

## A Culture of non-citation: Assessing the digital impact of British History Online and the Early English Books Online Text Creation Partnership

Jonathan Blaney <jonathan\_dot\_blaney\_at\_sas\_dot\_ac\_dot\_uk>, Institute of Historical Research, University of London  
Judith Siefring <judith\_dot\_siefring\_at\_bodleian\_dot\_ox\_dot\_ac\_dot\_uk>, Bodleian Libraries, University of Oxford

### Abstract

This article discusses the culture of digital citation within the humanities, with specific reference to research done on the citation of two well-used digital resources: British History Online and the Early English Books Online Text Creation Partnership. Because these two resources are available in both print and digital form, they provide a good test case of whether academics prefer to cite print sources when they have used digital resources in their research.

“The initial effect of printing was not that of an increased distribution of identical copies being put into numerous hands...it was, first and foremost, the very transformation of the ethics of reading and writing” [Compagnon 1979, 245]  
[1]

### The glamour of print

There is no doubt that print sources, as opposed to digital ones, still carry immense cultural cachet. As this article will explore in detail, many scholars prefer to cite the print version of a work, even when they have only seen a digital surrogate, in part because of this perceived prestige. We shall see that such practice is problematic in a number of ways, and causes difficulties even when using resources which are widely used and respected. For example, reference works such as the *Oxford English Dictionary* (OED) and *Encyclopedia Britannica* (particularly the 11th edition of 1910-11) are viewed as paragons of scholarship. Those who have worked in reference publishing (such as the authors of this paper) may have a slightly more nuanced view: mistakes and misjudgements inevitably occur and a large pool of contributors creates problems of consistency. The larger the project (the OED runs to 20 volumes plus supplements as of last printing<sup>[2]</sup>) the more difficult amendments, at least in the print copies, become.

An example of this unevenness is the OED entry for *hubris*, which falls far below its usual standards. The definition is “Presumption, orig. towards the gods; pride, excessive self-confidence” [OED 1933]. This is flatly contradicted by the first quotations, which have nothing to do with gods. The etymology is short by OED standards, “<Greek ὕβρις” and inadequate: which parts of the definition equate to Greek ὕβρις? All of them? Arguably none of them do, in fact. Liddell and Scott’s Greek Lexicon, still the standard dictionary of classical Greek for English speakers, gives a number of meanings of ὕβρις: “wanton violence; insolence; lust, lewdness; (of animals) violence; an outrage; violation, rape; serious injury to the person; a loss by sea; an overbearing man”. The sense development from Greek to English has been ignored by the OED, except insofar as it has confused the definer.

A much better treatment of *hubris* is to be found in Wikipedia [Wikipedia 2016]. Here the complexities of the term are addressed much more fully<sup>[3]</sup>. The OED entry was written in 1933 and, it seems, has not been revised since, other than to add some newer citations. There is no doubt that when the OED team comes around to revising *hubris* they will produce a superior entry<sup>[4]</sup>. However, it is likely that there will never be another print edition of the OED [Daily Telegraph

1

2

3

2014], so any improved *hubris* will never be citable in a print format. Advocates of print citation over digital citation would be condemned by their stated principles to forever cite a piece of work that is flawed.

It is clear that we are at a critical juncture in the culture of citation. Reference sources and journals are increasingly becoming online only, or exist online in expanded versions of print. Yet, for reasons this paper will explore, there is still a widespread aversion to the citing (as opposed to the using) of digital resources. It remains a brave researcher who will cite Wikipedia in preference to the OED.

In this article we will explore the resistance to citing digital sources that is still widespread in the humanities. We have found that this resistance is particularly prevalent where a print version of the source also exists; for this reason we have focused on two digital projects which can be consulted online but whose citations can be silently “converted” to a print-only reference. As well as presenting the results of several surveys of users, we will suggest some solutions to the culture of non-citation which we uncovered.

## A neglected topic

Despite our discovery of this pervasive culture of non-citation, we have been able to find little mention of it as a research topic. It is noticeable, in reviewing the literature on digital citation, that cross-disciplinary studies of the subject tend to *exclude* humanities subjects<sup>[5]</sup>, preferring to measure disciplines from the sciences and social sciences. One such study says, “Humanities papers were excluded from the analysis because of their long citation windows and high uncitedness rates” [Lozano et al. n.d.]. Effectively, digital citations in the humanities are so low as not to register in comparison with science and social sciences: “The very low percentage of articles cited at least once may be a reflection of the tendency of humanities researchers to cite books instead of articles” [Larivière and Gingras 2009].

Note that here we are talking about citation of articles. The focus of this article, the citation of digital resources, appears to be the unloved son of the unloved son. This impression is borne out by the present authors’ experience when surveying journal editors about their digital citation practices. Many editors assumed, in an understandable but illuminating example of *déformation professionnelle*, that by digital citation we meant simply citation of online articles. This is not primarily what we mean here, but it is important to note that citation of articles is more familiar to humanities scholars than citation of digital resources such as digital libraries. This is in part because tools such as Crossref automatically offer digital citations for articles alongside print citations and in part because citation of online articles has a certain congruity and familiarity with their print siblings.

Dalbello et al. have studied the digital citations in five journals in Classical Studies and English and come to the same broad conclusions as the present article: “supporting argument by means of electronic document [*sic*]...is not considered to be evidentiary... and... [t]hese citation practices point to the still invisible nature of the electronic document that is now ubiquitous in supporting the actual research practice” [Dalbello et al. 2006, 4].

A small-scale study (with 16 participants) examined the use of e-texts, defined as “any textual material in electronic form, used as a primary source” [Sukovic 2009, 1002], and found that e-texts used in research were largely excluded from citation in the written output: “references clearly did not represent the extent of e-text used in the research” [Sukovic 2009, 1007]. Sukovic’s study, although not statistically significant, also found more likelihood of citing e-texts where no corresponding print version exists [Sukovic 2009, 1009]. What is unique about the present study is that we have concentrated on digital resources where a print copy does exist and where a digital citation is very easily converted to a print citation. This situation provides the sternest test of a scholar’s attitude and citation ethics, which in itself reflects the prevailing attitude to the scholarliness of each type of citation.

## British History Online and EEBO-TCP as case studies

This paper studies the citation of two digital resources: *British History Online* (BHO) and *The Early English Books Online Text Creation Partnership* (EEBO-TCP). These are highly instructive case studies for two reasons: they both offer digital versions of print originals (which means that users can choose to cite print, even when they have used digital) and both resources proved amenable to being analysed using the Oxford Internet Institute’s *Toolkit for the Impact of Digitised*

*Scholarly Resources* (TIDSR), which uses quantitative (analytics, bibliometrics, log file analysis, surveys, webometrics) and qualitative (focus groups, workshops, interviews, user feedback) methods to evaluate the impact of scholarly digital resources [TIDSR n.d.].

