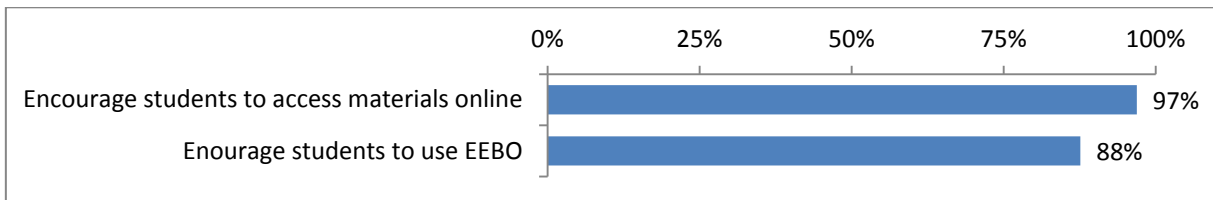
Use of online resources in teaching

How often do you use online resources in your teaching?



N=97

Of the respondents who spend at least one-fifth of their time teaching, the majority use online resources in their teaching either daily (20%) or several times a week (40%). This result indicates that use of online resources is now commonplace in teaching – of those with teaching responsibilities, none said that they never use online resources in their teaching. Creators of online content should therefore consider how they can best help teachers make use of their material, to help ensure that a particular resource keeps being used in teaching in the future. Teachers actively encourage their students to access material online and content creators need to make their resource one that is pointed to and recommended.



N=97

Nearly all of the teachers reported encouraging students to access materials online (97%), and none of the remaining 3 respondents actively discouraged the use of online materials. Of our sample (which was specifically recruited from EEBO-aware audiences), a high proportion (88%) encourage their students to use EEBO in particular.

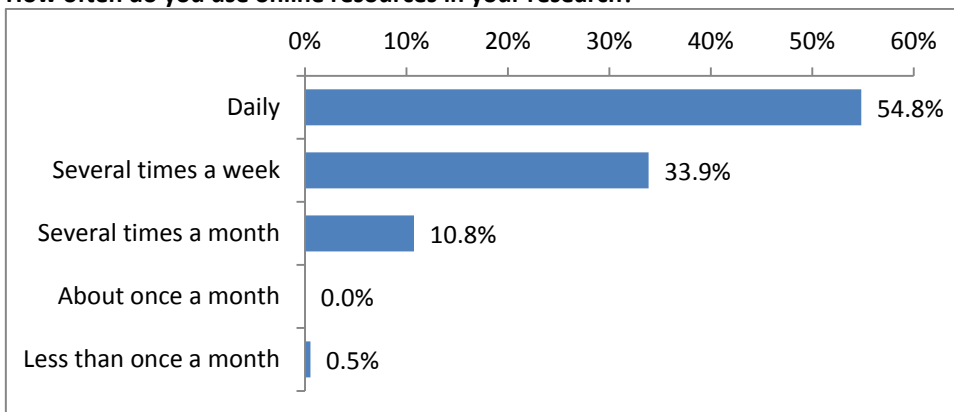
The 12 respondents who said they didn't encourage their students to use EEBO were asked to give their reasons. The free text responses were as follows:

- I was unsure how to answer this question: I *would* encourage my students to use EEBO, but I don't teach courses for which it's pertinent.
- Rarely relevant to them.
- Current university of employment does not subscribe.
- The lack of full-text searching; and also the fact that some of the images are of very poor quality.
- It is not available at my university.
- Not available at my institution.
- I'm currently teaching first year courses that do not require Early English at all.
- I teach a first year module, where accessing primary materials via EEBO is not required. If I was teaching second or third year students, I would encourage them to use EEBO.
- I teach archaeology but use EEBO occasionally for my research
- Teaching mainly nineteenth-century literature - does not seem relevant
- My library doesn't have a subscription!
- Have never heard of it until this email.

For the most part, then, teachers are not encouraging their students to use EEBO either because their institution does not subscribe or they teach courses where it is not relevant. The one critical response may indicate either lack of awareness of the possibilities afforded by EEBO (full-text searching *is* available) or may reflect dismay that not everything is available for full-text search (completeness of the corpus being particularly desirable for many scholars).

Use of online resources in research

How often do you use online resources in your research?

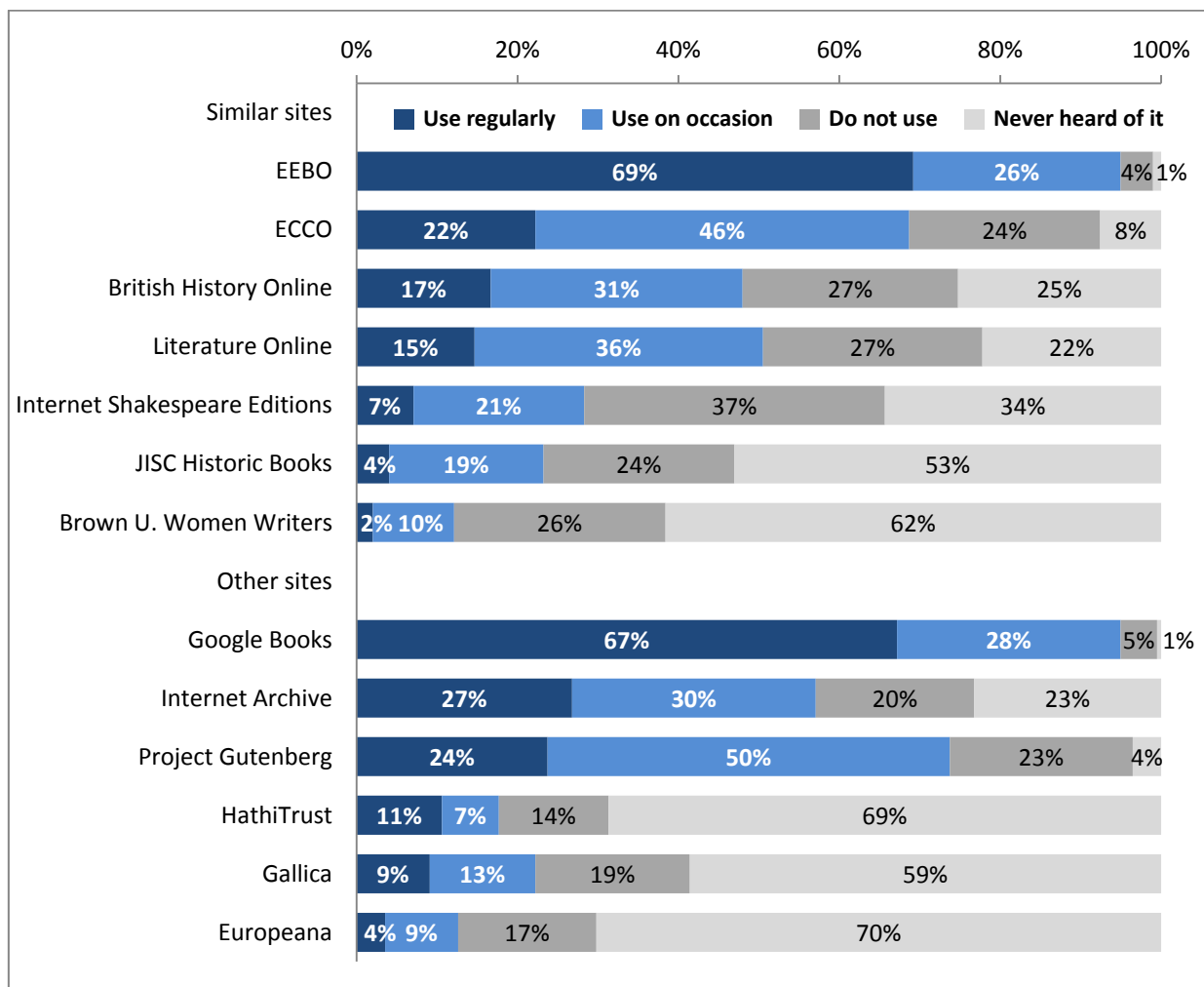


N=186

Of the 91% of respondents who spend at least one-fifth of their time on research, we can see that online resources are important tools. This result indicates just how heavily online resources are now used in research. Careful catering to the needs of the research community is therefore a key priority for creators of digital content.

Use and awareness of other resources

Respondents were asked to describe their use of or awareness of a number of resources which have some similarities with EEBO, and were then asked to indicate their use or awareness of particular general web sites that enable researchers to access electronic texts.



N=198

These results indicate that amongst this (admittedly self-selecting) audience, EEBO is by far the best known and used, but that ECCO, BHO and LION in particular are used by significant numbers of people who use EEBO. A comparison of the benefits and disadvantages of these resources would be a useful exercise to suggest possible improvements to EEBO-TCP.

These results from the general purpose sites suggest that amongst the EEBO user-community, Google Books has very high usage and recognition, Project Gutenberg and the Internet Archive are reasonably well-used and known, while Gallica, Europeana and the HathiTrust are little known or used. An exploration of the attractions of the top three “rival” resources, and of Google Books in particular, would provide useful insights into user preferences.

Finally, respondents were asked to list any additional resources which they use regularly or frequently. There were 98 free-text responses, which included all of the following 120 unique resources (with the number in the first column indicating if multiple respondents mentioned the resource):

- | | |
|---|--|
| 2 | 1641 Depositions |
| | Academic journals through university library website |
| | Allegro |
| | Alumni Cantabrigensis |
| | Alumni Oxoniensis |
| | Amazon |

Amazon Kindle
 American Memory
 3 America's Historical Newspapers
 Anglo-American Legal Tradition
 2 Archive.org
 Artcyclopedia
 2 ARTStor
 Bavarian State Library
 Bayerische Stadtsbibliothek [Munchen] (BSB)
 2 BBTI
 Bethlem Royal Hospital Archives and Museum Gallery
 Bibles online
 2 Bibliography of British and Irish History
 BibliotecaCervantesVirtual
 BnF
 Bodleian Library Broadside Ballads
 BrePolis
 Bridewell online
 2 British Book Trade Index
 2 British History Online
 British Images Pre-1700
 3 British Library
 British Literary Manuscripts Online
 British Museum prints
 3 British Periodicals Online
 British State Papers (home and colonial series, but published by different online vendors)
 Brotherton Manuscripts
 2 Burney Collection - 17th and 18th Century Newspapers online
 Calendar of State Papers Online
 CCED
 2 CERL
 Connected Histories
 Copac
 3 Database of Early English Playbooks (DEEP)
 DCB
 Dictionary of Irish Biography
 Dictionary of Literary Biography
 digilib.usm.edu
 Digital Scriptorum
 3 DNB online
 3 Early American Imprints
 2 EBBA
 EBO
 ebooks
 Ebrary
 Edit16
 e-journals
 e-rara.ch
 10 ESTC (English Short-Title Catalogue)
 2 Ethos
 folger.edu
 Gale State Papers Online
 GeoNames
 Google Maps
 2 Google Scholar
 History of Parliament
 http://www.lib.rochester.edu
 ICDL
 Individual research library's digital collections that I know of
 ISTC

- JISC Ireland
- 2 John Foxe Online
- 26 JSTOR
- KVK
- 2 London Lives
- Lost plays database - Matt Steggle
- Luminarium.org
- Manuscript facsimile websites, such as the site run by Harvard Houghton Library
- Many of the Diocese of York cause papers dating from the 17th century have been digitised, and are accessible via the Borthwick Institute website.
- Memory of the Netherlands
- Milton Reading Room
- Museum and Gallery Sites
- Museum catalogues
- 2 National Archives
- National Archives at Kew
- ODL Bodley
- 20 ODNB (Oxford Dictionary of National Biography)
- 18 OED (Oxford English Dictionary)
- 2 Old Bailey Proceedings On Line
- Online dictionaries
- online journals and newspaper databases
- Open source Shakespeare
- Oxford Scholarly Editions Online
- PARES
- Parker on the Web
- 2 Perdita Manuscripts
- Persée
- 2 Perseus Online Library
- Perseus project
- plre.folger.edu
- 2 Project Muse
- REED
- ricorso.net
- Sabin Americana
- SBN
- ScienceDirect
- Scottish Dictionaries Online
- Shakespeare Collection
- 2 Shaw-Shoemaker
- 7 State Papers Online
- Statutes of the Realm
- The Agas map
- The Burney Collection
- The Cecil Papers
- The Holinshed Project
- The Latin Library
- Thesaurus
- UCSB Ballad Project
- University of Edinburgh online resources
- 3 USTC
- V&A
- Various linguistic corpora, grammars and dictionaries
- VIAF
- Wright American Fiction

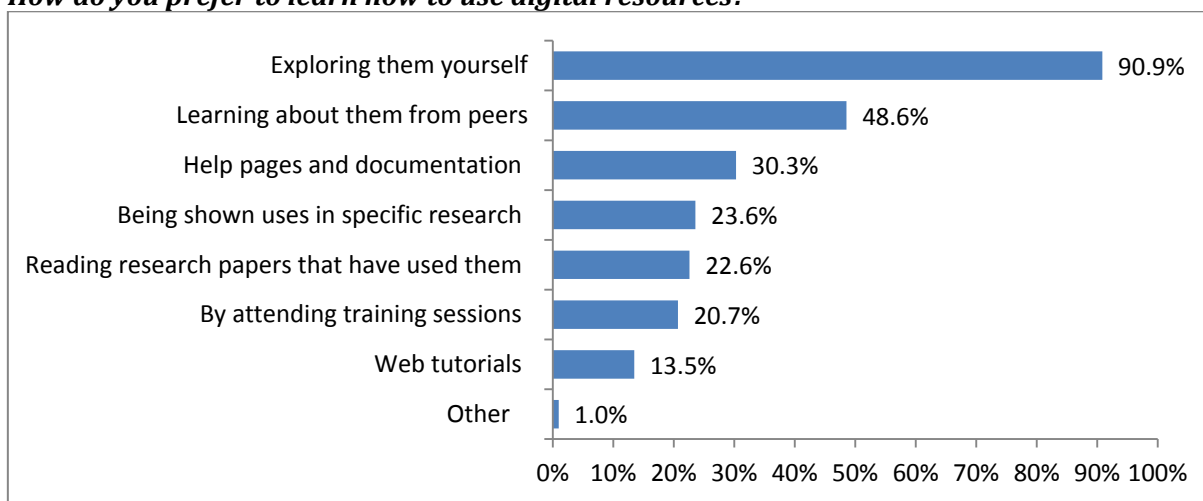
These very full responses are a very helpful indication of the kind of collections most commonly used by EEBO's user community. This sense of their being a "suite" of resources commonly used by researchers suggests that work could profitably be done exploring the potential for links with other resources, thus helping scholars to find their way through the collections they need. The

responses given above suggest that EEBO-TCP could most profitably consider the potential for links with (in no particular order):

- JSTOR
- ESTC
- OED
- ODNB
- CERL
- British Book Trades Index
- Database of Early English Playbooks
- British History Online
- EBBA

Finally in this area, participants were asked about learning to use digital resources.

How do you prefer to learn how to use digital resources?