British History Online [BHO 2015] and the Early English Books Online Text Creation Partnership [EEBO-TCP 2015] have both become key resources in their fields. BHO, begun in 2003, is a digital library of core sources for the history of the British Isles; it is run by the Institute of Historical Research (IHR) and the History of Parliament Trust (HoP). It contains about 1250 transcribed texts to date. EEBO-TCP is a collaboration between the Universities of Oxford and Michigan and the commercial publisher ProQuest, which creates searchable XML-encoded editions of early English books based on ProQuest's image resource. EEBO-TCP began in 1999 and will ultimately create around 70,000 full texts which are accessed chiefly via ProQuest's interface, although there are also other, local implementations. EEBO-TCP is currently available only through academic subscription at institutional level. However, all EEBO-TCP texts will ultimately be freely available in the public domain, and restrictions were lifted on the first batch of materials on 1 January 2015<sup>[6]</sup>. By contrast, in BHO, while some of the texts are behind a subscription paywall, the majority of the site is freely available.

11

EEBO-TCP aims to create an XML-encoded edition of every monographic text printed in English or in England in the period 1473-1700. If there is more than one edition of a particular work, then the first edition is selected unless there is a compelling reason to choose a later one (if the first edition is badly damaged or incomplete, for example). The texts for BHO are chosen by an academic advisory group from the IHR and HoP, with a view to meeting the needs of its core users; BHO is aimed at research-level historians (one university, after a trial of BHO's subscription content, declined on the basis that it was too advanced for its undergraduates). Similarly, EEBO-TCP's core users are generally scholars at postgraduate level and beyond, although many undergraduates do use the resource too<sup>[7]</sup>.

12

BHO receives about 10 million page views per year. Having carried out the TIDSR analysis in 2010-11, it became clear that, despite this high usage, it is very little cited in academic literature [Webster and Blaney 2011]. The bibliometrics (citation analysis) that allowed this conclusion are corroborated by qualitative data: feedback sent to BHO from scholars who are irritated by the site's practice of not displaying inline page numbers. These scholars complain that it makes the process of converting what they are reading on BHO to a print citation more difficult.

13

As set out in Webster and Blaney (2011), a search for citations of British History Online in journals (carried out in 2010) found 14 results using the Scopus service and 17 results using Google Scholar. By contrast, blog posts for just the period June to November 2010 showed that British History Online was referred to 84 times.

14

Qualitative data suggested the reason for under-citation: analysis of site feedback over the period 2003-2010 found a number of complaints about the non-display of the print page numbers from the digitized books. For example, this feedback message from 2007 is representative:

15

With great pleasure, I have been going through your most excellent online version of the 'Thurloe State Papers' for a scholarly paper which I am writing. After consulting a particular transcribed document, I would then click on to its approximate 'Page' number in the heading above, so as to search and find the exact citation for my Endnotes. However, I now see that your current online version does not display any original page images, thus preventing me from determining a precise citation. How can I access the images, so as to find the correct page-number for any given document?

EEBO-TCP's application of the TIDSR analysis was carried out in 2012-13, as part of the Jisc-funded project *SECT: Sustaining the EEBO-TCP Corpus in Transition* [SECT 2013], a study of the impact and sustainability of the EEBO-TCP corpus. The TIDSR analysis was used to inform a set of recommendations and practical implementations for the future curation and development of the corpus<sup>[8]</sup>.

16

Bibliometric analysis suggested that EEBO-TCP is having a steadily increasing impact on scholarship in relevant fields. The analysis surveyed EEBO and EEBO-TCP-related publications in databases such as JSTOR and Scopus,

17

demonstrating a steady growth in such publications over the decade 2002 to 2012. The Scopus data allows us to see the country of the authors of the publications, indicating that authors from USA and the UK are most likely to mention their use of EEBO, followed by two other English-speaking countries, Canada and Australia. If we look at the journals that these articles were published in, we find a range of journals chiefly in the fields of English Literature, Language and History of the medieval and early modern periods. The bibliographic data, therefore, supports the assertion that EEBO has had an increasing positive impact on scholarship, particularly in English-speaking countries [Sieftring and Meyer 2013, 29–31].

However, user feedback (particularly via a user survey discussed in detail below), as with BHO, indicated that many scholars, particularly in the humanities, fail to cite or otherwise acknowledge their use of digital resources. The quantitative data accumulated during bibliometric analysis can therefore only be partial; if users are not citing their use of EEBO and EEBO-TCP then any numbers-based demonstration of their impact could be significantly lower than the true impact on scholarship. This disparity raises the issue of citation practice – what are users citing if they are not pointing to their use of digital collections? The image sets in EEBO and the full texts based on them in EEBO-TCP are based on actual printed books from libraries around the world: do users cite these original print copies even though they have never actually seen them?

18

This raises questions for all creators of digital resources: Why do users avoid citing the digital copies? What implications does this have for creators of digital resources, particularly when they need to demonstrate impact? And what measures can content creators introduce to combat the problem?