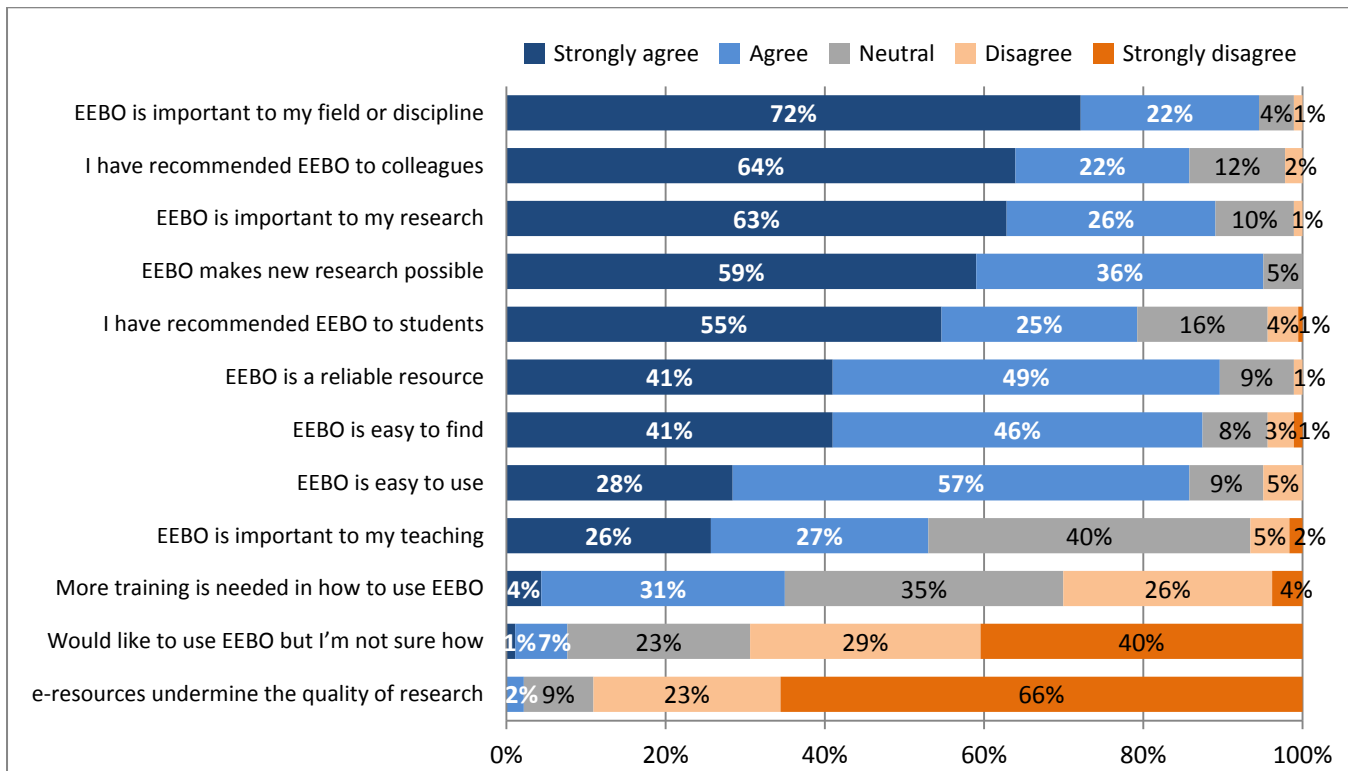


N=208

The answers to this question suggest that library training sessions and web tutorials aren't particularly effective ways of carrying out user education. Most people prefer to explore resources themselves or on the advice of their peers.

EEBO's reputation

Participants were asked to indicate their level of agreement with various statements. The results are outlined in the table below.

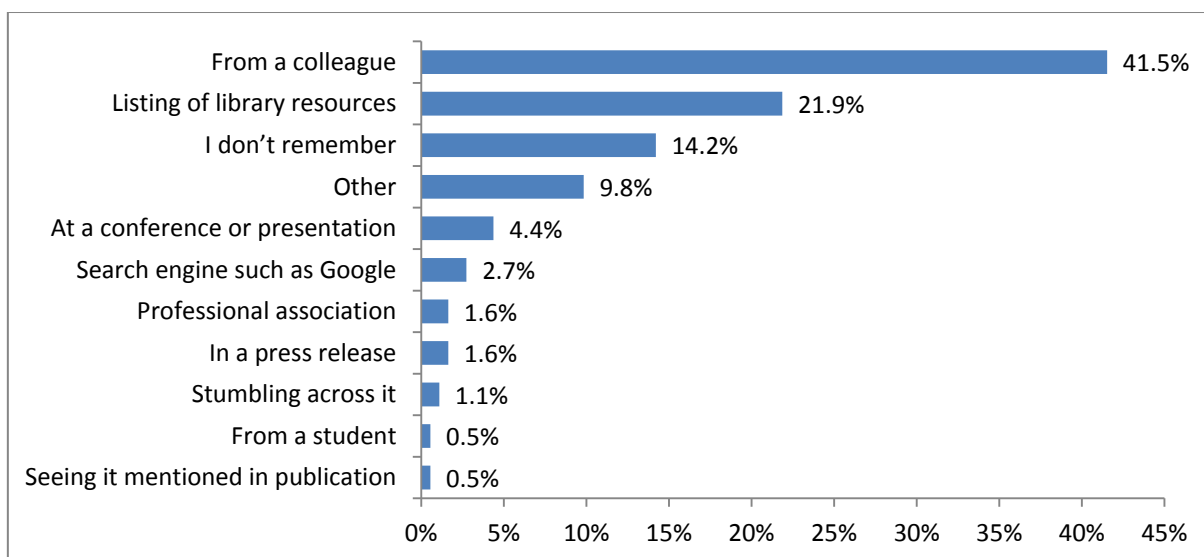


N=183

We can conclude that EEBO’s reputation is very high amongst the user community – it is considered reliable, easy to use and easy to find. Users consider EEBO to be important for their own research and teaching, but particularly for research. They strongly believe in its importance to their field or discipline, and value its contribution to new research possibilities. Users recommend EEBO to both their colleagues and students, but of these more often to colleagues. The user community appears to be aware of how to make use of EEBO in their work, and there doesn’t seem to be an overwhelming need for more training, although this would broadly be welcomed. The user community very strongly believes that electronic resources like EEBO do not undermine the quality of humanities research.

Finding the EEBO corpus

Participants were asked, if they remember, where they first learned about EEBO.



N=183

Those who responded “other” gave the following free-text responses:

- I originally used the microfilms. :)

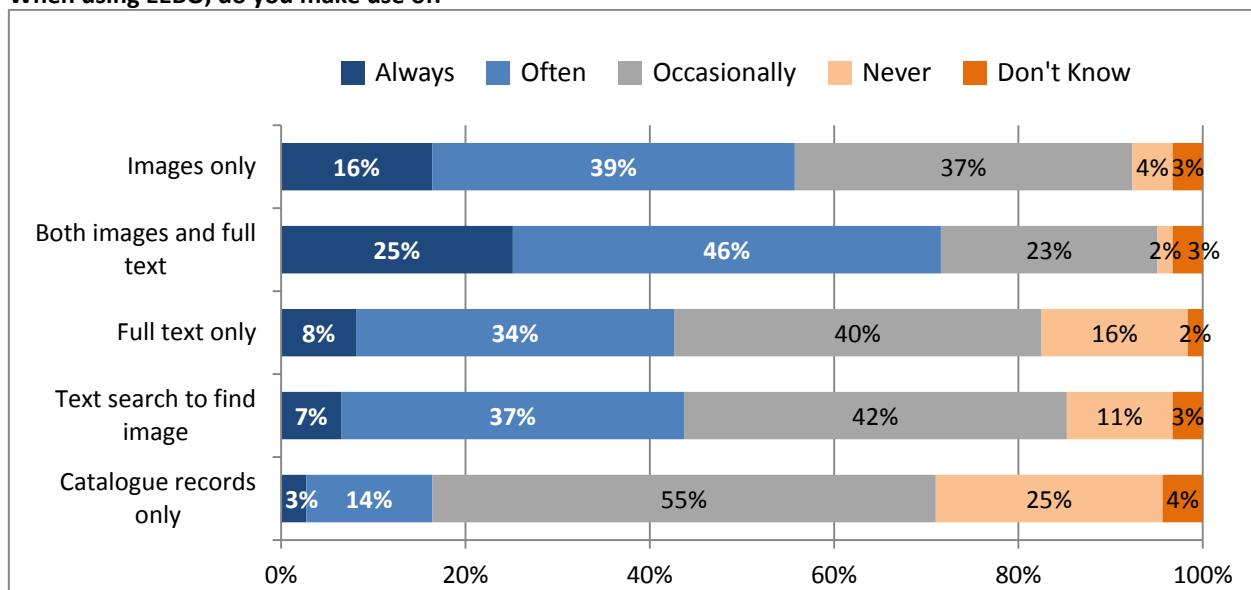
- probably as a trial of the uni library, maybe plugged by faculty members
- From a lecturer, when I was a student
- From a teacher
- Mentioned in an undergraduate lecture.
- Tutor
- In my time as a graduate student, by a professor's recommendation.
- When researching my ancestor's George Thomason Collection
- From a professor during undergraduate studies
- Mentioned by a prof
- From teachers when I was a postgraduate student
- From a graduate supervisor
- Contributor
- Folger Shakespeare Library
- Probably from hearing lecturers mention it during my undergraduate study.
- In a grad course for an assignment I had to do
- Mentioned in a postgrad course description.

It is enlightening that while library documentation seems to reach many people, the most common method of learning about EEBO is through a colleague. This may have implications for the kind of outreach work that the project pursues. The free-text responses indicate the importance of academics passing on information about EEBO to their students. Supporting these academics would therefore have knock-on benefits for their students.

Using EEBO

Participants were asked to describe how they use EEBO.

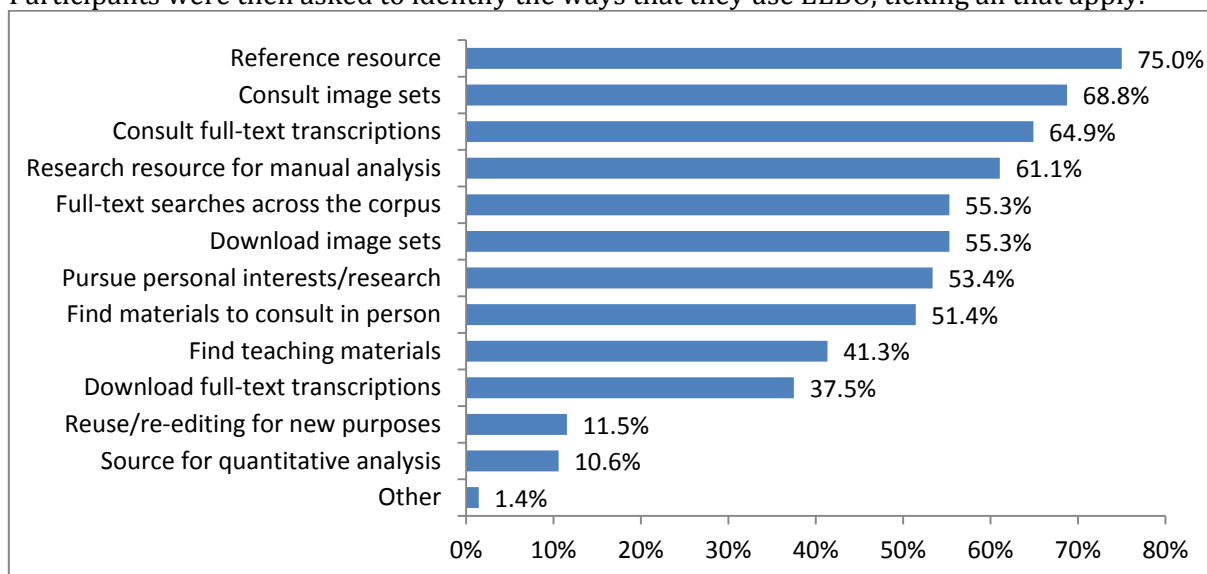
When using EEBO, do you make use of:



The results of this question indicate the value of all three aspects of the EEBO resource – the image sets, the transcribed texts and the catalogue records – working together in tandem. It also indicates the significant added value that the TCP texts give to EEBO, both as transcriptions and as finding aids.

Scholars will need different parts of the EEBO resource depending on the particular task or work they are undertaking at any one time. The value of this interconnectivity must be considered when planning for the texts being made available in the public domain.

Participants were then asked to identify the ways that they use EEBO, ticking all that apply.



N=208

Those who ticked “other” gave the following free-text responses:

- to identify fragmentary text
- to refresh my memory on resources I’ve already seen
- to supplement the OED to provide a wider sample set of how keywords were used

These figures once again strongly illustrate that users use the different aspects of EEBO in tandem and for different types of work. Large numbers of users both consult and download the full-text transcriptions. A very high number carry out full-text searches across the corpus. 12% of the users surveyed look to reuse or re-edit the texts for new purposes, which gives a snapshot of the potential for future development of the corpus.

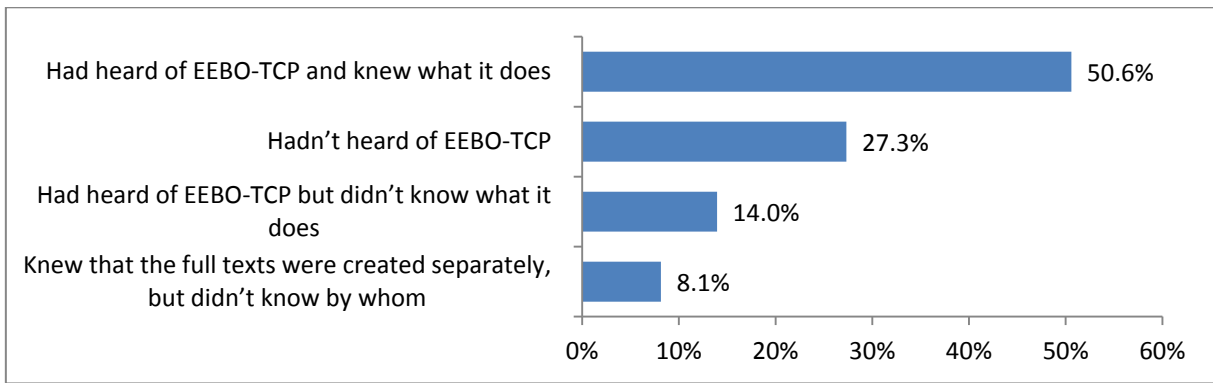
The 4 respondents who did not use EEBO were asked to give reasons why, ticking all that apply.

Reason	n
It is not relevant to my research	4
I just never got around to using EEBO	3
I don't have a subscription that allows me to access EEBO	2
It is not relevant to my teaching	1
I tried to use it, but found it too difficult to use	0
There are other similar resources I find more useful	0
I prefer working with physical materials over electronic materials	0
Other	0

This small sample suggests that those who don’t use EEBO, don’t use it because it is not directly relevant to their work or because they cannot access it. Some simply have never got round to it. It does not appear that people aren’t using EEBO because they don’t consider it useful or reliable or because they prefer other resources.

Use and awareness of EEBO-TCP

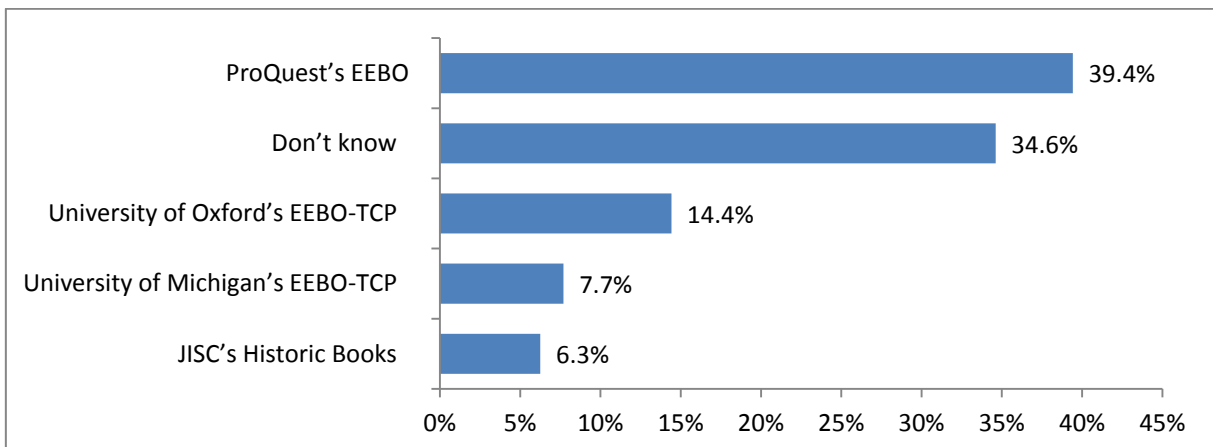
Respondents who earlier answered that they **had used** EEBO texts (either by themselves or in conjunction with images) at least occasionally were asked if, before completing the survey, they had heard of EEBO-TCP, and whether they were aware that EEBO-TCP creates the full texts available on the main EEBO site.