19

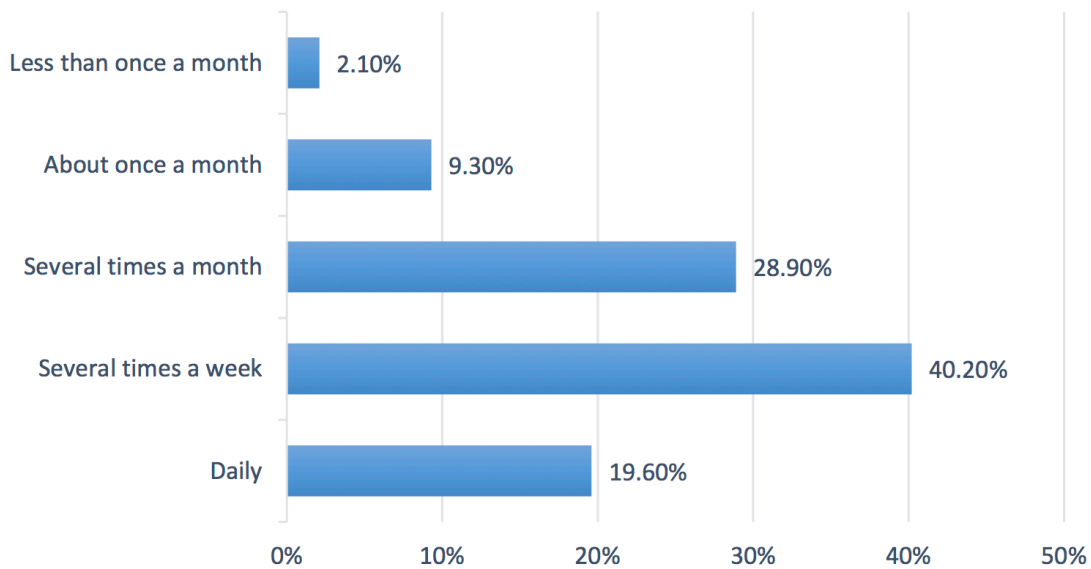
## Survey of citation practice

As part of the TIDSR analysis, the SECT project conducted an online survey of EEBO-TCP users. The survey was run for around four months from the summer into the autumn of 2012. The survey was advertised on the project website and via Twitter, and was highlighted at the EEBO-TCP Oxford conference in 2012. Details were also sent to faculty administrators at units specialising in the early modern period at institutions across the UK, for circulation to their students and staff. 220 people in total started the EEBO-TCP survey, 208 completed at least part of the survey, and 185 completed it in full. The survey asked participants for lots of information about their use of EEBO and EEBO-TCP, and some of the questions asked pertained to citation practice [Sieftring and Meyer 2013, 7–26].

20

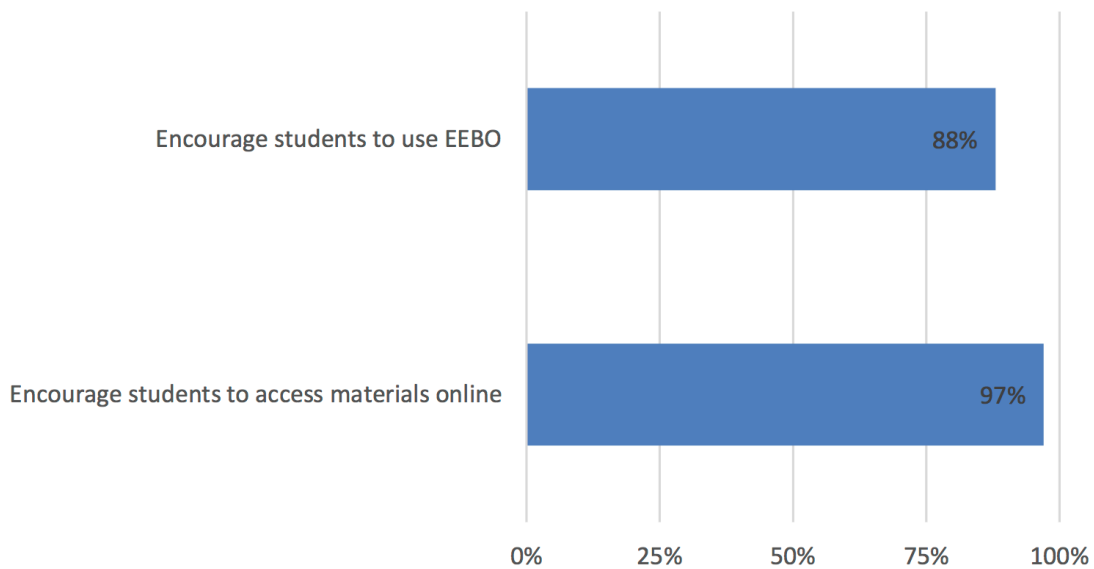
The survey sought to establish the impact of EEBO-TCP in both teaching and research, and revealed interesting attitudes to citation in both areas. First we asked users who identified themselves as spending at least one-fifth of their time teaching [Sieftring and Meyer 2013, 7]:

21



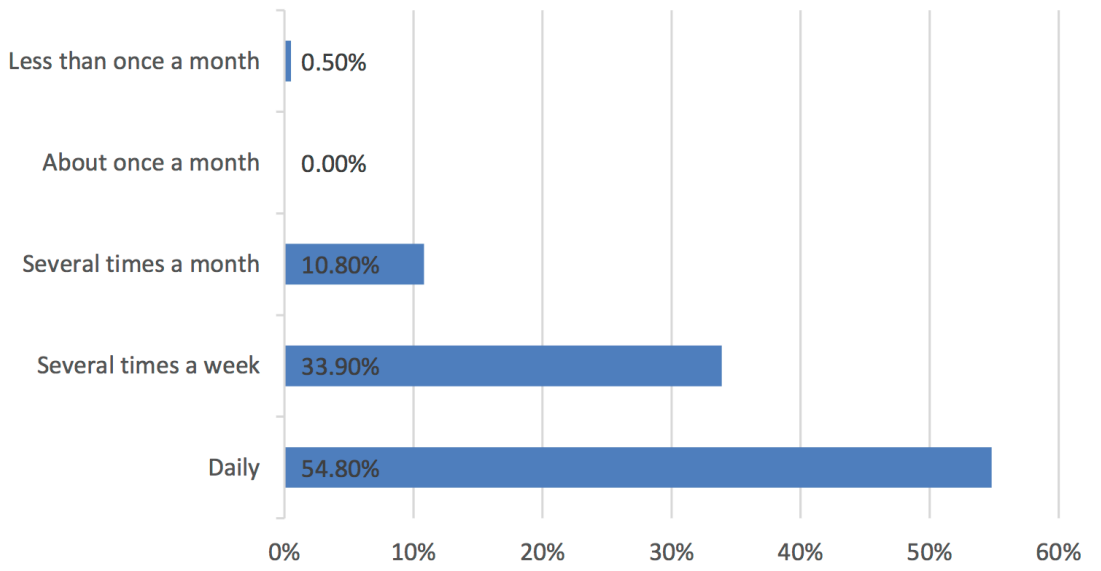
**Figure 1.** How often do you use online resources in your teaching? N=97

Most respondents use online resources in their teaching either daily (20%) or several times a week (40%). Teaching academics not only use online resources for teaching themselves but actively encourage their students to use them for their own work. As we would expect from a survey that set out to reach EEBO-TCP users, a high number of respondents encourage their students to use EEBO in particular [Siefiring and Meyer 2013, 8]:



**Figure 2.** Do you encourage your students to use online materials? N=97

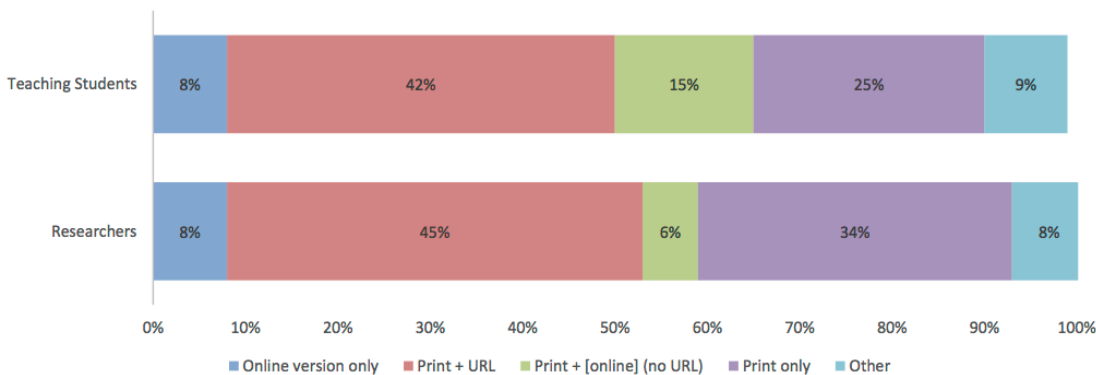
Almost all of these teachers encourage students to access online materials (97%), and none of the remaining 3% of respondents actively discourage their use. The survey also revealed that use of online resources in research is similarly ubiquitous [Siefiring and Meyer 2013, 8]:



**Figure 3.** How often do you use online resources in your research? N=186

It is clear that online resources are now heavily used by most teaching academics and researchers. EEBO-TCP in particular is very widely used in early modern studies. But is this enormous weight of use reflected in citation practice? In order to explore this question further, researchers were asked how they themselves cite materials from EEBO-TCP. Those who teach were asked how they would instruct their students to acknowledge resources that they have consulted online [Siefring and Meyer 2013, 17].

24



**Figure 4.** How do (or would) you cite materials from EEBO-TCP? Researchers, n=172; Teaching students, n=97

The use of EEBO is now commonplace in research and teaching and yet 34% of respondents fail to acknowledge that they have used an EEBO text and instead cite the print version only. A quarter of respondents actively teach their students to cite only print. These students are being taught to ignore or disguise their use of digital resources. The responses to this question suggest an additional problem: many (and in the case of the EEBO-TCP-aware audience for the survey, most) researchers want to acknowledge their use of online material but, as there is no single established way of doing so, there is considerable variation in practice. Some cite both print and online sources, some online only, some simply place “[Online]” after their citation. Further illustrative examples can be found in the free-text responses of those who answered “other” to the question above.

25

Examples of “other” answers to the question of how to teach students to cite:

26

- “Cite original print version, using originating library’s citation system (e.g., Huntington Library) and then

indicate that they read the copy on EEBO.”

- “Like in the EEBO guidelines - cite the print version as seen in EEBO, including ref to EEBO as well as the library the digital version is scanned from.”
- “Cite the original and the source they've used (e.g. EEBO, Online)”
- “Cite the original where there is one or the online one where there is not”
- “It really depends on the specific situation”
- “Good question!”

Examples of “other” answers to the question of how researchers do or would cite EEBO-TCP:

27

- “(In the case of my recent book, I had in fact seen physical copies of all the EEBO resources I used.)”
- “Cite the resource, cite the online source name but not URL”
- “Cross-check against image and then cite print version only”
- “Cite fully both electronic and print versions”
- “Due to recent discussions with colleagues, I am in transition as to how I would cite documents obtained from EEBO.”
- “Don't know”

These responses suggest the uncertainty that many feel when confronted with the issue of how to cite their online sources. But the deeper problem remains that many apparently don't want to reveal that they used online sources at all.

28

## The culture of non-citation

The EEBO-TCP user survey suggested that around a third of researchers fail to indicate their use of digital resources at all. Is this indicative of wider practice? What are the reasons for users failing to cite digital material? Why are (some) authors reluctant to cite digital if they can change to a print citation?

29

In order to try to find out more about the culture of citation, in April and May 2013 the authors sent via email a short survey to 60 UK-based print journals covering the fields of literature and history. Representatives from each journal were asked to send their responses also via email. An attempt was made to balance the selection by surveying journals covering different time periods, geographical focus, and thematic approach. In order to maximise the likelihood of reply, the survey asked three simple questions:

30

1. If you receive mss with digital citations do you change them to print if possible or leave them as digital?
2. If you receive mss with print citations do you change them to digital if possible or leave them as print?
3. Do you have a policy on digital versus print citation in your authors' guidelines?

37 replies (a response rate of 62%) were received. 97% said they would not change a print citation to digital, and 78% would not change a digital citation to print. Nine asked for clarification on what was meant by “digital citation.” Some responses treated the question as only pertaining to the DOI of a journal article, not appearing to have in mind citations to online text resources.

31

More interesting than the bald figures are the comments included by some editors with their replies. For a number of journals digital citation by authors was mentioned as a rare or non-existent occurrence, for example, “[our journal] doesn't receive submissions with digital citations, so these questions are not relevant to us.”

32

This raises the question of cause and effect. Do authors eschew digital citation because they think it would be frowned upon? If a journal never prints digital citations then potential authors reading it may think this is a deliberate policy (although the editor mentioned above did not say that they would not include digital citations, only that they do not receive them).

33

So, it may be that journal editors receive few digital citations and researchers rarely see such citations in articles they are using for their research. What assumptions are fuelling this cycle? Why do many shy away from citing digital, whether they be authors or editors?<sup>[9]</sup>

34

Practically, including URLs is seen to be a problem. Many fear that a particular URL will no longer be active in five or ten years' time (see, for example, [European Science Foundation 2011]). While it is true that some URLs might not remain the same, this problem is increasingly being addressed by libraries and archives. 35

Legislation change in 2013 mandated the UK's copyright libraries to capture the UK's public web domain as part of their remit [UK Web archive 2016]. Additionally The National Archives have addressed this problem in their UK Government Web Archive [UKGWA 2016] (prompted by complaints that Hansard references were suffering from link rot) by providing a bridging page that takes users from the defunct URL to the archived web page. 36

The Internet Archive has released a Firefox extension, "No More 404s". If the extension is enabled the browser hitting a broken link will offer to take the user to an archived version of that page in the Wayback Machine [Firefox 2016]. This cannot be comprehensive (because the Internet Archive is not comprehensive) but it offers a broad solution to link rot that only this particular archive is equipped to provide. 37

Harvard Law School Library leads a group of libraries and others in maintaining Perma.cc, a free service which allows anyone with an account to create a link to an archival version of a web page [Perma.cc 2016]. In principle this makes it possible for scholars to create digital citations which will never break. 38