N=172

Of the people that answered this question, there is about an even split between people who didn't know or were unsure of what EEBO-TCP does (49%) and people who had heard of EEBO-TCP and knew what it does (51%). This is particularly striking as this was an EEBO-TCP survey and therefore to some extent a self-selecting and aware sample group. This might suggest that EEBO-TCP need to do more awareness and profile-raising work to make more people more informed about the nature of the corpus they are using and the processes via which it was created.

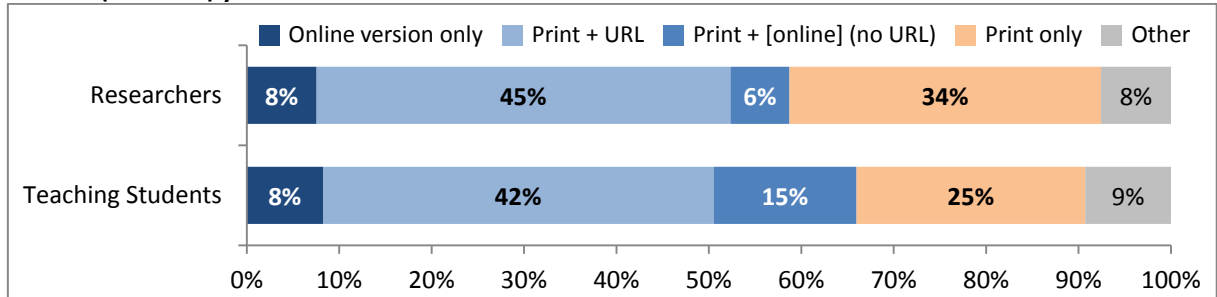
The next survey question asked, *when accessing EEBO-TCP, which of the following interfaces have you used:*



Most people, as we would expect, use the ProQuest site. However, some also have used the local implementations of the corpus at Oxford and Michigan, and a small number have used JISC Historic Books (although we'd expect this figure to rise in the future). It is notable that a significant number of people don't know which interface they use – this suggests that most users just go to EEBO, probably as linked to via their local institution's network, and don't think about who is providing the resource. Reaching such people with additional support materials could be tricky, as they may tend simply to visit sites via static bookmarks or by always using the same pathways via institutional websites.

Respondents were then asked about their citation habits with regard to EEBO-TCP.

How do (or would) you cite materials from EEBO-TCP?



Researchers, n=172; Teaching students, n=97

Researchers were asked about their habits with regard to citing materials from EEBO-TCP. Earlier in the survey, those engaged in teaching had been asked how they would teach students to cite resources consulted online; those data are also included here for comparison.

These responses reveal the variety of approaches to digital citation current within the scholarly community, as well as the startling number of people who fail to indicate their use of digital resources at all.

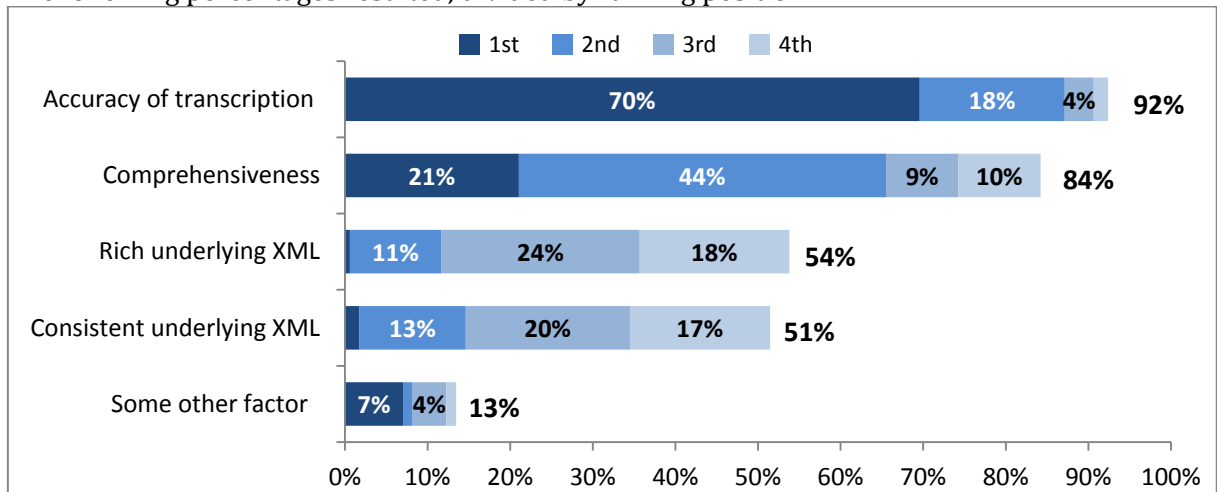
The student data is an indication that not only do many teachers teach their students effectively to hide their use of digital resources, but also that those who do teach their students to acknowledge their use, teach them to do so in different ways.

The implications of these are discussed more fully in the SECT report on digital citation, available separately.

Importance of EEBO-TCP Features

Participants were asked to rank various factors in order of importance to their work.

The following percentages resulted, divided by ranking position



Overall, then, accuracy of transcription is considered by most to be the most important factor, followed by completion/comprehensiveness of the corpus. The consistency and richness of the XML encoding are considered important but less so than the first two factors.

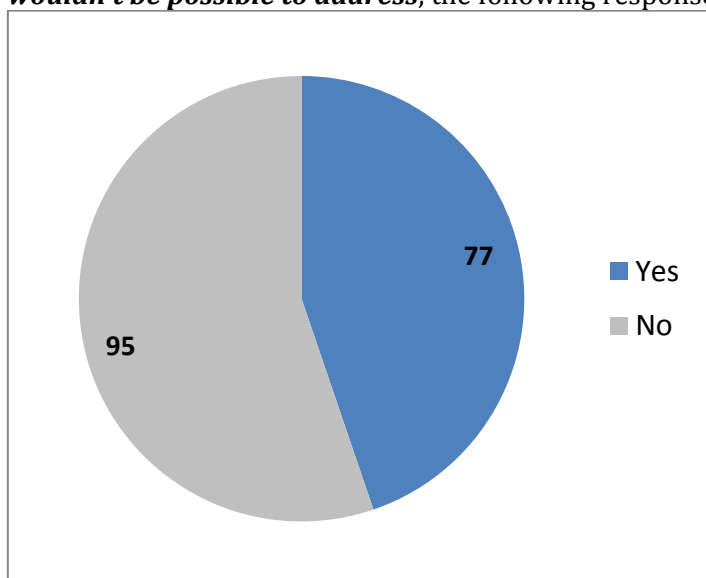
Participant who answered "other" provided the following free-text descriptions of other factors that their work depends on:

- I'd actually like to see some built-in links to OED definitions. I put that ahead of XML bc I'm not XML literate.
- Links to images are correct
- access, searching, downloading, using offline
- time

- Accessibility--i.e. it is not accessible to individual users who do not have access through an institution
- Images
- The accuracy and comprehensiveness of the search functions across corpus.
- Downloadability for texts
- User-friendliness
- Ease of relating transcription to image of corresponding original text
- Access to texts not available in hard copy
- Number of texts covered.
- The quality of the texts
- Unavailable on Google or Gallica
- I haven't used it and that was the most neutral answer
- High quality reproductions of illustrations, covers, title pages etc
- I am most appreciative of the opportunity to see an approximation of how the words appear on the page. The TCP transcriptions are a close second.

These responses indicate other areas of importance that the TCP must consider: the availability of images alongside the text, the ability to download materials and use them offline, search functionality and user-friendliness, and the ability to access texts unavailable elsewhere.

When asked *if EEBO-TCP had allowed them to ask any research questions that otherwise wouldn't be possible to address*, the following responses were given:



This result suggests the truly transformative nature of EEBO-TCP: when you take away those who didn't see or answer the question, 45% of respondents (77 of 172) have found EEBO-TCP to enable new, otherwise impossible, research questions to be explored. These respondents were asked if they could provide one or two examples of new research questions that EEBO-TCP and the responses are both varied and fascinating:

- I spent some time working on providing annotations for a historical text, and I was able to trace the sources the author used to see where he transcribed a text precisely and where he modified it. Without the ease of full-text searching, I probably would not have pursued these questions because they would have taken too much time to investigate.
- What kind of genres are there in early English printed texts? How do they change over time?
- How a specific set of works were cited across a much broader field of contemporary publishing than I would otherwise have been able to locate.
- Broadly speaking, it has allowed me to attain a greater understanding of the usage of various words over a given time period.
- The frequency / distribution of particular words between particular dates. This can be especially useful when the results run contrary to the information contained on the Oxford OED.
- Locating the use of particular words.
- Generally: Use of the full text as a subject index to texts that otherwise would not be known to discuss a

- specific person, topic, etc.
- locating a passage quoted but not cited elsewhere - finding out about the collocation of two terms
 - E.g. it allows me to track phrases, proverbs, etc. across playtexts for purposes of commentary in a modern critical edition
 - I have been able to trace printed books possibly read by early modern readers from fragments of text transcribed in manuscript commonplace books and memoranda. Working on early book lists and catalogues, which often use not the actual title of a book but a subtitle or other familiar description, I use EEBO to identify books not identifiable in any other way (for example using ESTC). This is particularly useful if a catalogue lists something which is a part/section of a bigger work.
 - Tracking the development of particular terminology across different texts.
 - how do readers read early modern books? 2) how do texts differ?
 - easier to compare newspapers texts by using electronic versions
 - Responses to minor French poets, later reception history of Du Bartas (see my paper from the EEBO-TCP 2012 conference for more)
 - It is useful for searching for key words or concepts across large swathes of text.
 - Which contemporary printed sources referenced a particular author or work, suggesting ways a work was received.
 - 'Is this word/phrase common or rare in the period?'
 - the use of the term originality in art texts the use of the term connoisseurship in art texts
 - Clarity of print.
 - It allows you to make links between apparently disparate texts that you may not otherwise think to connect.
 - The most important development that EEBO has made possible in my research has to do with the prevalence and registers of circulation of works in which particular words were present. In other words, previous research may have lead me to consider more "high literary" contexts, but EEBO helps me begin to gauge the broader contexts in which key terms and concepts may have been considered.
 - it has pushed my research in a more linguistic direction because I can search for keywords across a large corpus.
 - Searching for particular key words relating to palaeography in my research. Searching for particular key words relating to shorthand in my research.
 - Searching for references to particular writers in the newsbooks of the 1640s and 1650s. Using frequency analysis and word clouds to investigate writers' styles and preoccupations.
 - Though one must approach the TCP with care, as not all texts are available and those that are were not chosen at random, I've looked for history of usage information for various words, phrases.
 - How common was a phrase like ___ in the corpus? How early was such a phrase available in print?
 - It has helped me find very esoteric citations in books I never knew existed since these are searchable across the internet. These helped me make connections from my work to a much more global consideration.
 - How and where did writers reference bookmarks? How and where did writers reference enclosure?
 - appearance or not of specific vocabulary in author's work.
 - Access to texts I wouldn't be able to use otherwise
 - full-text searches allow for discovery of patterns of allusion
 - It helps me search full-text to find references in print sources. Sample research Q: "What early modern books include Shakespeare's proverb "fat paunches make lean pates"?"
 - How do literature scholars use electronic primary sources? What influences literature scholars' desires to use electronic primary sources?
 - Access to search allows for identifying synergies, eg on gender, recipes, ingredients etc
 - It has enabled me to search across works for brief references to objects, allowing me to build up a broader picture of how these seemingly slight allusions might have acted and been received culturally.
 - I has allowed me to think through some of the implications of text-mining and distant reading.
 - A complex set of q
 - Searching for allusions to particular names or places across a corpus. Comparing the use of such a thing in prose vs in verse.
 - was able to see language trends across time as well as determine frequency of specific attributes
 - I have been better able to look at the use of classical authors in Early Modern English geographical works
 - it is less a specific question in general and more the opportunity to constantly ask very specific questions of the primary sources because they are always available on eebo, unlike in a library where you only get a certain amount of time with them.
 - Not sure what EEBO-TCP is, but in terms of EEBO, it allows me to pursue thematic topics across the corpus irrespective of genre or other limitations.

- It has allowed me to work on images of the material text, which was important for my undergraduate thesis (and consulting the copy in the Huntingdon Library as an undergraduate was not feasible!)
- Better idea of integration of text & image - eg. how images are facing in a volume.
- ability to trace diffusion and evolving meanings of terminology through title and full text keyword searching
- As I received the email to participate in the survey, I was in fact in the middle of conducting an EEBO search to find out the history of perceptions of a relatively obscure substance (ambergris); I wouldn't have been able to begin without EEBO.
- Compare frequency of use of certain crucial terms before and after given dates. Find particular terms in texts I would not have thought to look at.
- Frequency of punctuation use in drama.
- I use it a lot to trace words, proverbial phrases, poem titles, etc. where dictionaries and other resources would give me a distorted and less full sense of their origins and history It has enabled me to identify the sources of many things in a current editing project that I would not otherwise have found. I'm not sure either of these was a question EEBO-TCP allowed me to ask, so much as a question it enabled me to hope to answer
- 'Noli me tangere' is a phrase used in Wyatt's famous sonnet 'Whoso list to hunt ...' It means 'Do not touch me'. Biblically, these words were spoken by the resurrected Christ. In Wyatt's poem they are spoken by an attractive but unattainable courtly woman, possibly Anne Boleyn. EEBO enabled me to discover that 'Noli me tangere' was also the name of a medical disease that, like syphilis, rotted the nose. So EEBO helped me to discover a new aspect of the poem, a vengeful ambiguity that attributes to this woman a particularly virulent degree of unpleasantness.
- The dispersal of new words, mostly. .
- Helping to trace early usage of certain words Helping to explore perceptions of Scotland and Scottishness in early modern period
- By giving me access to a wider wealth of early English literature it has given me a more comprehensive view of the period and consequently I feel all my research is much indebted to the resource.
- Using full text search enables retrieval of references to specific terms (eg. Psalm, Psalms) within a variety of texts which would otherwise be impossible to find. Searching titles of texts to discover range of associated meanings (eg. Diary) and diversity of texts produced under such a title.
- How often certain keywords have appeared in texts over a period of time.
- text analysis
- It has allowed me to analyse a single term across a wide period of time and range of literature.
- Dating of usage (getting round limitations of OED). Associations of words in phrases. Recurrence of textual material between works by different authors. But the primary value is as a way of reading early modern texts and doing word-searches quickly to enrich the answers I can give to relatively traditional research questions.
- It frees up time and money previously needed for physical travel to access these resources. It has allowed me to browse more resources, more quickly and thereby granted me access to new questions.
- How often does a specific term appear in early modern texts
- I am an editorial musicologist based in north-west Wales, 200 miles or more from any of the major libraries in London, Oxford and Cambridge; EEBO makes it possible for me to examine, search and trawl through far more materials than I could possibly access on any affordable research trip undertaken in order to consult original materials. It also facilitates the elimination of materials that prove irrelevant -- for example, if I am seeking copies which contain manuscript emendations.
- It has given me access to writing on cunning folk by Protestant theologians that were unavailable in my university library. As well, it provided me with transcripts of plays from the 16th-17th centuries.
- Was this abnormality present in multiple editions of this text? How were woodcuts utilised in other editions of this text?
- It allowed me to find a particular phrase difficult to find in the image set.
- Any research question that I ask has been facilitated by EEBO TCP because I am unable to spend long periods in the archives due to family commitments.
- EEBO has allowed the researcher to save large amounts of time in his/her search for various items. It is an incredible resource.
- EEBO-TCP and ECCO and similar online resources have enabled me to pursue a PhD at Trinity College Dublin that I may not have been able to pursue at a university in another country. I am disabled and suffer from serious medical conditions so the easy access allows me to engage in high quality research that I would otherwise not be able to do.
- The range of spelling variants in foreign words (French, Italian, German, etc) used in a thematically or generically defined corpus of English texts (e.g. plays from a particular decade or on a particular theme).

- Much easier access to word patterns, frequency of particular words etc.
- Shakespeare and philosophy
- What kinds of attitudes and appraisals there were towards literary publication in the 16th century. How do the translators of early print products apply the dedicatory device and other paratexts they were producing in order to justify their translation choices and evaluate the text and their own work.
- The ability to follow a particular intertextual discourse across 70 years, including both canonical and non-canonical texts, allowed me to answer a question that a very careful scholar had raised but been unable to answer 30 years ago.

These responses give us a snapshot of the breadth and depth of research enabled by EEBO-TCP. Participants views on what improvements could be made to EEBO-TCP are just as enlightening.

Respondents were asked:

What one thing would you change about EEBO-TCP to improve it, if anything?

75 people answered this question. Their answers were:

- I'd like to see more texts available--there aren't always transcriptions for the texts I'm interested in.
- Make it free
- More works keyed in!
- Include more texts
- The availability of full text searching within all texts would make the single biggest difference.
- More consistent comprehension of abbreviations + Greek characters
- Add collaborative, open, crowdsourced corrections and additional layers of markup/annotation in a stand-off manner for publicly accessibly volumes.
- Make it freely available *along with* digital images.
- Completing the full-text transcriptions for all items in EEBO.
- Speed
- Make it clear (especially for infrequent and student users) that EEBO does not contain all ESTC titles. I'm staggered by how many researchers assume that they can use EEBO to find everything written by/printed by a particular individual. There should be a large health warning somewhere! A particular frustration is that if you reorder search results by date, every time you look at one record's images the results list returns to the original alpha order - very annoying!
- Be less conservative over illegibility or find a way of including 'guesses'. Sometimes when you look at a transcription, a word that is transcribed as illegible can be deduced or is legible bar one or two letters--but by marking it or some letters as illegible, it's no longer searchable.
- Better metadata
- Accurate transcription of Latin is needed: both of texts that are entirely in Latin, and of bits of Latin embedded in English texts. The corpus of texts currently offered is biased in favour of the vernacular, for no good scholarly or intellectual reason.
- I'd improve the quality of the transcriptions.
- Make the process of downloading .pdfs easier. The marked list system is clunky and time consuming. ECCO is better. I know this is two things, but it's important that you record the shelfmark of the physical item the facsimile comes from where possible.
- Speed up the server. Searching is very slow. Pages take a long time to load.
- Greater coverage (ie ALL the EEBO books)
- I find the search interfaces complex, clunky and not always reliable
- I would suggest ways to link the image sets to the full-text transcriptions a bit more readily. (It may be that these links do exist already, and I am not familiar with them--if so, perhaps making the links from one to the other more explicit would be helpful!)
- Add more texts 2) improve the transcriptions 3) create modern transcriptions underlying the texts so that searching is rationalized
- The search function - really clunky to use compared to Google searching.
- Move toward more random or at least carefully selected set of texts to establish a more representative sample
- Eliminate genre distinctions in the defaults -- defaults should always search the maximum set.
- more texts from other literary traditions
- Price--my current institution can't afford either EEBO or EEBO-TCP.
- Make it more affordable

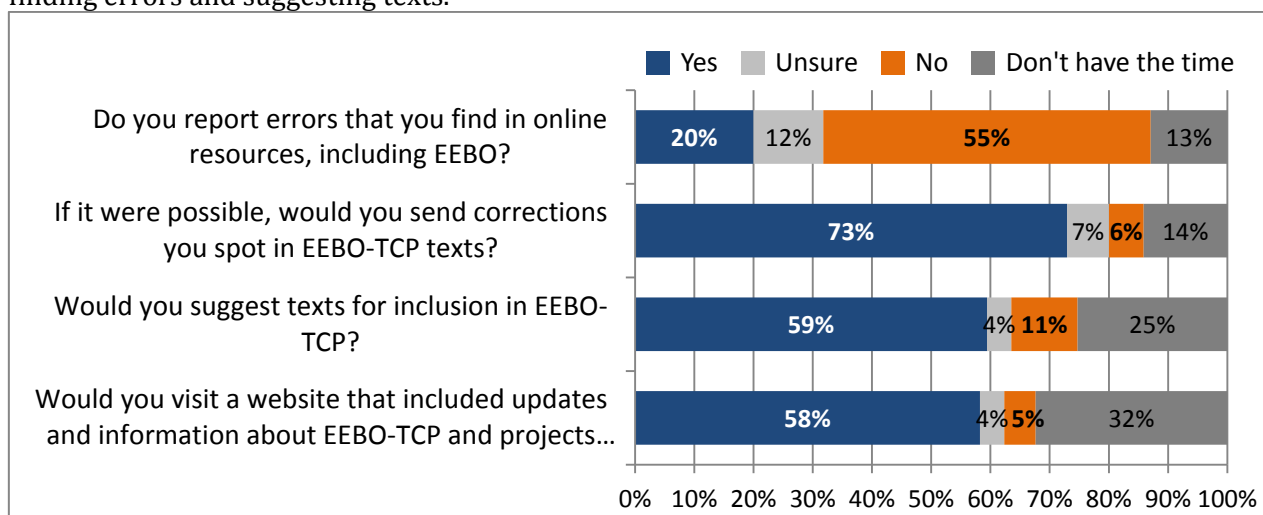
- open access
- The XML should be freely available at one click.
- Being able to search the images.
- Full EEBO texts are difficult to download for saving and printing.
- More full-text transcriptions, completed more accurately.
- A complex set of questions about forms of imprints.
- Accuracy of transcriptions and CATALOGUING information!
- still not clear on "TCP" EEBO
- haven't used it yet, but will now
- Increase the number of full-text transcriptions, so that I can do full text searches on more documents. And make the searching faster - it's very time consuming waiting for (basic) searches to complete.
- I want to know more about EEBO-TCP
- Make it easier to access information on authors without linking to LION (perhaps also linking to ODNB?)
- improvement of transcription accuracy, particularly ability to recognise abbreviated forms in black letter
- It would be useful if it could show exactly which spelling variants are being searched for with particular words.
- Allow searches that combine terms in a certain proximity (e.g. within a given paragraph)
- It can sometimes be cumbersome having to go through the marked list in order to download images.
- Accuracy of transcription.
- COMPREHENSIVENESS - there are STC items missing from it altogether (I can give an example if you want), so one must still use ESTC. - and then comprehensively good images (there are still image sets from defective copies which need replacing - again, I can give an example) - and then comprehensive availability of etexts
- All texts should have a transcription.
- quicker and easier access for remote users with my login always remembered
- Sometimes it is very slow, and if I'm doing an extensive methodical search it is frustratingly slow! I've had difficulty trying to download and save whole books - single images much easier. Seem to be unable to download full text versions which would be very helpful. Have to select/copy these - and as a result text carries on the line out of the page, so I have to redo each line.
- Accuracy of transcription
- I would make it possible for users to download searchable copies of the text with the formatting intact and I would make the online searching capacity for where one is looking for a word or phrase easier to use and faster. I would also make the formatting more consistent - there are frequently unnecessary spaces. I would consider allowing privileged users to correct inaccuracies in the full text. Most academics will check against the original images if they are going to quote an excerpt of text and then discover inaccuracies in the full text; if they were given permission to alter the originals (as in Wikipedia) this would vastly improve accuracy. Offload onto your well-educated userbase! You could moderate this if you felt it necessary. Page signatures should also be given where there are not page numbers to help users locate text in the originals.
- It would be great if the texts were properly digitised rather than reproductions from the old STC microfilms
- Somehow make it easier to navigate through the screenshots.
- More transcriptions.
- It's probably my fault but I can't always save and download EEBO TCP texts in the layout as it appears via EEBO. The sparse, text-only downloads are hard to read.
- The searching capabilities seems not consistent at times.
- It is very slow to work on most browsers. Whether searching or retrieving an image, it would be nice to have faster access
- Allow display of image and text concurrently
- I HAVE FOUND IT VERY SLOW AND LOTS OF THE IMAGES FUZZY
- better quality of transcription
- Some catalogue listings are a little opaque as regards detailed contents.
- Accuracy of transcription, including Latin and Greek,
- An improved tutorials system for new users
- I would like to have the possibility to download a pdf searchable of the text
- I found a text that was inaccurately recorded. The text differed from the record relating to it. I'd improve accuracy of data recording, in other words.
- The Search Engine does not accurately portray the depth of the catalogue and should be more simpler
- Scans of originals as opposed to microfilm to enable machine OCR. The user can to this armed with some decent software so you don't have to.

- I have always found that EEBO appears to run slower when I am browsing through image sets of periodicals. It is always much quicker to download these in full than to browse them in EEBO. I'm not sure if this is relevant to the survey though.
- More images to be made freely available
- I would like to have more information on the production choices of the full text. On what grounds is it decided if the section of text is considered prologue, dedication or 'to the reader'? Is one word code switch enough to have the work listed among those marked as containing the language in the search, etc.
- I love the way the ESTC has added links to the EEBO texts; it would be nice if that went in the other direction, and I could quickly jump from the EEBO metadata to ESTC. (I didn't give the obvious answers: free access to all and full-colour images a la Early European Books! I assume those as givens!)

All of these responses provide valuable input for EEBO-TCP, but we might make particular note of the repeated requests for comprehensiveness of coverage, improved transcription accuracy, and free open access to text and images.

EEBO-TCP feedback mechanisms

Participants were asked a series of questions about engagement with online resources, such as finding errors and suggesting texts.



N=170

This suggests that only a fairly small number of people actively report errors found in online materials. This may be because they don't have the time or inclination, but it may also be the case that they don't know how or where to report errors, or whether such reports would be welcome.

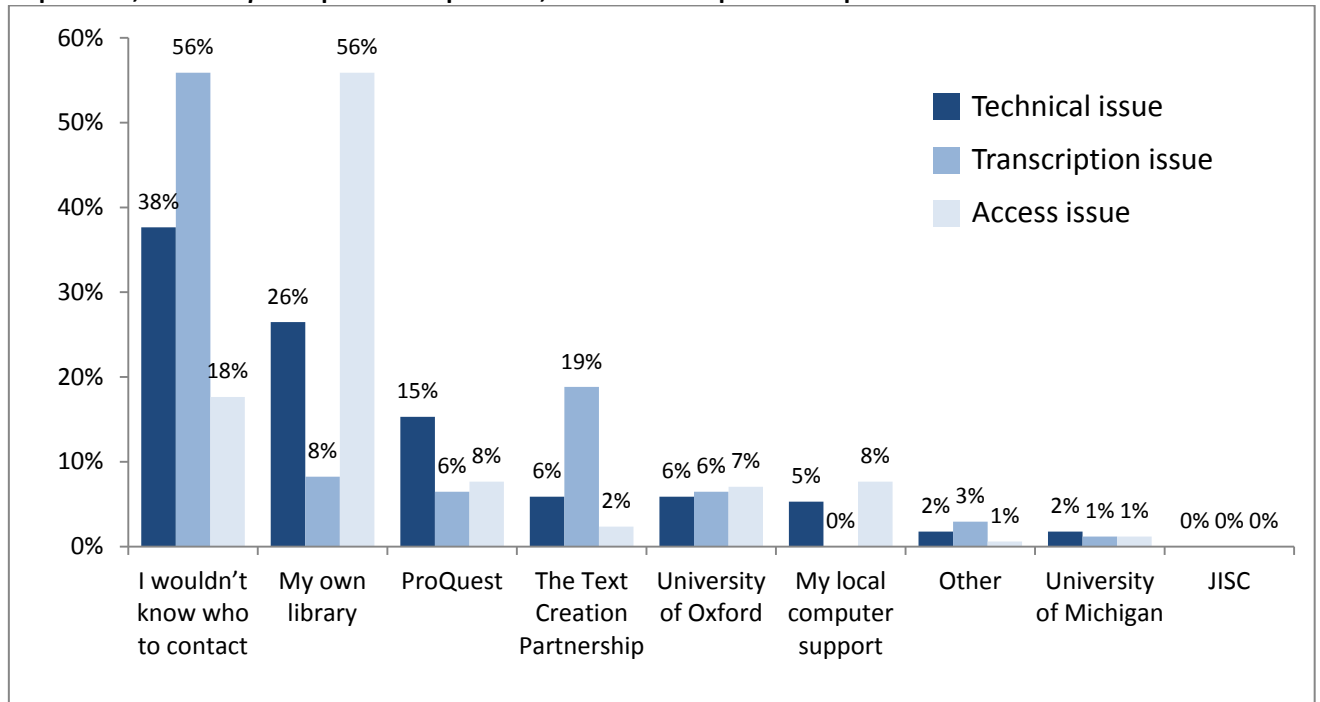
Respondents were asked: If it were possible, would you send corrections you spot in EEBO-TCP texts? This very encouraging response suggests that a very significant number of users would become involved in error correction were such a mechanism available.

Respondents were also asked: Would you suggest texts for inclusion in EEBO-TCP? Again, the positive response (59% saying yes) suggests a willingness on the part of the user community to become more involved in the work of the TCP.

With regard to whether scholars would visit a website that included updates and information about EEBO-TCP and projects related to it, a large number responded positively to this question, or were at least unsure. There were few outright "no"s. Of course, the TCP already has a centralised website, but this answer also suggests the importance of a centralised space in the future, perhaps particularly after the texts enter the public domain.

Participants were asked to identify who they'd contact in the event of different kinds of problems with using EEBO-TCP.

Who would you contact first if you needed to get in touch with someone with regard to a *technical* question or problem, a *transcription* question or problem, and an *access* question or problem with EEBO-TCP?



N=170

The answers to these questions suggest that it is not very clear who should be contacted in the event of different kinds of problem. Access and to a lesser extent technical problems are perhaps most likely to be addressed to the user's own library. For the TCP's purposes, the most revealing result is of the question on transcription errors – most users wouldn't know who to contact, and only 19% (32 respondents) would think to contact the TCP (although an additional 13 would contact either Oxford or Michigan). This once again suggests that more outreach and publicity work on the work of the TCP would be of value to the user community.

Who are EEBO-TCP's users?

Participants were asked:

Please choose the title that best describes your role when you use EEBO (or your main role, if you don't use EEBO)

Role	n	%
Undergraduate student	2	1%
Postgraduate student	71	38%
Librarian	8	4%
Professor	35	19%
Lecturer	21	11%
Researcher – academic	37	20%
Researcher – independent	8	4%
Reader	0	0%
Other	3	2%

N=185

Those who answered other:

- book conservator
- PhD Retired
- assistant lecturer

The people who answered this survey are overwhelmingly either academics or students at postgraduate level. With the caveat that more people at this level may have seen the survey, it is interesting that only 2 undergraduates responded. We might take it as indicative of EEBO-TCP's core users and be aware of the needs of this core user group – postgraduate and academic researchers – but should also consider how to reach undergraduate users better than is currently the case. The question about how people first learned about EEBO confirmed the importance of academics passing on their knowledge of EEBO to their students.

Participants were asked:

What is your main field of specialization?

Field	n	%
Languages and Literature	94	51%
History	58	31%
Art History	10	5%
Library Science	6	3%
History of Science	2	1%
Linguistics	2	1%
Archaeology	1	1%
Other	12	6%

N=185

Those who responded “other” answered:

- Bibliography
- Computer Science
- Digital Research Support
- History of book, book conservation
- History of philosophy
- Intellectual History
- Liturgy
- Music
- Music
- Political science
- Theology

As we would expect, there is a very strong showing here for Languages & Literature and History, which may be thought of as EEBO-TCP's core subjects. However, the range of disciplines represented reflects the range of materials available via EEBO. Art History makes a notable showing – an area where TCP might consider doing further outreach.