A more pertinent barrier to digital citation might be URLs that are too long and cumbersome. Journal editors and print publishers dislike them because they look ugly and are hard to typeset or format. Academics and students too dislike their appearance, and the fact that, together with a full citation, they can affect the word or page count of a piece of work. They are unattractive for readers more generally. They have often been generated for technical reasons by the content creators with little thought for the needs of eventual users. 39

Philosophically, some researchers may feel that there is little difference between the database where one accesses a text and the library where one reads a book. Such researchers wouldn't cite the library, so why cite a database? Many scholars, especially early modernists, do of course cite the source library. Those who give this as a reason for failure to cite have perhaps never considered the question of whether it is honest to hide their use of digital material, or thought about how such material is funded. We would further argue that there is a clear difference in the reading experience between manuscript, print and digital, and to elide this in citation does an injustice to each [Grafton 2009, 310f] [Edwards 2013]. 40

This "hiding" of the use of digital resources gets to the heart of the problem of non-citation: many believe digital resources are, or are perceived to be, less reliable or not as good as "real" (i.e. conventional print) resources. Some seem to believe that they need to hide their use of digital surrogates in order to appear to be a better scholar (see [Sukovic 2009] for some instructive examples from interviews with researchers). 41

It may be that leading scholars, secure in their jobs and reputations, would be less nervous about citing digital resources, if they think it appropriate. There is some anecdotal evidence that this is the case. Historians Paul Kennedy (internationally famous for *The Rise and Fall of the Great Powers*) and Norman Davies (internationally famous for *The Isles*) have both approvingly cited Wikipedia, sometimes in explicit preference to print sources, in more recent books [Kennedy 2013] [Davies 2012, 143–149]. Michael Screech's *Laughter at the Foot of the Cross*, written as an emeritus professor, mentions his being unable to find a reference despite reading the complete works of St Jerome, but that a "young friend" found the passage easily by using a database [Screech 1997, 70]. But these are senior academics with nothing left to prove. 42

In a trenchant article, the historian Tim Hitchcock points to a deeper scholarly problem that lack of transparency over digital sources is obscuring. Researchers are trained in traditional source materials: libraries, archives, conventional reference works. They are, usually, inexpert in using digital resources: "We have not established the necessary new systems of reference and validation that would make our use of these resources transparent and repeatable." [Hitchcock 2013, 18] 43

Hitchcock is making a broader point than this article seeks to address: digital resources are problematic, often deeply 44



so, and in claiming to use print where they have used digital, researchers are seriously misrepresenting their methodology. In order that scholarly work receive due scrutiny, it is essential that scholars be clear, open and honest about their use of digital resources.

## Solutions? Changing perceptions and practice

What, then, can be done to improve the reputation of digital resources and to encourage users to acknowledge their use? Such change must start with the actual content creators themselves.

Fundamentally, content creators need to make it easy for their users to be open and to properly acknowledge their use of a particular resource. If it is easy to cite a digital resource, more users will do so. Digital resources should make URLs as short as possible and, if possible, human-decodable, and should include a clear link to an automatically-generated citation from the main page of a text, image, or entry. In this way, content creators can make it as easy as possible for their users to cite (or as difficult as possible for them not to), with the result that citation rates should improve. A number of high profile sites, such as the OED, Oxford DNB, Wikipedia, and indeed British History Online, do allow users to automatically generate citations in multiple formats, although even these excellent resources use URLs that are not obviously decodable for readers.<sup>[10]</sup>

Some scholars are concerned about how to cite something that may change or be updated. Digital editors must, therefore, make it clear how to date content accessed via their resource. Release information and/or editorial updates should be made as obvious as possible. By dating digital items in this way, online resource managers can help users feel comfortable about how to clearly refer to the evidence that they are citing and when they are citing it. Sites should encourage or guide users always to give a date of access whenever they cite a digital resource, and should include such a date in automatically generated citations.

Indeed, while this article was being researched and written, British History Online was relaunched with a new citation generator and using a completely new format for its URLs, to make them more immediately meaningful to the user. (British History Online had for a number of years offered citation help, allowing the automatic generation of a citation for any page of content: this has not had a discernible effect on citation habits). As a secondary benefit, if British History Online completely disappeared the new URLs would enable a researcher to trace the URL to a portion of a print book. In practice it might be easier to locate the URL in the UK Web Archive maintained by the British Library on behalf of all copyright libraries in the UK, although this is not currently publicly available; the Internet Archive's crawls of the site are, it seems, not comprehensive.

What was previously a database-generated number has now been converted to a human-readable series and book. Further specificity is provided by the page range of the book; where this is not possible, for born-digital content and dictionary-like material, a meaningful subsection name has been chosen. For example the old URL for the Survey of London, Volume 46, pages 280 to 293 was:

<http://www.british-history.ac.uk/report.aspx?compid=119422>

This has now become:

<http://www.british-history.ac.uk/survey-london/vol46/pp280-293>

An additional advantage here is that the user who only wants the volume level, or the series level, can strip off parts of the URL intuitively:

<http://www.british-history.ac.uk/survey-london/vol46>

<http://www.british-history.ac.uk/survey-london>

This process was carried out semi-automatically for about 100,000 URLs, using the original database and simply concatenating database fields. Although the results are surely a great improvement, the process was not onerous.

There was consultation in the team about the best forms of abbreviations to use for the best trade-off between shortness of URL and clarity. The decisions here were, first, to use standard abbreviations where they exist. For example, for the Victoria County History (known to historians as the VCH) standard county abbreviations were used:

<http://www.british-history.ac.uk/vch/oxon/vol6/pp7-14>

The process could only be semi-automatic because some fields of the database inevitably contained characters which are not allowed in URLs (such as quotation marks in the titles of Acts of Parliament), which had to be located with a regular expression and treated on a case-by-case basis (a side-benefit was that this process exposed some metadata errors which could be fixed). Further, the URLs had to be tested for uniqueness. Some of the new URLs were not unique because they pointed to the single page of the same book. This was most frequently the case with the folio volumes of the journals of the House of Lords and House of Commons, where several days' sittings might occur on the same page, but had been separated for digitisation.