Finally, participants were asked to indicate in which country they live:

Country	n	%
England	91	49%
Scotland	9	5%
Wales	7	4%
Northern Ireland	1	1%
Republic of Ireland	25	14%
USA	32	17%
Canada	6	3%
Australia	2	1%
Other	12	6%

N=185

Again as we should expect, most of the survey respondents came from the UK and the Republic of Ireland, with a number of representatives from other English-speaking countries. The “other”

answers indicate that EEBO is also valued and used by scholars from non-English speaking countries, notably Finland and Italy (although this may tell us more about the distribution of the survey than about usage in Europe):

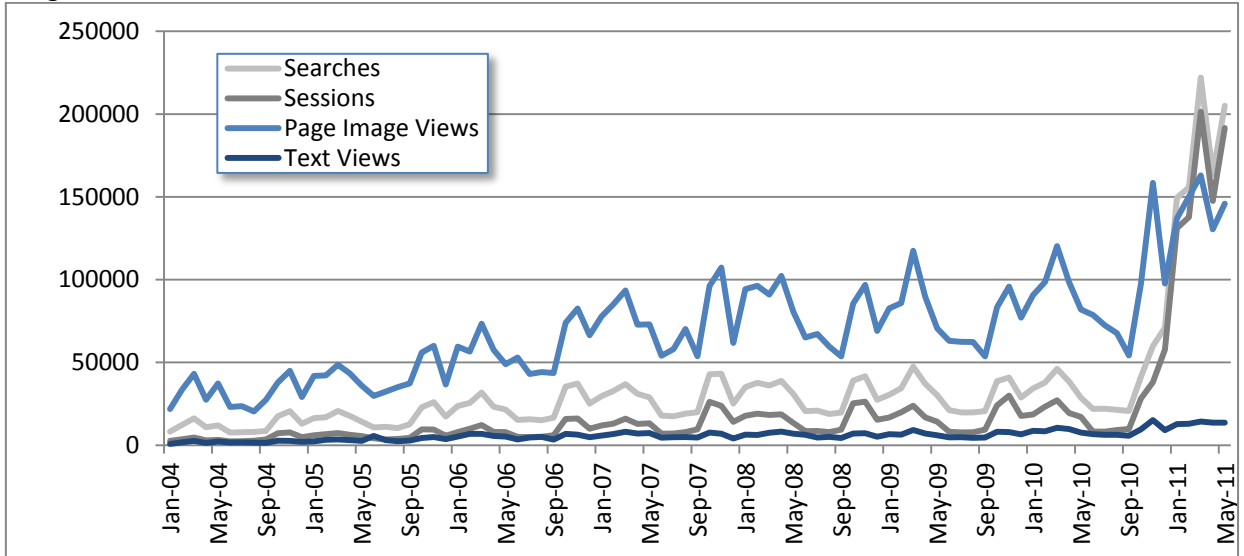
6	Italy
4	Finland
2	Switzerland
1	Netherlands

99 of the survey respondents asked to be added to our mailing list to receive EEBO-TCP updates. Making efforts to increase the EEBO-TCP mailing list may be an effective way to disseminate more information about the work of the TCP.

Analytics

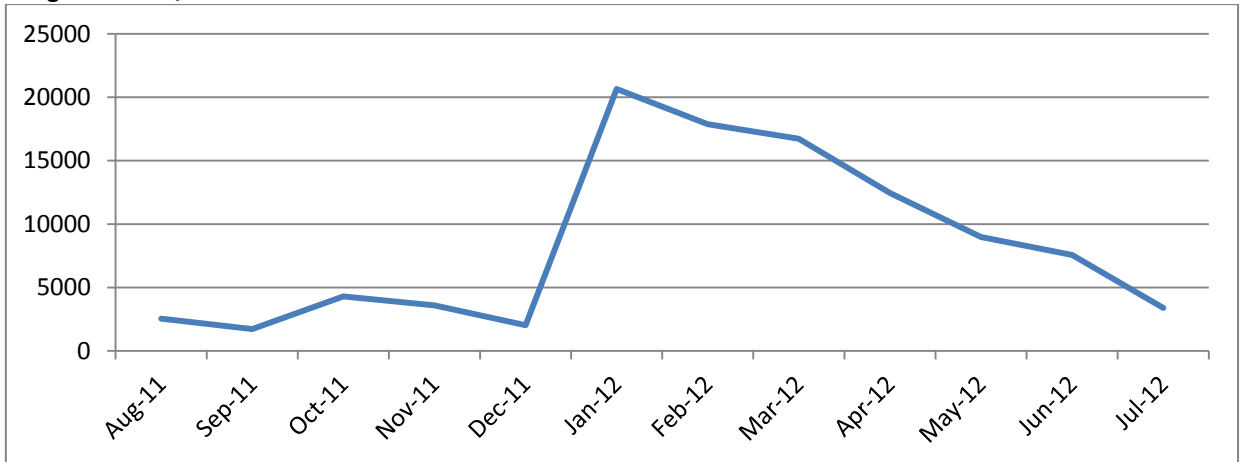
The analytics in this report are based on usage statistics from three sources: JISC Historic Books (for the period August 2011-July 2012), ProQuest (for the period January 2011-February 2012), and University of Michigan (for the period August 2009-December 2012). Selected usage statistics are presented below.

Usage statistics, ProQuest



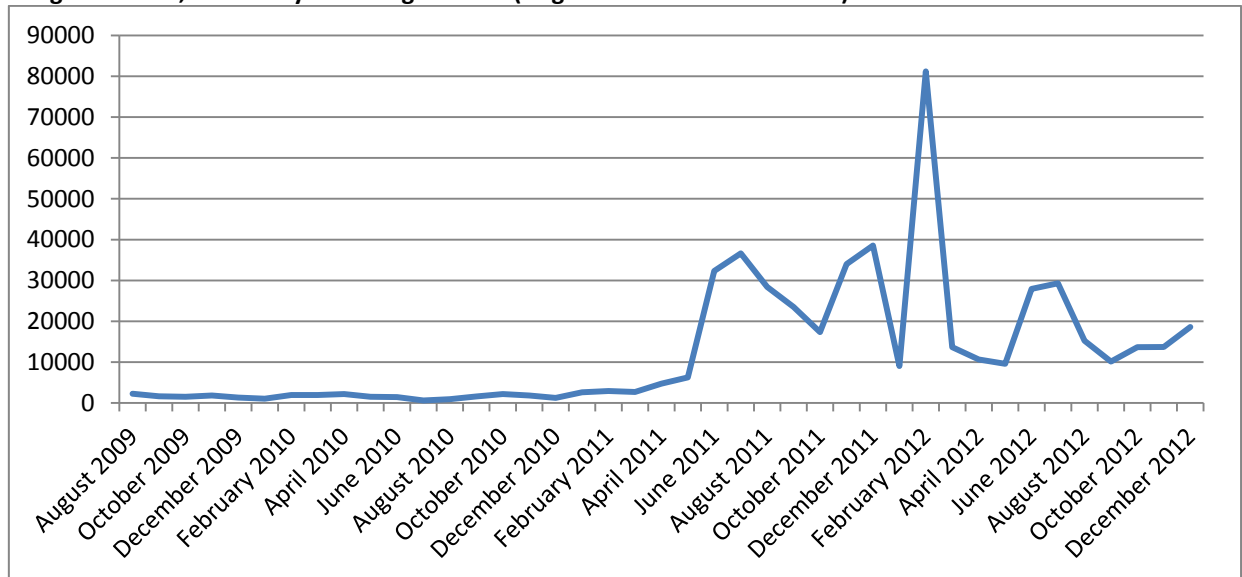
The ProQuest statistics, which cover the longest period, show a steady increase in usage from 2004-2011. There are no really surprising numbers here, which generally show a trend of steady growth with an uptick in 2011. Text views via the ProQuest interface are consistently viewed at a rate about 10% of the page image views.

Usage statistics, JISC Historic Books Collection



The JISC Historic Books Collection has much lower usage than ProQuest, and shows a worrying decline throughout 2012. This trend should be investigated further to determine if there is a reason for this decline, and if anything can be done to reverse the trend.

Usage statistics, University of Michigan EEBO (August 2009-December 2012)



While usage of the Michigan site has historically been low compared to ProQuest, numbers improved dramatically in April 2011 (when Phase II launched) and even more in December 2011-February 2012 (when the most recent batch of texts was released to the public). In general, the material that has been made public is (unsurprisingly) the content which gets the most usage.

In general, all these basic usage stats show a mixed picture: while usage in general is growing, there are considerable bumps when new collections are announced, but these seem to be often followed by rather steep declines. It is important to determine whether these declines are simply because many of the people who first come to the resources after an announcement are just window shoppers, as it were, or whether there are features of the resources which are making it difficult to turn interested audiences into regular users.

Bibliometrics

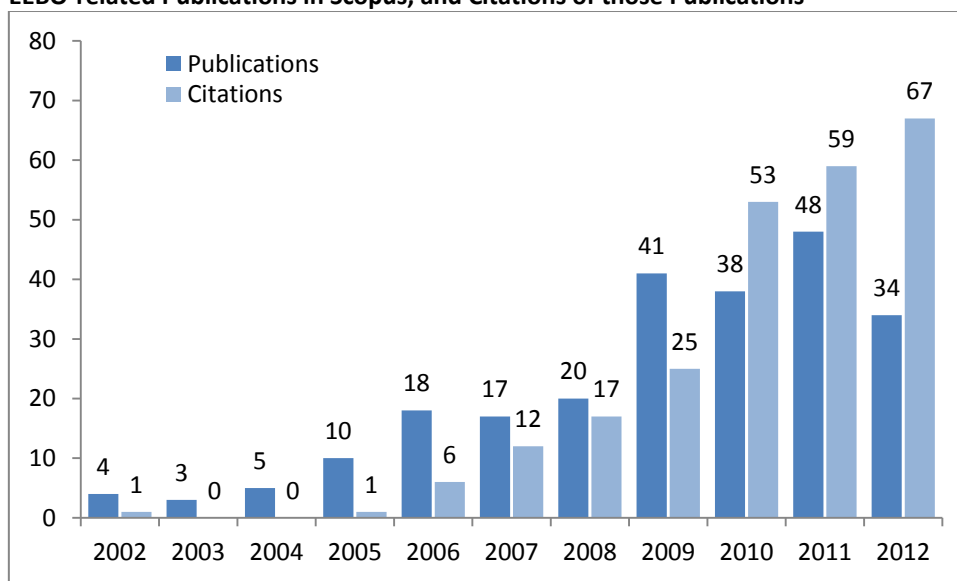
Bibliometric analysis of sources citing EEBO-TCP and EEBO were performed using data from Google Scholar, Scopus, and JSTOR. These results show the extent to which EEBO and EEBO-TCP are mentioned and cited in the literature, although they do not necessarily find all uses of the resources that did not cite or mention them by name.

Publications related to Early English Books Online

	n
Google Scholar ¹	3,450
JSTOR ²	299
Scopus ³	239
Scopus Theses & Dissertations	1,875
ProQuest Dissertations & Theses ⁴	458

It is apparent in the data shown above that Early English Books online is 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 journals 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, unpublished documents hosted on academic servers, and presentations.

EEBO-related Publications in Scopus, and Citations of those Publications



Only Scopus allows one to readily count the publications per year, but 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. In addition, we see that citations to the articles that mention EEBO/EEBO-TCP are also growing, which indicates a growing secondary impact on scholarship. One would expect the data in Google and JSTOR to follow a similar pattern.

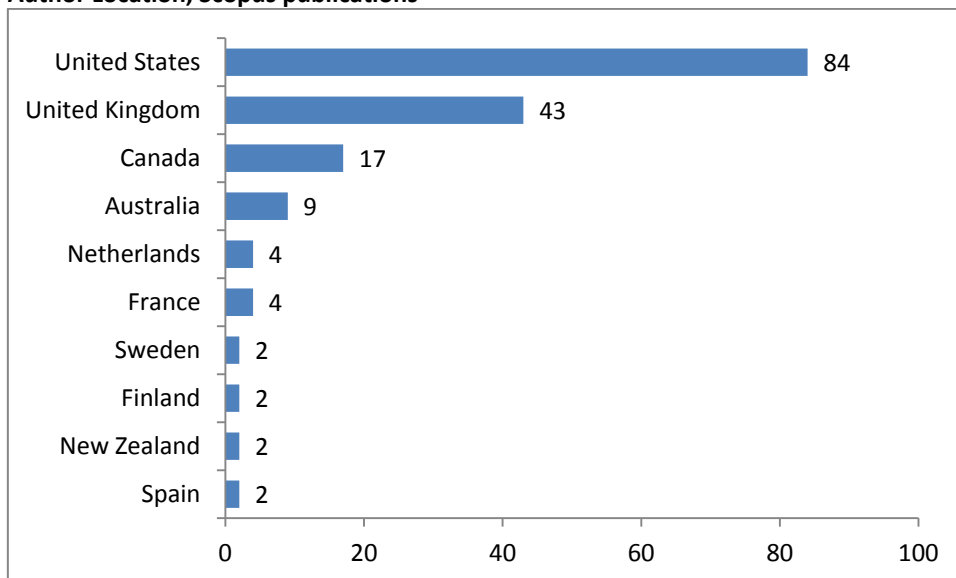
¹ Google Scholar search term: "eebo-tcp" OR "eebo tcp" OR eebo OR "early english books online"

² JSTOR search term: eebo-tcp OR "eebo tcp" OR eebo OR "early english books online" in full-text, including all content

³ Scopus search term: ALL("eebo-tcp" OR "eebo tcp" OR eebo OR "early english books online")

⁴ ProQuest Dissertations & Theses: The Humanities and Social Sciences Collection search term: "eebo-tcp" OR "eebo tcp" OR eebo OR "early english books online" in full-text

Author Location, Scopus publications⁵



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 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.

Journals of Publication, All Scopus journals with more than one article⁶

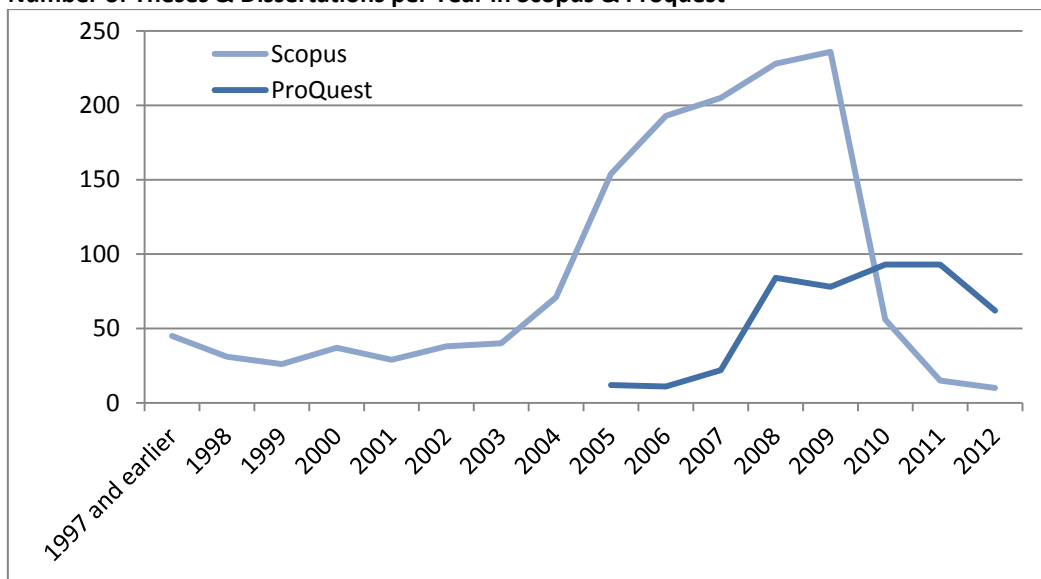
20	Studies in English Literature
9	Notes and Queries
7	Renaissance Studies
6	PMLA
5	Library
5	English Literary History
4	Studies in Philology
4	Review of English Studies
4	Literary and Linguistic Computing
4	English Studies
3	Spenser Studies
3	Shakespeare
3	Shakespeare Quarterly
3	Comparative Drama
3	Historical Journal
3	Journal of Medieval and Early Modern Studies
3	Modern Philology
3	Northern History
3	Renaissance and Reformation
2	Serials
2	Eighteenth Century Studies
2	Agricultural History
2	Reference Services Review

⁵ Only countries with more than one publication shown. 13 additional countries are also represented in the data by one publication each.

⁶ 115 additional journals only appear once each in the dataset

- 2 Journal of the History of Medicine and Allied Sciences
- 2 Milton Quarterly
- 2 Journal of the Early Book Society
- 2 Rhetorica Journal of the History of Rhetoric
- 2 Huntington Library Quarterly
- 2 Explicator
- 2 Papers of the Bibliographical Society of America
- 2 Literature and Theology
- 2 Print Quarterly
- 2 ANQ Quarterly Journal of Short Articles Notes and Reviews

Number of Theses & Dissertations per Year in Scopus & Proquest



The thesis & dissertation data shows a somewhat unusual pattern that is likely an artefact of the database. There is considerable growth from 2004-2009 in the number of post-graduate theses and dissertations mentioning the Early English Books Online resource (from 40 in 2003, to 71 in 2004, and to a high of 236 in 2009), followed by a rather remarkable apparent decline. It is likely, given the precipitous nature of this decline, that some repositories are either behind in making their data available to Scopus, or stopped indexing certain kinds of documents. This is particularly likely since the ProQuest data, which has a smaller set of dissertations, does not show a similar pattern. Instead, the ProQuest data increases dramatically in 2008 and then stays relatively steady (keeping in mind that 2012 data will still be updated in the early months of 2013).

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

Web 2.0 Impacts

Two types of data were collected and analysed: Twitter data, and data from Google Blog Search. Both results are reported below.

Twitter

Twitter data was automatically collected⁷ once a day from February 2012-February 2013.⁸

	n
Total Tweets	628
Retweets	281
Links in tweets	253
Unique Twitter accounts	225

These data show a moderately active Twitter presence, and a reasonable volume of re-tweets of content related to the resource. However, there is certainly room for improvement to increase the visibility of EEBO in the Twittersphere. This is particularly true in that the UK Twitter account for EEBO-TCP (see below) has 1,321 Followers. This shows a strong level of interest in EEBO -TCP that could be better leveraged by thinking through the EEBO-TCP Twitter strategy.



Oxford EEBO TCP
@OxfordEEBOTCP FOLLOWS YOU

The University of Oxford partners of the Early English Books Online Text Creation Partnership. Based at the Bodleian Libraries.
eebotcp@bodleian.ox.ac.uk
Oxford, UK - <http://www.bodleian.ox.ac.uk/eebotcp/>

438 TWEETS 1,931 FOLLOWING 1,321 FOLLOWERS Following

Total tweets for all twitter accounts with greater than 10 tweets

Top Tweeters	n	Short description
heatherfro	42	PhD student in Glasgow studying gender in early modern London
thefrozensea	27	"Early Modern Dialogues" blogger
OxfordEEBOTCP	26	The EEBO-TCP project Twitter in the UK
SgWingo	20	PhD student in Michigan studying rare manuscripts
jamescummins	19	U of Oxford TEI expert
HistoricBooks	15	JISC Historic Books collection
carenmilloy	15	Head of projects at JISC Collections
TCPstream	14	The TCP at the U of Michigan
perayson	12	Senior lecturer at Lancaster U using Natural Language Processing
TraceLarkhall	10	English Literature faculty at Bath Spa U
Pipwillcox	10	Bodleian staff member working on EEBO-TCP

⁷ Using an automated Google spreadsheet running the TAGS (<http://mashe.hawksey.info/twitter-archive-tagsv3/>) template set to gather data once per day.

⁸ Twitter search terms: @OxfordEEBOTCP OR #tcpSECT OR @TCPstream OR #eebo. Data from 13 February 2012 to 13 February 2013.

Visualization of the top 150⁹ terms used in Tweets



The visualization above of the words from the Twitter collection related to EEBO and EEBO-TCP show lots of attention to the September (2012) conference co-hosted in Oxford by the SECT project. Notice also the positive affected terms occurring in the data (e.g. excited, delighted, interested, great) as well as the terms that indicate the impact of EEBO/EEBO-TCP on research (e.g. revolutionizing, future, advance). Again, this collection of words can suggest ways for the team to capitalize on perceived strengths to increase the visibility of the resource in the Twittersphere more widely.

⁹ Some words removed for readability, including @OxfordEEBOTCP, RT, @TCPstream

Google Blog Search

Google Blog Search found 32 sources mentioning EEBO or EEBO-TCP. This is a relatively modest number of blog posts, and suggests that efforts could be made to make the resources more visible in the blogosphere. As you can see below, blog posts came from the project itself in some cases, and there were also a number of posts related to the conference organized by the project. Others announce the availability of new material. It would be good to see more blogging about innovative uses and new lines of research inspired by EEBO and EEBO-TCP (such as those which were discussed at the 2012 conference), and it might prove valuable to encourage scholars who we know have done interesting work to blog about it if they have blogs.

[Conference Organizers reflect on “Revolutionizing Early Modern ...](#)

www.textcreationpartnership.org/

25 Jan 2013 by rebecca

Conference Organizers reflect on “Revolutionizing Early Modern Studies”? *EEBO-TCP* in 2012. January 25, 2013 - Posted in Conference Report. This conference report was contributed by Judith Siefring, a TCP editor at the University of Oxford, with contributions from Pip Willcox, also an ... The *Early English Books Online* Text Creation Partnership in 2012, held in Oxford on the 17th and 18th September 2012, was a cause for great celebration for those of us involved in its organization.

[15 More results from Text Creation Partnership](#)

[EEBO-TCP 2012 Conference Proceedings | EEBO-TCP](#)

blogs.bodleian.ox.ac.uk/eebotcp/

7 Dec 2012 by Judith S

As the proceedings illustrate, the conference was a stimulating meeting where work and ideas using *EEBO-TCP* were shared through a series of excellent papers, posters, and discussion. The event provided a wealth of ...

[Erica Zimmer presents at recent EEBO-TCP conference » Editorial ...](#)

www.bu.edu/editinst/

18 Oct 2012 by Katherine A Evans

Congratulations to Erica Zimmer, PhD candidate in the Editorial Institute, for having presented at the recent *EEBO-TCP* conference, “Revolutionizing Early Modern Studies”?: The *Early English Books Online* Text Creation ...

[Early English Books Online-Text Creation Partnership \(EEBO-TCP ...](#)

cucataloging.blogspot.com/

7 May 2008 by Cataloging and Metadata Services

This week, CMS loaded 11,462 *EEBO-TCP* records into Chinook. The collection is described on the *EEBO-TCP* homepage as: “The University of Michigan, the University of Oxford, the Council on Library and Information ...

[Early English Books Online Text Creation Partnership: User Survey ...](#)

earlymodernonlinebib.wordpress.com/

8 Oct 2012 by Eleanor Shevlin

Posted on behalf of the *EEBO-TCP* project Please help the *Early English Books Online* Text Creation Partnership plan for the future by filling in our user survey, and be entered into a prize draw to win one of ten £50 Amazon ...

[5 More results from Early Modern Online Bibliography](#)

[Text Creation Partnership Releases Over 4,000 New EEBO-TCP Texts](#)

publishing.umich.edu/

7 Apr 2011 by admin

We are pleased to announce the release of 4180 texts from the second phase of our *Early English Books Online* Text Creation Partnership (*EEBO-TCP*) project. These texts were produced in collaboration with ProQuest, and ...

[Early English Books Online – Text Creation Partnership \(EEBO-TCP\)](#)

blogs.unimelb.edu.au/eresource/

4 Jul 2012 by Admin

Electronic resources, new databases and key research tools at the University of Melbourne.

[Call for papers: “Revolutionizing Early Modern Studies”? - Bodleian ...](#)

historyatox.wordpress.com/

15 May 2012 by iholowaty

Call for papers: “Revolutionizing Early Modern Studies”? Conference: *EEBO-TCP* 2012. 15/05/2012 by iholowaty · “Revolutionizing Early Modern Studies”? The *Early English Books Online* Text Creation Partnership in 2012. University of ...

[Collection Development Blog » 25,000 EEBO-TCP Texts Now Live](#)

obelix.lib.hku.hk/cdblog/

4 Oct 2009 by electronic resources coordinator

As one of the partner institutions of the *Early English Books Online* Text Creation Partnership (*EEBO TCP*) project, HKU Libraries is pleased to announce the completion of the first production phase of *EEBO TCP*. The project ...

[ECCO-TCP and EEBO-TCP: A new way of exploring texts from 1400 ...](#)

litlanglibrary.wordpress.com/

7 Mar 2011 by litlanguic

Tired of squinting at the scanned page images in *Early English Books Online* (*EEBO*) and Eighteenth-Century Collections Online (*ECCO*)? Those days are now gone: The *EEBO-TCP* and *ECCO-TCP* project databases offer ...

[Project Curriculum/Work Plan: Week One » Early Modern Digital ...](#)

emdigitalagendas.folger.edu/

23 Oct 2012 by Owen Williams

At the end of the day, exercises will be assigned introducing the most widely used digital corpus in early modern English studies, *Early English Books Online* (*EEBO*). *EEBO* is a commercially available collection of digitized full-text facsimiles. It currently ... Participants will break into small groups to find examples and discuss applications of *EEBO-TCP* for research and classroom use. On Friday afternoon, discussion returns to the principles of STCs by examining those ...

[Partnership makes 18th century texts available to public | The ...](#)

ur.umich.edu/feed

1 May 2011

... publishers, and university libraries to produce scholar-ready text editions of works from digital image collections, including *ECCO*, *Early English Books Online* (*EEBO*) from ProQuest, and Evans Early American Imprint from Readex. ... TCP Outreach Coordinator Ari Friedlander says the *EEBO-TCP* project is much larger than *ECCO-TCP* because pre-1700 works are more difficult to capture with optical character recognition than *ECCO*'s 18th-century texts, and ...

[Reading rare books online. « Vade Mecum](#)

andrewkeener.wordpress.com/

5 Jan 2013 by andrewkeener

For readers with access, electronic databases including *Early English Books Online* (*EEBO*) offer thousands of early and rare printed materials that can be downloaded to a home computer, printed out, consulted in a PDF ... For instance, you can page through the artifact in its entirety;

you can download it to your computer; you can peruse the ASCII text (although *EEBO's TCP* project currently only has available first-edition keyed texts, so this one would not be there).

[University of Michigan Library opens ECCO – Eighteenth Century ...](#)

www.libraries.wright.edu/noshelfrequired/

28 Apr 2011 by spolanka

... and university libraries to produce scholar-ready (that is, TEI-compliant, SGML/XML enhanced) text editions of works from digital image collections, including ECCO, *Early English Books Online (EEBO)* from ProQuest, and Evans Early ... According to Ari Friedlander, TCP Outreach Coordinator at U-M, the *EEBO-TCP* project is much larger than ECCO-TCP because pre-1700 works are more difficult to capture with optical character recognition (OCR) than ECCO's ...

[From the Director – April 13, 2012 – OSUL Odds and Ends | From ...](#)

<https://library.osu.edu/blogs/director/>

13 Apr 2012 by batts.8@osu.edu

This article from the Chronicle of Higher Education provides more detail – <http://chronicle.com/article/Language-and/127122>. *Early English Books Online – Text Creation Partnership (EEBO-TCP)*. Many years ago, Ohio State ...

[Update on TCP--full text searching of EEBO & ECCO | LCR Collections](#)

libcollections.blogspot.com/

3 Apr 2007 by Helen Clarke

Update on TCP--full text searching of *EEBO* & *ECCO*. I just wanted to ... Most notably, you can now search all 3 collections (*EEBO-TCP*, *Evans-TCP*, and *ECCO-TCP*) ... *Early English Books Online - TCP - 13,634*. *Evans Early* ...

[English at Reading · Mark Hutchings: recent and forthcoming ...](#)

<https://blogs.reading.ac.uk/english-at-reading/>

2 Nov 2012 by Cindy

In September I presented a paper at the University of Oxford's *EEBO-TCP* conference on the use of the database *Early English Books Online (EEBO)* in teaching, drawing on my Part 3 module *Editing the Renaissance* ...

[Early Modern Digital Agendas | HASTAC](#)

hastac.org/users/edgaradams122

15 Dec 2012 by zhoel13

... can historicize, theorize, and critically evaluate current and future digital approaches to early modern literary studies—from *Early English Books Online-Text Creation Partnership (EEBO-TCP)* to advanced corpus linguistics, ...

[The Future of Primary Texts Online is Almost Here - ProfHacker ...](#)

chronicle.com/blogs/profhacker/

25 Apr 2011 by Prof. Hacker

It's no exaggeration to say that life in the humanities has been radically transformed over the last decade or so as a result of the release of databases of primary texts, including *Early English Books Online (EEBO)*, *Eighteenth-Century Collections Online (ECCO)*, and *Early* If your institution is a TCP partner, then you can go to the *EEBO-TCP* Website and access all the keyed texts (and if your library subscribes to *EEBO* as well, the corresponding images are pulled in).

- [References](#)

[Launch of website | The Early Modern Blog](#)

blogs.reading.ac.uk/emrc/

13 Jul 2012 by Leigh Blount

Alice Eardley and Michelle O'Callaghan will be launching the website at the *EEBO-TCP* conference, "Revolutionizing Early Modern Studies"? The *Early English Books Online* Text Creation Partnership in 2012, which will be ...

[where material book culture meets digital humanities » Wynken de ...](#)

sarahwerner.net/blog/

29 Apr 2012 by Sarah Werner

Thanks to *EEBO* (*Early English Books Online*), ECCO (Eighteenth Century Collections Online), and Gallica (the digital collection of the Bibliothèque nationale), among others, digital facsimiles are available for us to consult and download entire works from the early modern printed world. There are limitations, of ... *EEBO-TCP* can make research a bit easier if you're interested, say, in sassafras and want to find instances of it being discussed. In the right hands, you can ...

[The Permissive Digital Archive – Copious but not Compendious](#)

blogs.helsinki.fi/kaislani/

20 Nov 2012 by kaislani

[11] The following example is from John Lavagnino, "Scholarship in the *EEBO-TCP* Age", talk by John Lavagnino at the conference Revolutionizing Early Modern Studies? The *Early English Books Online* Text Creation ...

[APPOSITIONS: Studies in Renaissance / Early Modern Literature ...](#)

appositions.blogspot.com/

30 May 2010 by noreply@blogger.com (whow)

We propose to annotate existing texts created by the *Early English Books Online* Text Creation Partnership (*EEBO-TCP*), enriching these texts by providing detailed information about their prosodic structure. We would use a ...

- [References](#)

[Humanist Discussion Group, Vol. 24, No. 906. - Renaissance Humour](#)

renhum.blogspot.com/

26 Apr 2011 by Renaissance Humour

... commercial publishers, and university libraries to produce scholar-ready (that is, TEI-compliant, SGML/XML enhanced) text editions of works from digital image collections, including ECCO, *Early English Books Online* (*EEBO*) from ProQuest, and ... According to Ari Friedlander, TCP Outreach Coordinator, the *EEBO-TCP* project is much larger than ECCO-TCP because pre-1700 works are more difficult to capture with optical character recognition (OCR) than ECCO's ...