53

For example, page 14 of the print version of the *Journal of the House of Lords*, volume 1, contains very short summaries of sittings on various days in January 1550. Each sitting is a different file on BHO, with a different URL. These have been given a suffixed letter to disambiguate what would otherwise be the same URL:

54

<http://www.british-history.ac.uk/commons-jrnl/vol1/p14a>

<http://www.british-history.ac.uk/commons-jrnl/vol1/p14b>

As this example shows, although page ranges are easily added automatically to URLs, they are not necessarily best practice. Better still would be a meaningful string chosen by an editor: in the Lords and Commons journals this would be the date of sitting; in the VCH example above, the parish or other unit under discussion. This could not be done retrospectively in this case, but can be done conveniently as content is created. Future material digitised on the site will use an editorial decision to create each URL. For example, since relaunch the site has published Proceedings in Parliament 1624, each with a bespoke URL encoding the date of sitting:

55

<http://www.british-history.ac.uk/no-series/proceedings-1624-parl/mar-03>

The new version of the site went live in December 2014 and so it is still too early to say if this change of URL convention will make a difference to the under-citation of BHO discussed above. But one of the impediments given to digital citation has, for this resource, been definitively removed. Indeed a further refinement was added in October 2015, in response to the objection that a page range was too imprecise for an academic reference. When the user mouses over a paragraph of text, a pilcrow appears in the left-hand margin (very faintly, so as not to be distracting); clicking on the pilcrow inserts the relevant paragraph number into the URL bar, allowing a more precise reference than a traditional print one.

56

These practical measures could help users feel more comfortable with the practicalities of digital citation. However, the more philosophical discomfort of according digital materials the same weight as print must also be addressed. Those working in the field of digital content creation are doubtless aware that some digital resources seem to be held in particularly high esteem. Despite the nuances of editorial practice discussed at the beginning of this article, the OED or the Oxford Dictionary of National Biography, for example, are understood to be built on sound scholarship and their medium is considered unimportant. With such examples in mind, other digital content creators could usefully consider how best to promote the scholarly rigour and importance of their own resource as a way of gradually eroding residual beliefs in the lack of respectability of the digital. As a first step, web resources should provide easily accessible editorial documentation at the point of accessing texts and images (rather than solely on project-descriptive websites), enabling users to fully understand the nature of the material they are accessing and the assumptions that they can make about it. By making clear the nature of their materials, content creators encourage researchers to use them in appropriate ways and thereby enhance their own scholarly reputations.

57

All of these mechanisms may help change individuals' scholarly citation practices. In turn, these scholars, should they become teaching academics, will pass on their habits and expectations to their students. The EEBO-TCP user survey

58

asked participants how they prefer to learn about digital resources [Sieftring and Meyer 2013, 12]:

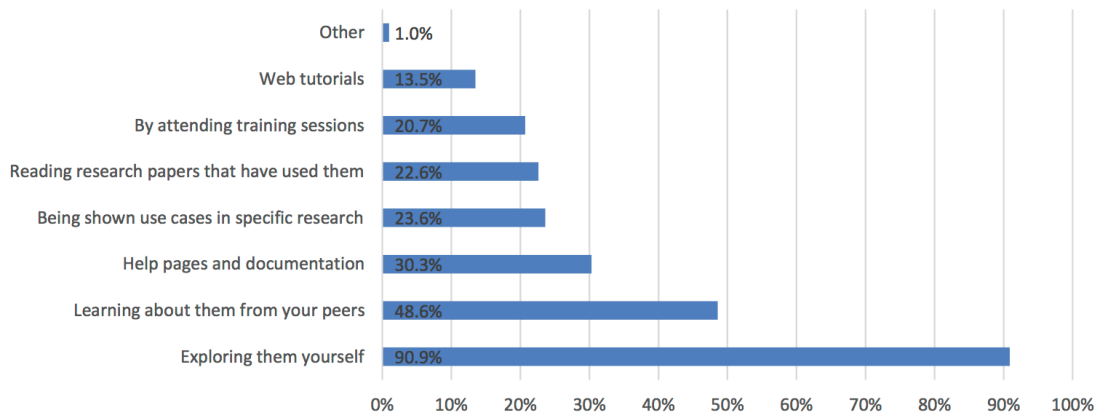


Figure 5. How do you prefer to learn how to use digital resources? N=208

Overwhelmingly participants prefer to explore digital resources themselves or learn about them from peers. Uptake for library training sessions tends to be low. Uptake for web tutorials seems rather low too – but this may be due to lack of publicity or planning for dissemination. However, participants indicated in the free text responses “other” ways of learning about digital resources. Of the 17 free text responses to this question, 11 of them describe recommendations by a tutor, professor or lecturer, at both undergraduate and graduate level. This reveals a particularly important means of learning about online collections: from those who teach [Sieftring and Meyer 2013, 12].

59

Teaching academics play a vital role in disseminating scholarly practices. Online resources could usefully prepare citation guidelines and editorial documentation that could be circulated to academic departments and subject administrators for inclusion in local documentation given to students as they begin their studies. Although beyond the scope of this article, a survey of what guidelines are currently used would provide useful data on the current generation of students and how they are expected to cite. 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. By integrating awareness of the nature of digital resources and the importance of citing them directly into teaching, digital content creators can work together with academics to help shape the practices of the next generation of scholars.

60

An increasing problem for contributors and editors who prefer print citation for journal articles will be that journal publishers themselves are steadily moving towards online-only publication for their journals (for obvious economic reasons). Alice Meadows, of Wiley Publishing, argues in a blog post that one sticking point — the provision of print journals to members as a key benefit of the membership of a learned society — is beginning to become less important as an issue [Meadows 2012]. Another driver will be the moves towards open access in journal publishing: open access presupposes a digital format. Likewise the changes to the UK Research Excellence Framework (in operation after 1 April 2016) will require scholars to deposit all scholarly outputs in an institutional repository (by their nature, primarily digital) [HEFCE 2015].

61

As mentioned above, most survey respondents conceived our questions about digital citation generally as being about DOIs specifically. As also mentioned earlier, this may be a form of *deformation professionelle*: they were thinking as journal editors, so they thought only of journal citation. But it seems plausible too that the ubiquity of DOIs has acted as a forcing function: editors have had to engage with them and make policy decisions about them. There has been no such forcing function for other digital sources, as is evidenced by survey responses such as: “I would change them [digital citations] to print, but have actually not had the need to, since authors seem to choose print anyway.”

62

Online-only journals can, obviously, only be cited digitally. It will become more and more confusing if – as scholars increasingly use three-fold citation types: items which exist only in print, digital versions or surrogates of originally print

63

items, and items which are born-digital – researchers cite some online items and not others. As citing online-only material becomes commonplace, it should become equally commonplace to cite all material accessed online, regardless of whether a print version exists or not.

While active solutions can be undertaken by individual projects and content creators, a gradual shift in culture and practice may already be being led by respected institutions. But as Patrick Dunleavy argues in a thought-provoking blog post, change cannot be effected unilaterally but will be “a long process of driving out legacy citation systems” [Dunleavy 2014]. Contra Dunleavy, however, we would argue that the principle should be: *cite what you used*. This does not preclude giving an additional digital or print citation to aid discovery.