[DigitalKoans » Blog Archive Text Creation Partnership Project ...](#)

digital-scholarship.com/digitalkoans/

9 Jun 2011 by admin

... 80,000, representing a substantial portion of the nearly 300,000 books contained in the subscription databases from which they are transcribed: *Early English Books Online* (*EEBO*), Evans Early American Imprints, and Eighteenth Century Collections Online (ECCO). ... Through 2014, the primary focus of the TCP is to produce around 44,000 texts for a second phase of the *EEBO-TCP* partnership (the first phase, which ended in 2009, produced around 25,000 texts).

[Peter Scott's Library Blog: The University of Michigan Library ...](#)

xrefer.blogspot.com/

7 Apr 2011 by Peter Scott

The University of Michigan Library has announced the release of 4,180 texts from the second phase of its *Early English Books Online* Text Creation Partnership (*EEBO-TCP*) project. The Text Creation Partnership produced ...

[More results from Peter Scott's Library Blog](#)

[Early Modern Digital Agendas at the Folger Institute | The Early ...](#)

www.emintelligencer.org.uk/

28 Nov 2012 by Karen

... current and future digital tools and approaches in early modern literary studies—from *Early English Books Online*-Text Creation Partnership (*EEBO-TCP*) to advanced corpus linguistics, semantic searching, and visualization ...

[Washington College News: WC Alum Wins National Award For ...](#)

washingtoncollegenews.blogspot.com/

10 Feb 2005 by Washington College

Chestertown, MD, February 10, 2005 — Heidi Atwood, a 2004 graduate in English from Washington College, has received the Grand Prize in the 2004 *Early English Books Online/EEBO-TCP* Undergraduate Essay ...

[p-herbals-msg - Stefan's Florilegium Archive](#)

www.florilegium.org/

15 Jun 2000

To: sca-cooks at ansteorra.org. Subject: Re: [Sca-cooks] Gerard's Herball. From: "Christina L Biles" <bilescl at okstate.edu>. Date: Tue, 13 Nov 2001 10:54:35 -0600. The URL for *Early English Books Online* is. <http://www.lib.umi.com/eebo/>. Unfortunately, you have to be a member institution for full access to the project. Gerard is Plants (Deluxe Clothbound Edition) (Hardcover) on the shelf. It's also up on *EEBO* and *EEBO-TCP* for those with academic connections. Hope this helps, ...

- [References](#)

[Bodleian Libraries secure £1 million JISC award « German Friend's ...](#)

germanbodfriends.wordpress.com/

11 Mar 2011 by iholowaty

This funding enables the *Early English Books Online* Text Creation Partnership (*EEBO-TCP*), led by the Bodleian and the University of Michigan, to make available a further 43,500 texts as part of their project to offer all ...

[HoBo: Events](#)

www.english.ox.ac.uk/hobo/

16 Nov 2009 by Ian Gadd

... full-text transcriptions of works in *Early English Books Online* (*EEBO*), we invite proposals for research papers and posters reflecting the various ways in which TCP texts are being used. Is *EEBO-TCP* revolutionizing research and teaching in ...

- [References](#)

[Bodleian Libraries - Digify - miriam mueller](#)

blog.miriammueller.net/

9 Mar 2011

This grant extends earlier work done to create the *Early English Books Online* (*EEBO*) resource through ProQuest. *EEBO* "provides access to digital facsimiles of over 125,000 works published in England or English between 1473 and 1700." The *EEBO-TCP* (Text Creation Partnership) has made available more than 25,000 documents that allow users to view either the original document or a fully searchable and browsable full-text version. the new grant will make available another ...

Qualitative Impacts

The quantitative results in the preceding section give one view of the impact of EEBO/EEBO-TCP, but to put these raw numbers in context, we also gathered extensive qualitative information using focus groups, interviews, and an analysis of user feedback. These results are reported here.

Focus Groups

SECT Workshop

The first workshop for the SECT project was held on the 20th of April, and was attended by 19 people, including academics, editors, project developers and digital technologies specialists. A full report on this workshop is available via the project website, www.bodleian.ox.ac.uk/eebotcp/SECT.

Digital Humanities Summer School focus group

A short focus group on EEBO-TCP was held on the 3 July 2012 and was attended by around 25 people. The group was made up of delegates for the Digital Humanities Summer School which is multidisciplinary and not period-specific, and so provided a useful audience, not restricted to those with a particular interest in literary studies, history or the early modern period. Attendees were asked about their awareness of EEBO-TCP and about their use of digital humanities resources.

The main points to emerge from this focus group were:

1. EEBO-TCP is reasonably well-known in a general humanities context. About a third of attendees said they'd heard of it, and one compared it with the OED in terms of the importance of the resource in his department.
2. When asked about other resources comparable to TCP, participants suggested
 - a. Broadside Ballads at Santa Barbara (modern text, old layout; only 4000 of 10000 ballads done; in some ways better than EEBO, in others worse)
 - b. ECCO (participants aware that it was OCR-based)
 - c. Old Bailey Online
 - d. Allegra catalogue ("not efficient to search")
 - e. GoogleBooks
 - f. Internet Archive (user liked that they can make their own collection and search it)
3. What improvements/developments would users like from EEBO-TCP?
 - a. Make it easier to find particular genres, e.g. ballads
 - b. diachronic presentation / n-gram-like qualities
 - c. ability to separate the paratext, e.g. front matter
4. No one seems to go to library training sessions. Participants were unaware of EEBO training sessions. When asked for a show of hands, about half preferred online training and about half face to face, but some wanted both.
5. One participant suggested that librarians are against publisher training; they think it's not appropriate. So if EEBO-TCP offers training, it should be made very clear that it is EEBO-TCP editors who are doing the training.

Digital Citation Focus Group

On the 20th November a focus group on digital citation and research methodologies in the Humanities was held, attended by 15 people, including academics, editors, online and print publishers, and digital content creators. The focus group was held in response to themes which

developed at the EEBO-TCP conference in September 2012 (more details are given under “User feedback” below). A full report on digital citation will be made available separately.

The focus group explored the lack of adequate citation of digital resources and the variation in practice amongst those who do cite or otherwise acknowledge their use of such material. The group discussion indicated that there is unlikely to be a quick or easy way to change either the perception that digital resources are somehow less “respectable” or scholarly than traditional print ones or the variation in citation practice. However, there are measures which could gradually improve the citation situation in the Humanities.

The group suggested that the main measures likely to improve the situation are:

1. **Publicizing the issue.** We must continue to talk about it, formally and informally. Conference papers, blog posts, articles and presentations which focus on the problem of digital citation will keep the issue current and will encourage users to consider their own practices. It may be particularly productive to target the subject associations as they often run associated journals.
2. **Making citation easy.** Creators or curators of digital content must make it as easy as possible for their users to cite (or as difficult as possible for them not to), including making useful URLs.
3. **Incentivizing citation.** Researchers and projects could be incentivised to make clear their use of resources, e.g. by publicizing articles or projects on the EEBO-TCP website, or even by some kind of monetary incentive – everyone who has cited a text could enter some kind of competition (like the EEBO-TCP essay competition of old).
4. **Dating digital items.** Digital collections must make it clear how to date content accessed via their resource. Release information and/or editorial updates should be made as obvious as possible.
5. **Interdisciplinary knowledge exchange.** We must look for input from other areas of study where there is philosophical overlap – for example, citation for audiovisual material (the work being done by Sian Barber at Royal Holloway) or for music.
6. **Respected institutions leading change.** We may already be seeing a gradual shift in practice led by respected bodies – recently the Royal Society announced their move to continual publication, whereby they will give a DOI but no page numbers. Such moves will change the focus within digital scholarship.

In addition to the broad measures outlined above, we should consider specific measures in relation to teaching and training. Uptake for library training sessions tends to be low. Uptake for web tutorials, anecdotally, seems rather low too – but this may be due to lack of publicity or planning for dissemination. Participants agreed that it would be helpful to draft citation guidelines for digital resources that could be circulated to academic departments and subject administrators for inclusion in local documentation circulated to students as they begin their studies. Making such teaching and training materials easily available on project websites would also be helpful. One-off project-led training sessions could be worthwhile, providing that they are properly promoted to encourage good attendance.

What other, more specific, recommendations emerged from the focus group discussion?

1. Make URLs as short as possible and, if possible, human-decodable.
2. Include a clear link to a citation from the main page of a text, image, etc.
3. Encourage/guide users always to give a date of access whenever they cite a digital resource, and include such a date in automatically generated citations.
4. Provide easily accessible editorial documentation at the point of accessing texts and images (rather than solely on project – descriptive – websites).
5. Digital content creators should consider how best to raise and develop the scholarly reputation of their resource, and promote that resource accordingly.

6. Where content (such as, from 2015, EEBO-TCP Phase One texts) is in the public domain and not tied to one point of access, citation information should be tied to individual texts (perhaps by including a citation in the TEI header, if possible).

Overall, the group felt that the best way to tackle the problem of digital citation is to continue to raise it as an issue and prompt users to reassess individual and institutional practices.

Interviews and Opinion-Gathering

Librarians

Librarians play an important role as mediators of digital content; they are often the key people responsible for introducing users to online resources and for helping these users resolve any problems they encounter. Indeed EEBO-TCP itself is the product of libraries; text production is conducted at the University of Michigan Library and the Bodleian Library at Oxford and draws on library special collections around the world. Library representatives were present at the SECT workshops and focus groups, and at the EEBO-TCP conference. Additionally, individual interviews with specialist librarians were conducted for the project: Isabel Holowaty (History Subject Librarian at the Bodleian Library), Sarah Wheale (Bodleian Library Rare Books Cataloguer), Sean Hughes (Assistant Librarian at Trinity College Dublin library), and Teresa Pedroso (Disability Librarian at the Bodleian).

Some of the issues raised in discussions with librarians are outlined below.

EEBO is presented as a key resource for the study of History and English Language and Literature. However, students don't know the difference between EEBO and EEBO-TCP – they don't understand why some books have full text and some don't (some believe that all image sets have full-text). Academics and more advanced researchers better understand the distinction, and are better aware of the nature of the collection.

Many Oxford students come across EEBO via the MARC records available through SOLO. They may find a reference to a text in a book they are reading and then search for it in SOLO, which pulls up the EEBO text. They like SOLO because it uses a single search box, like Google. A shocking number of students find things via Google, which doesn't find things in collections like EEBO. Younger students assume that if they find something on the internet it exists, if they don't it doesn't exist – they don't, for example, ask rare books specialists about uncatalogued or card-catalogued material.

Readers don't understand some very basic things about EEBO. Some don't know that you can't full-text search images – confusion with Eighteenth Century Collections Online (ECCO) could be an issue here as ECCO uses OCR technology and allows results to be highlighted on the images. Many readers don't understand why EEBO is different.

Documentation is extremely important – users need to be able to establish clearly the nature of a collection and what they can do with it. Many students do not read help pages or editorial policy documents; perhaps a click-through button on every page of full-text would be better, saying something like “how was this text created?”

Placement of help pages and documentation also needs to be taken into account when considering the needs of disabled users – it is important not to label any help pages as “for the disabled”. This reduces the number of people who will look at it to a minority, while people who could benefit from the information ignore it as they feel it is not for them. For example, an elderly researcher whose eyesight isn't what it used to be could benefit from larger font sizes, but would not think to read documentation intended for disabled users. Calling such documentation something like “user preferences” would be more useful to more people. The language level of help pages is also important – overly complicated textual language should be avoided.

In general, online collections should ideally hold usability sessions to look in detail at how they cater for disabled users, looking for such things as clearly labelled buttons, clear form fields, easy to access text versions for the visually impaired, clearly marked steps to get to download text, and appropriate labelling.

More project updates would be welcome – these could be made available via websites or blogs and then librarians could disseminate them amongst the student body. Web tutorials would also be helpful in this respect – these need to be short and specific, e.g. how to do a proximity search, a Boolean search, wild card searching, etc. Students tend not to attend library training sessions. Training and user education materials would be better sent to academics who could give them to their students – students do attend compulsory sessions in their academic syllabus. One

librarian estimated that of the direct reader queries that she gets, around 90% of them involve online resources in some way. Readers think they understand how to use print collections and are more likely to ask about online resources. This illustrates the importance of making sure that librarians are kept up-to-date about online collections. A hub for information about EEBO, EEBO-TCP, and projects based on or related to it would be very useful in this context.

Encoding Experts

Sebastian Rahtz and James Cummings of IT Services at Oxford University are active very members of the Text Encoding Initiative (TEI) community and they have consulted on numerous projects which have made use of EEBO-TCP texts.

Some of the issues raised by James and Sebastian were:

- A primary concern about EEBO-TCP is that it is a one-way workflow; there is no mechanism to feed in corrections. Crowdsourcing was identified as offering significant potential in this area.
- The metadata associated with TCP files can be confusing. There are multiple file identifiers and unhelpful file names, plus the headers are separate from the files. Which is the canonical identifier? Such issues have implications for citation.
- Some collections (like the Oxford Text Archive) allow corrected and enhanced versions to be filtered back in to the main corpus, but this can produce a situation where multiple versions of a single text exist and have to be managed carefully.
- Sebastian and James have converted many of the TCP texts into TEI-P5, for use by projects and for development as e-books. They stress the importance in their view of converting all TCP files in this way, and suggest that funding ought to be sought to convert the texts en masse before the 2015 lifting of restrictions on Phase 1 texts.
- TC documentation is a problem. Projects should have as part of their start-up something like a wiki, which would allow all queries and resolutions to be accessible in a shared space. Users who want to work with the underlying tagging need to know what was meant by a particular piece of mark-up or why something was tagged in a particular way. Good documentation needs to be made available for users.

Projects

There have been many projects which have taken EEBO-TCP texts and have used and developed them for new purposes. Representatives of a number of these projects have participated in SECT events, notably the SECT workshop in April 2012 and the EEBO-TCP conference in 2012. Individual meetings and/or conference calls were also held with representatives from particular projects, and others were contacted via email and invited to share their views. Project perspectives appear in the workshop, conference and digital citation reports, and a list of projects based on or connected to EEBO-TCP text is included as an appendix to this report. This section will briefly deal with some additional issues raised by projects.

Many projects use the TCP transcriptions as a good starting point for further work. The John Donne Society's Digital Text Project, for example, checks and corrects errors and gaps in the transcriptions, before editing the XML mark-up to make it fully TEI-compliant before adding additional, more detailed mark-up. Such projects like to access the texts in both plain text and XML formats, and would appreciate being able to do so via FTP or a dedicated content management system that would allow them to get the text in various desired formats. E-book formats are also desirable for ease of reading, for example. As the results of the survey also suggest, accuracy and comprehensiveness are seen to be the most important elements of TCP, as XML encoding can be edited and updated automatically, and for different purposes.

Some project personnel expressed a desire to be able to feed back some of their work into the TCP corpus – many projects work carefully to perfect the particular texts and would like to be able to share the results of their hard work with other EEBO users. A mechanism to include enriched mark-up would also be attractive, although it may be problematic.

A desire for improved metadata was also expressed – better metadata would allow researchers to more effectively isolate the works or authors or genres that interest them.

EEBO-TCP Editors

Arguably the people who know the pros, cons and potential of the EEBO-TCP collection best are those who work most closely with it – the EEBO-TCP editing teams based at Michigan and Oxford. For this reason, the editors at Oxford were met with individually, and the editors at Michigan invited to contribute via email, in order to put together a picture of how the corpus is viewed by those who are creating it.

Editors' views focused in the main on the topics outlined below – these issues were identified as important, although opinion varied on the specifics.