64

Recently the Royal Society announced their move to continual publication, whereby they will give a DOI but no page numbers [Royal Society n.d.]. Such moves will change the focus within digital scholarship. It might help, too, to bear in mind historical precedents. For example, we might consider when it became standard practice to include colophon data or title page printing information during the move from manuscript to print culture. We should also bear in mind that there is an element of serendipity in which citation methods take off and which don't. For example, in classical studies, Stephanus pagination has become a standard way of referencing the work of Plato. The system is based on a 1578 edition of Plato's works by Henri Estienne [Wikipedia 2016]. Estienne is also responsible for the division of Bible chapters into verses that have become so conventional today that many do not realise that verses, and indeed chapters, of the Bible are bibliographic conventions that have nothing to do with the earliest known manuscripts. It was pure chance that these arbitrary systems, eschewing page numbers and even editions, became the standard. It is easy to imagine other, equally arbitrary but effective, citation norms arising for the citation of digital texts.

65

## Conclusion

Change of practice at individual scholarly level reflects and promotes change at a wider cultural level. As more and more established academics are open about their use of online resources, the belief that digital content is less scholarly should lessen, and citation should improve. Open access and changes to the REF in the UK will, as mentioned above, sharpen the imperative to cite digital versions of content. Like the OED, many journals are abandoning print. Young researchers, not just luminaries like Kennedy and Davies, should find it increasingly easy to cite what they used in research without apology; their critics will find their position increasingly untenable.

66

In the meantime, it falls primarily to digital content creators to heighten awareness of the issues involved. Simply raising the issue is often enough – in conversation with them, the authors of this article have found that many scholars have never thought carefully about their digital behaviour and simply require some prompting to reconsider their citation habits. 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<sup>[11]</sup>. We must take the opportunity to look beyond our own specialisms to interdisciplinary knowledge exchange. Text and image-based resources, for example, should look for input from other areas of study where there is philosophical overlap – such as citation for audiovisual material [BUFVC 2013] or for music. Raising awareness of digital citation should become part of the public engagement strategy for all digital resources.

67

Digital citation is important because it is a reflection of how digital resources are valued. It is important because it helps build cases for further funding and enhancement based on evidence of use and impact. It is important because it allows readers of published research to trace and discover sources, both known and new to them, as accurately as possible. It is also honest.

68

We hope and expect that, in time, the currently too widespread practice of citing a print work which has neither been seen nor used will come to seem an unfortunate historical interlude, one in which the practice of scholarly transparency was briefly and lamentably abandoned. Digital resources are here to stay – it is time that they received the credit that is their due.

69

## Acknowledgements

## Notes

[1] "L'effet initial de l'imprimerie n'est pas celui d'une diffusion accrue d'exemplaires identiques d'un même texte mis ainsi dans beaucoup de mains...il est d'abord et surtout la transformation même de l'éthique de la lecture et de l'écriture".

[2] The Oxford English Dictionary (Second Edition), edited by John Simpson and Edmund Weiner, was published by the Clarendon Press in 1989. The first edition was printed in parts between 1884 and 1928, and the dictionary has been revised, updated and supplemented over the years. In 2000, the OED moved from print to online publication, [www.oed.com](http://www.oed.com), and since then the dictionary has been undergoing its first major revision. A concise account of the history of the OED may be found at <http://public.oed.com/history-of-the-oed/> [Accessed November 2015]

[3] One of our peer reviewers pointed out that this definition is derived from Merriam-Webster Online, which had escaped our notice. We are grateful for the information and, more generally, to the comments of the anonymous reviews which have enabled us to improve this article.

[4] Compare, for example, the etymology in the OED's (revised) entry for *omphalos*: "< ancient Greek *ὀμφαλός* navel, centre, hub, round stone in the temple of Apollo at Delphi supposed to mark the centre of the earth, knob or boss (ultimately cognate with navel n.). In quot. 1847 at sense 2a via German *Omphalos* (1830 in the source translated)" [OED 2004].

[5] For example, [Hajjem et al. 2005], which does not include a humanities subject in its 10 candidate disciplines.

[6] This first public release amounted to 25,363 texts. Eventually, around 70,000 texts will be made freely available.

[7] It is notable that only 1% of respondents to an EEBO-TCP user survey (discussed in more detail below) identified themselves as undergraduates, compared to, for example postgraduate students (38%), professors (19%) and academic researchers (20%).

[8] A full discussion of the TIDSR analysis of EEBO-TCP can be found in [Siefiring and Meyer 2013].

[9] The views on digital citation expressed in this and subsequent paragraphs reflect the thoughts and opinions of scholars, editors, publishers and digital humanities specialists expressed at a number of forums, chiefly in relation to Blaney's paper on citing digital at the Sheffield Digital Humanities Congress in September 2012 [Blaney 2012], a digital citation focus group held by Siefiring in Oxford in November 2012, and Blaney & Siefiring's joint presentation on assessing the impact of digital resources at the Digital Humanities@Oxford Summer School in July 2013 [Blaney and Siefiring 2013].

[10] The URL included in the Oxford DNB's automatically generated citation for the writer Thomas Churchyard (1523?-1694) is <http://www.oxforddnb.com/view/article/5407>, for example, while Wikipedia's entry for the same writer has the URL [http://en.wikipedia.org/w/index.php?title=Thomas\\_Churchyard&oldid=576213384](http://en.wikipedia.org/w/index.php?title=Thomas_Churchyard&oldid=576213384) [both accessed November 2015].

[11] An excellent example is Victoria Van Hyning's blog post written in response to this article's authors' joint presentation at the Digital Humanities@Oxford Summer School in July 2013 [Van Hyning 2013] [Blaney and Siefiring 2013].

## Works Cited

**BHO 2015** British History Online 2015, University of London and the History of Parliament Trust. Available from <http://www.british-history.ac.uk/> [14 March 2016].

**BUFVC 2013** British Universities Film & Video Council 2013 "Audiovisual Citation Guidelines." Available from <http://bufvc.ac.uk/projects-research/avcitation> [11 March 2016].

**Blaney 2012** Blaney J. 2012 "The Problem of Citation in the Digital Humanities." Available from <http://www.hrionline.ac.uk/openbook/chapter/dhc2012-blaney> [11 March 2016].

**Blaney and Siefiring 2013** Blaney J. and Siefiring J. 2013 "EEBO-TCP Measuring Impact and Making Changes." Available from <http://podcasts.ox.ac.uk/04eebo-tcp-measuring-impact-and-making-changes> [14 March 2016].