Encoding

Some felt it very important that the TCP texts should be converted to TEI P5-compliant XML, on the understanding that this would be an automated process that would require some compromise. Maintaining a corpus that is in up-to-date TEI has significant benefits in terms of interoperability and future development. Many projects already take the XML provided by the TCP to TEI experts in order to have the data converted. It would be very useful to such users were our data already compliant. Others are much less convinced of the importance of the TEI – it was suggested that to those outside the TEI community, the encoding scheme doesn't mean much. When the texts enter the public domain, they will go in all directions (uncontrolled by the TCP) and many users will access them in e-book format via Amazon or the Internet Archive. It was argued that the underlying encoding scheme in such a context is of little interest to most users.

Conservation

Some argued very strongly that the TCP and/or Oxford and Michigan have a cultural responsibility to conserve the data that has been created. There should be a simple, plain data set held in untouchable ur-text form. There should be a time-stamped archive copy (several copies ideally) of every text created by the TCP – physical back-up tapes would be desirable from a preservation point of view. This basic principle of preservation was felt by some to be fundamentally important and of significant long-term value.

Documentation

The state of the TCP documentation was strongly felt to be a problem. Much (most?) of the editorial policy decisions are not written down in a way that users can access. It is strongly felt that users need to know as much information about how the texts were created as possible, and a clear editorial policy needs to be made available. One view was that TCP documentation should be (have been) date stamped. It should be possible for researchers to establish when a text was reviewed and therefore what policies were being followed at that time.

One suggestion was that the best way to address this problem would be to write a TCP manual with a full index (and that specific funding might be sought to carry out this work). A supplementary "history of the TCP" document would also be useful. Such complete and careful documentation would literally sustain the expertise and decision-making of the project – complete transparency on behalf of the TCP will affect how people use the resource.

Accuracy and correction

Overall editors feel that the TCP is a fantastic corpus, but acknowledge that there are inevitable inadequacies in the corpus, due to the production processes and the nature of the underlying images, but also due to the unavoidable differences between individual editors and therefore texts. In an ideal world there would be a second pass through texts, but given that this is highly unlikely if not impossible, the TCP must consider the potential offered by correction.

Some feel that there definitely should be a mechanism for correcting errors and illegibles in the corpus. EEBO-TCP will be a key resource for decades to come, and accuracy of data is very important to most users. However, it is unclear what the business model for such a process could or should be. Others were uncertain as to how much value a correction process would

actually have – how much would it really improve the data? Some suggested that a pilot project to look at correction could answer this question. The idea of shared applications is growing in currency (see Project Bamboo, for example). Involving the community of users would be an interesting approach to correcting the data. Crowdsourcing initiatives often encourage users to build a resource – like Transcribe Bentham, Your Art, What’s the Score at the Bodleian? – rather than to correct existing data. Corrections to the TCP data might be better served by the development of a plug-in for textual projects – essentially a form (plus image) for users to fill in with corrections and submit for manual checking by a digital editorial team.

However, some felt very strongly that a correction process based on crowdsourcing would be a bad idea for quality assurance reasons – it was suggested that a better model would involve teams at various libraries who could correct from the originals or one central team who could travel to various libraries.

Thoughts on correction focused on transcription rather than encoding – standardizing encoding across all texts doesn’t seem to be something editors consider workable. It was argued, however, that front and back matter might be one area that would particularly benefit from standardized encoding.

If any correction and editing to the corpus is to be carried out, it was strongly argued by some that an edition-based model be employed. An editions model would be easier to cite, and would allow the important ability to date stamp research results.

Completeness

Completeness of the corpus was acknowledged to be of importance for scholarship – research results have to be heavily caveated. The collection is inherently partial – it contains only what survived and is held by 200 or so libraries. The EEBO-TCP selection process has added another layer of partiality. Ideally the TCP corpus would be expanded to include all books in EEBO, but it was conceded that the funding for such a process would be very difficult to obtain. There were differing views on what should be done if the TCP had some more funding and therefore had to prioritise the inclusion of some texts over others. Some argued that the TCP should look at incorporating single editions of everything including foreign language texts in order to correct the huge TCP bias in favour of English language material. If multiple editions of works *are* done, however, they should be linked together. An alternative view held that the lack of Latin, French, and multiple editions was not too problematic. The TCP policy – that one edition of everything (as it were) is better than multiple editions of some things – was felt to be sound. In this view, the order of priority would be 1. Finish all English language texts, 2. Finish multiple editions of English works, 3. Work on non-English material. The latter was considered to require a different skill-set than that developed over the course of the TCP project, and would require input from specialists in the various languages concerned. The completion of non-English texts might be better done in a collective European context.

Metadata

Improved metadata was considered important by some. One view is that the TCP should add ESTC numbers and the shelf mark of the physical copy to the catalogue records and to the TCP metadata (i.e. the TEI header), because of the increasing volume of work concentrating on copy-specific research. Another suggestion was that we need to consider how users will search metadata when the texts separate from ProQuest’s image resource (i.e. when they enter the public domain).

Copyright

Some felt that while it is understandable, it is regrettable from an open-access point of view that the images and metadata are copyrighted by ProQuest. Might ProQuest be willing to change their copyright statement to allow use for professional (but not commercial) use? This would help users who might want to use an image on a blog or a presentation – they may not even realise that they ought to request appropriate permissions from the publisher.

Public domain in 2015

The Phase One texts will enter the public domain in 2015. Some editors think that the TCP has a responsibility to manage the availability of the texts at this point. A plan for providing access and a publicity strategy need to be devised.

In terms of the textual data, it was felt by some that if any improvements or developments were made to the Phase 1 texts before 1 January 2015, then there should also be a commitment to making the same changes across the Phase 2 data set. Internal consistency of this sort, across all the TCP texts, is very important.

General

Editors have many ideas about potential enhancements to and developments of the data – they believe that the editorial teams who created the texts should look into ways to further develop the data for the benefit of users.

Collectively, editors see the TCP as a brilliant project. For some, the practicalities of creating it can make one forget what a fantastic and important resource it is. It will be transformative for generations to come, especially as other resources come online. The more high-quality resources there are available to search across, the more important and enlightening research will come to light. The project needs to focus on user needs and user understanding. Our user group will need an authoritative version of the texts, and we should have a user education strategy to publicise the corpus in the future.

User Feedback

Conference

The SECT project co-hosted, with Oxford EEBO-TCP, the conference “*Revolutionizing Early Modern Studies*”? *The Early English Books Online Text Creation Partnership in 2012*, held in Oxford on the 17th and 18th September 2012. The conference coincided with the tenth year of production for the TCP in Oxford and it allowed SECT and EEBO-TCP to reflect on the impact that the corpus has had on research and teaching in the early modern period and to explore planned and potential developments in the future. The conference provided invaluable evidence for SECT of just how valued the corpus is amongst its users, and the many ways that it enhances and develops research in the Humanities. Proceedings of the conference are available via the Oxford University Research Archive, at <http://ora.ox.ac.uk/objects/uuid:4e64ddb6-f919-4cb0-8faf-85507a33af60>.

The conference was opened by Richard Ovenden, Associate Director of the Bodleian Libraries, who has been an important advocate for the TCP since its inception. Richard introduced keynote speaker Dr John Lavagnino of King’s College London, who delivered a superb survey of “Scholarship in the EEBO-TCP Age”. John set the tone for the whole conference by exploring the philosophical questions and practical challenges of digital scholarship. He explained the importance of the TCP production model – transcribed rather than OCRred text – and considered the kinds of work that the corpus allows scholars to do, either uniquely in digital rather than print form or very significantly faster than previously possible. John also introduced what would become a recurring theme in the conference – that EEBO-TCP is “everywhere in early modern studies, though largely hidden: overt citation and discussion are minimal”. This citation problem was followed up in the SECT focus group on digital citation ¹⁰ (see <http://www.bodleian.ox.ac.uk/eebotcp/sect/2012/12/digital-citation-focus-group/>), while research methodologies in the humanities remains a topic ripe for further discussion.

The first panel, *EEBO-TCP: Practice and Potential*, was opened by Becky Welzenbach, the TCP’s Outreach Librarian, who gave delegates an overview of the current state of the TCP. She was followed by Martin Mueller who gave a thought-provoking talk outlining his work on linguistic annotation of the TCP corpus, looking in particular at his work on MorphAdober ¹¹. Martin subsequently wrote a very interesting blog post on the conference and on his ideas for the future of EEBO-TCP which will be discussed further under “Projects” below. Martin was followed by Marie-Helene Lay who provided a detailed exploration of how to deal with spelling variation in early modern French and English. Panel One was completed by Elizabeth Scott-Baumann who discussed her database, which she has created together with Ben Burton, which offers early modern poetry marked up by form and metre. This first session gave a real sense of the careful and detailed development work currently being undertaken, based on EEBO-TCP materials, and of some of the challenges presented by using early modern materials for research.

Peter Auger opened the second panel, on Early Modern Reception and Response, with a fascinating discussion of how EEBO-TCP has allowed him to explore early modern English responses to French poets. The TCP corpus has allowed Peter to build on earlier research by allowing him to identify additional sources. This sense of building on and reinforcing earlier research was itself reinforced in Simon Davies’ discussion of his work on early modern demonology. Mary Erica Zimmer, like Peter and Simon, gave a clear sense of the kind of detailed and specialist work that the TCP enables in her talk on Spenser’s *Letter of the Authors*.

These stimulating panels were followed by a poster session, which showcased some of the projects which have used or are related to EEBO-TCP in significant ways. James Cummings illustrated the productive ways that EEBO-TCP materials can be enhanced and reused for new purposes. James also joined Ian Gadd, Giles Bergel, and Pip Willcox in a poster exploring a project which will be of great interest to EEBO-TCP users, the digitization of the Stationers’ Register. Jayne Henley provided a striking poster showing her work on editing texts in Welsh for the TCP. Jim Kuhn, Sarah Werner and Owen Williams of the Folger Shakespeare Library showed

¹⁰ See <http://www.bodleian.ox.ac.uk/eebotcp/sect/2012/12/digital-citation-focus-group/>

¹¹ See <http://morphadorner.northwestern.edu/>

a poster focusing on their plans for interoperable digital editions of early modern drama. Judith Siefring's final poster on *SECT: Sustaining the EEBO-TCP Corpus in Transition*, described the project's focus on assessing the impact of the TCP corpus, for which all of the posters and panels at the conference have supplied such valuable input.

The third and final panel of the first day of the conference was a superb illustration of the ways in which the TCP is being used in teaching. Heather Froelich presented her paper, co-written with Richard J Whitt and Jonathan Hope, on the TextLab course run at Strathclyde University, which fosters collaborative working to explore text and language in detail. Mark Hutchings then spoke about his course which uses EEBO-TCP materials to teach his students editing theory and practice. Leah Knight surveyed her ten years' worth of experience in using the TCP in the classroom and the challenges that this has brought with it. This excellent panel provided a useful counterpoint to the impressive and detailed research work outlined earlier in the day.

Day Two of the conference opened with Panel Four, on the subject of the politics and practicalities of editing. Daniel Carey and Anders Ingram opened with an engaging paper on their work creating an edition of Richard Hackluyt's *Principal Navigations* based on the TCP transcription. Giles Bergel followed them with a timely and thought-provoking discussion on the politics and poetics of transcription. This very practical engagement with the challenges of digital editing was followed up by Michelle O'Callaghan and Alice Eardley's presentation of their own work creating digital editions for the *Verse Miscellanies Online* project. Sebastian Rahtz closed this fascinating session with an exploration of how he and James Cummings have worked to bring the TCP encoding into line with more recent versions of the Text Encoding Initiative guidelines.

Panel Five concentrated on the work being done by the Corpus Research on Early Modern English (CREME) team at Lancaster University. Alistair Baron, Andrew Hardie, Paul Rayson, Stephen Pumphrey, Alison Findlay and Liz Oakley-Brown gave a series of papers exploring the potential of the TCP corpus for linguistic and semantic analysis, and applications in the classroom. These very stimulating papers were extremely well-received by the conference audience.

The sixth and final panel of the conference, on Digital Research Methods, was opened by Jake Halford who discussed his work on the emergence of "new philosophy" in the seventeenth-century. Jake explored how EEBO-TCP has helped him in his research and graciously suggested that hearing the work of others explored during the conference has given him possibilities for his own work. Helen Sonner then gave a very engaging paper on the popular construction of meaning in early modern print, tracing the meaning and development of the word "plantation". Matthew Steggle closed the session with a charming discussion of how EEBO-TCP has enabled his work looking for "lost plays", concentrating, for this paper, on the work of Thomas Dekker.

The conference was brought to a close with a summary and plenary discussion, led by Emma Smith. Emma skilfully pulled together the themes of the conference, highlighting the range of work being carried out using EEBO-TCP and demonstrating the value of the conference in bringing scholars together to share their work and ideas. Emma led a discussion which considered how scholars can fully embrace the possibilities offered by digital technology, and how this changing digital landscape is prompting researchers and content creators alike to think about research methodologies. How are research methods changing? How can scholars explain and make explicit their methodologies? What role can content creators play in this process? This discussion of the changing nature of the research process, and of research goals, led on to a discussion of the role of libraries and in particular of rare book libraries.

By considering the state of the TCP in 2012, this conference enabled a stimulating exploration of the changing research landscape for scholars in the humanities and for those who endeavour to support such research. The important questions raised are ripe for further discussion in the future and have fed directly in to the work of the SECT project.

Conclusions

The TIDSR analysis of EEBO-TCP has provided an important opportunity to reflect on the impact of EEBO-TCP. This process of taking a step back and looking in detail at how users use the corpus, what they like and don't like about it, what the reputation of EEBO-TCP is and how can a good reputation be maintained in the future, has been enormously valuable for EEBO-TCP. Overall, the broad consensus seems to be that EEBO-TCP is a fantastic resource and is greatly valued by the scholars that use it, but there are improvements that could be made to ensure its central place in early modern scholarship for decades to come. Many themes developed over the course of the study and were brought up by a variety of different people in different areas. The material gathered through the TIDSR process will now be used to formulate some recommendations for improvements that could or should be made to EEBO-TCP and to establish how feasible these recommendations are in practice.

Appendix

List of projects based on or related to EEBO-TCP

This list demonstrates the range of projects which have made use of the EEBO-TCP corpus. It is not an exhaustive list. All URLs were accessed on 25/01/13.

Complete Works of James Shirley, <http://www2.warwick.ac.uk/fac/arts/ren/oupjamesshirley/>

Corpus Research on Early Modern English (CRÈME), <http://ucrel.lancs.ac.uk/>

Electronic Database of Poetic Form,
http://digital.humanities.ox.ac.uk/ProjectProfile/Project_page.aspx?pid=157

Great Writers Inspire, <http://openspires.oucs.ox.ac.uk/greatwriters/>

The Holinshed Project, <http://www.english.ox.ac.uk/holinshed/>

The Hakluyt Project, <http://www.rmg.co.uk/researchers/research-areas-and-projects/hakluyt-editorial-project/>

INKE: Implementing New Knowledge Environments, <http://inke.ca/>

JISC Historic Books, <http://www.jisc-collections.ac.uk/jiscecollections/jischistoricbooks/>

John Donne Society Digital Text Project, <http://community.itergateway.org/groups/john-donne-society-digital-text-project>

LEME: Lexicons of Early Modern English, <http://leme.library.utoronto.ca/>

Manuscripts Online, <http://manuscriptsonline.wordpress.com/>

The Map of Early Modern London, <http://mapoflondon.uvic.ca/>

The MONK Workbench, <https://monk.library.illinois.edu/cic/public/>

MorphAdorner, <http://morphadorner.northwestern.edu/>

Patterns of Reference, <http://www.internetcentre.imperial.ac.uk/project/por/>

PhiloLogic, <https://sites.google.com/site/philologic3/>

The Spenser Archive, <http://spenserarchive.org>

Verse Miscellanies Online, <http://www.reading.ac.uk/emrc/research-activities/emrc-miscellanies-project.aspx>

Witches in Early Modern England, <http://witching.org/>

Sustaining the EEBO-TCP Corpus in Transition: Report on the TIDSR Benchmarking Study
Judith Siefring & Eric T. Meyer
London: JISC
2013



JISC