**Compagnon 1979** Compagnon, A. *La seconde main - ou le travail de la citation*, Éditions du Seuil, Paris (1979).

**Daily Telegraph 2014** Flanagan, P. "RIP for OED as world's finest dictionary goes out of print." Available from <http://www.telegraph.co.uk/culture/culturenews/10777079/RIP-for-OED-as-worlds-finest-dictionary-goes-out-of-print.html> [13 September 2016].

- Dalbello et al. 2006** Dalbello, M. Lopatovska I., Mahony P. & Ron N. "Electronic texts and the citation system of scholarly journals in the humanities: case studies of citation practices in the fields of classical studies and English literature." Available from [http://arizona.openrepository.com/arizona/bitstream/10150/105648/1/Dalbello\\_posterrev.pdf](http://arizona.openrepository.com/arizona/bitstream/10150/105648/1/Dalbello_posterrev.pdf) [11 March 2016].
- Davies 2012** Davies D. *Vanished Kingdoms: The History of Half-Forgotten Europe*, Penguin, London (2012).
- Dunleavy 2014** Dunleavy P. 2014 "Academic citation practices need to be modernized so that all references are digital and lead to full texts." Available from <http://blogs.lse.ac.uk/impactofsocialsciences/2014/05/21/academic-citation-practices-need-to-be-modernized/> [11 March 2016].
- EEBO-TCP 2015** Early English Books Online Text Creation Partnership 2015, Bodleian Libraries, University of Oxford. Available from <http://www.bodleian.ox.ac.uk/eebotcp/> [14 March 2016].
- Edwards 2013** Edwards A. S. G. 2013 "Back to the real?." *The Times Literary Supplement*, 7 June 2013. Available from <http://www.the-tls.co.uk/tls/public/article1269403.ece> [11 March 2016].
- European Science Foundation 2011** European Science Foundation 2011 "Science Policy Briefing 42: Research Infrastructures in the Digital Humanities." Available from [http://www.esf.org/fileadmin/Public\\_documents/Publications/spb42\\_RI\\_DigitalHumanities.pdf](http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf) [11 March 2016].
- Firefox 2016** "No More 404s." Available from <https://testpilot.firefox.com/experiments/no-more-404s/> [12 October 2016].
- Grafton 2009** Grafton A. 2009, "Codex in Crisis: the Book Dematerializes" in *Worlds Made By Words: Scholarship and Community in the Modern West*, Harvard University Press, Cambridge MA (2009).
- HEFCE 2015** Higher Education Funding Council for England 2014 "Policy for open access in the post-2014 Research Excellence Framework." Available from <http://www.hefce.ac.uk/pubs/year/2014/201407/> [11 March 2016].
- Hajjem et al. 2005** Hajjem C., Harnad S. & Gingras Y. "Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact," *IEEE Data Engineering Bulletin* 28.4 (2005), 39-47. Available from <http://arxiv.org/ftp/cs/papers/0606/0606079.pdf> [11 March 2015].
- Hitchcock 2013** Hitchcock T. "Confronting the Digital: or How Academic History Writing Lost the Plot," *Cultural and Social History*, 10.1 (2013), 9-23.
- Kennedy 2013** Kennedy P. *Engineers of Victory: The Problem Solvers Who Turned the Tide in the Second World War*, Allen Lane, London (2013).
- Larivière and Gingras 2009** Larivière, V. and Gingras Y. "The decline in the concentration of citations, 1900–2007" *Journal of the American Society for Information Science and Technology*, 60.4 (2009), 858-862. Available from <http://arxiv.org/ftp/arxiv/papers/0809/0809.5250.pdf> [11 March 2016].
- Lozano et al. n.d.** Lozano, G. A., Larivière, V. & Gingras Y. "The weakening relationship between the Impact Factor and papers' citations in the digital age." Available from <http://arxiv.org/ftp/arxiv/papers/1205/1205.4328.pdf> [11 March 2016].
- Meadows 2012** Meadows A. 2012 "Moving Scholarly Society Members Online-Only – Are We Reaching the Tipping Point?." Available from <http://scholarlykitchen.sspnet.org/2012/12/13/moving-scholarly-society-members-online-only-are-we-reaching-the-tipping-point/> [11 March 2016].
- OED 1933** hubris, n. Available from <http://www.oed.com/view/Entry/89081> [11 March 2016].
- OED 2004** omphalos, n. Available from <http://www.oed.com/view/Entry/131290> [11 March 2016].
- Perma.cc 2016** "About Perma.cc." Available from <https://perma.cc/about> [12 October 2016].
- Royal Society n.d.** *Continuous publication*. Available from <https://royalsociety.org/journals/authors/continuous-publication/> [11 March 2016].
- SECT 2013** SECT: Sustaining the EEBO-TCP Corpus in Transition, 2013. Available from <http://www.bodleian.ox.ac.uk/eebotcp/SECT/> [14 March 2016].
- Screech 1997** Screech M. *Laughter at the Foot of the Cross*, Allen Lane, London (1997).
- Siefring and Meyer 2013** Siefring J. and Meyer E. *Sustaining the EEBO-TCP Corpus in Transition: Report on the TIDSR Benchmarking Study* (2013). JISC, London (2013). Available from SSRN: <http://ssrn.com/abstract=2236202> or <http://dx.doi.org/10.2139/ssrn.2236202> [11 March 2016].

**Sukovic 2009** Sukovic, S. "References to e-texts in academic publications," "Journal of Documentation", 65.6 (2009), 998-1015.

**TIDSR n.d.** "TIDSR: Toolkit for the Impact of Digitised Scholarly Resources." Available from <http://microsites.oii.ox.ac.uk/tidsr/> [11 March 2016].

**UK Web archive 2016** UK Web Archive 2016. Available from <http://www.webarchive.org.uk/ukwa/> [11 March 2016].

**UKGWA 2016** UK Government Web Archive. Available from <http://www.nationalarchives.gov.uk/webarchive/> [11 March 2016].

**Van Hying 2013** Van Hying, V. 2013 "Citing and Contributing to Digital Resources." Available from <http://snakeweight.wordpress.com/2013/07/22/citing-and-contributing-to-digital-resources/> [11 March 2016].

**Webster and Blaney 2011** Webster, P. and Blaney, J. 2011, "The Impact and Embedding of an Established Resource: British History Online as a Case Study." Available from <http://sas-space.sas.ac.uk/2819/> [11 March 2015].

**Wikipedia 2015** *Wikipedia*, "Stephanus Pagination." Available from [https://en.wikipedia.org/wiki/Stephanus\\_pagination](https://en.wikipedia.org/wiki/Stephanus_pagination) [11 March 2016].

**Wikipedia 2016** *Wikipedia*, "Hubris." Available from <https://en.wikipedia.org/wiki/Hubris> [11 March 2016].



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.