Measuring the web impact of digitised scholarly resources

Kathryn E. Eccles
Oxford Internet Institute, University of Oxford, Oxford, UK

Mike Thelwall
Statistical Cybermetrics Research Group, University of Wolverhampton, Wolverhampton, UK, and

Eric T. Meyer
Oxford Internet Institute, University of Oxford, Oxford, UK

Abstract

Purpose – Webometric studies, using hyperlinks between websites as the basic data type, have been used to assess academic networks, the "impact factor" of academic communications and to analyse the impact of online digital libraries, and the impact of digital scholarly images. This study aims to be the first to use these methods to trace the impact, or success, of digitised scholarly resources in the Humanities. Running alongside a number of other methods of measuring impact online, the webometric study described here also aims to assess whether it is possible to measure a resource's impact using webometric analysis.

Design/methodology/approach – Link data were collected for five target project sites and a range of comparator sites.

Findings – The results show that digitised resources online can leave traces that can be identified and used to assess their impact. Where digitised resources are situated on shifting URLs, or amalgamated into larger online resources, their impact is difficult to measure with these methods, however.

Originality/value – This study is the first to use webometric methods to probe the impact of digitised scholarly resources in the Humanities.

Keywords Webometrics, Digitised scholarly resources, Humanities, Impact, Digital documents, Information management

Paper type Research paper

1. Introduction

Measures of online impact are becoming increasingly important as scholars spend more time on the web. Online methods of judging academic impact have begun to add context to long established methods of measuring impact, such as journal-based citation indexes, particularly in the social sciences (Kousha and Thelwall, 2007a). Studies have assessed the value of different web sources for impact assessment such as Google (e.g. Kousha and Thelwall, 2007b; Vaughan and Shaw, 2003; Vaughan and Shaw, 2005), Google Scholar (e.g. Harzing and van der Wal, 2009; Kousha and Thelwall, 2008; Meho and Yang, 2007), Google Books (Kousha and Thelwall, 2009), and a range of formal and informal academic outputs (Kousha *et al.*, 2010). Assessing the impact of digital resources, however, is more complex. The need for such measures is pressing, as investment in digitised scholarly resources is heavy and continues to grow, at least in the UK. While enthusiasm for these resources appears to be high,

funding councils, project staff and other stakeholders need to gain a more concrete sense of what impact they are having, and where.

The study described here was funded by the Joint Information Systems Committee (JISC). It was designed to probe the usage and impact of five specific digitised scholarly resources funded by JISC in its first phase of digitisation[1]. The five resources were: Histpop – Online Historical Population Reports; 19th Century British Library Newspapers (phase one); British Library Archival Sound Recordings (phase one); 18th Century Official Parliamentary Publications Portal 1688-1834 at the British Official Publications Collaborative Reader Information Service (BOPCRIS); and the Wellcome Medical Journals Backfiles. The wider project of judging the usage and impact of these projects encompassed a wide range of quantitative and qualitative research methods that were tested for their suitability regarding digital scholarly resources (Meyer *et al.*, 2009). The webometric study described here was one of these methods. There were two main advantages in using webometrics in this context. Firstly, it is relatively easy to acquire this data, whereas a log file analysis (which would yield some similar results) requires access that is not always available to researchers. Secondly, webometric methods allow researchers to compare the target site(s) with a range of other sites, in order to benchmark the results from the main site(s) of interest. This method enables researchers to gain a more comprehensive view of the target site(s) as part of an ecology of digital resources.

2. Related studies
There have been numerous attempts to measure the impact and reach of online scholarly resources. The LAIRAH project at University College London (UCL) (Warwick *et al.*, 2006, 2008), MIT's OpenCourseWare evaluation (Carson, 2006), the Open Educational Resources Report (Atkins *et al.*, 2007), the Berkeley Digital Resource study (Harley, 2007a, b; Harley *et al.*, 2006), and the Rice University Fondren Library Impact of Digital Resources on the Humanities study (Spiro and Segal, 2007). The Wolverhampton-Loughborough digital repository analysis project for JISC (Zuccala *et al.*, 2007) used web link analysis in combination with interview and survey data to shed light on the web networks created by digital repositories. These repositories differed from the digitisation projects which formed the focus of our study in that they are places where users can both access and store data. Although the above projects used a range of methods, only the last one used webometrics and none evaluated sites against other sites. Hence, none of the projects could assess from an evaluative rather than an informative perspective. This is a problem because funding agencies in particular need to know whether a project could be regarded as successful or not.

A number of webometric studies have found limitations with the use of webometric data. Most importantly, hyperlinks can be created for a wide variety of reasons, both positive and negative, and therefore do not necessarily indicate that the person creating the link endorses or uses the link target (Bar-Ilan, 2004; Thelwall, 2003; Thomas and Willet, 2000; Wilkinson *et al.*, 2003). A second issue is that links may be created by computer programs or automatically replicated as part of web design decisions (Thelwall, 2002). Nevertheless, a number of studies of links targeting universities have shown that counts of links typically reflect a tangible aspect of the target: its research productivity (Li *et al.*, 2005; Smith and Thelwall, 2002; Thelwall and Harries, 2004). Similarly, studies of links to online academic publications or journal websites have shown that these tend to correlate with offline measures, such as citations (Kousha and

Thelwall, 2007c; Vaughan and Hysen, 2002). Taken together, these findings suggest that links to academic repositories should give useful indicators of their uptake and value, although results should be interpreted cautiously.

3. Research questions

The primary research question here was to investigate whether it is possible to measure one aspect of a resource's impact or impacts using webometric analysis. This method cannot show the whole picture of the impact of a web-based resource, but it does provide one view of the impact of the resource. Web links can be seen as indicators of intellectual impact, since links indicate awareness and uptake. In order to produce a more nuanced picture of impact (asking for example, whether such resources allow researchers to produce better research, write more books or papers, or ask different questions), a mixed methods approach would be more appropriate. This was the remit of the wider study, from which this webometric study was drawn, and which sought to analyse a range of methods for judging the usage and impact of digitised scholarly resources[2]. Having completed such an analysis, successful methods were reviewed and recommended to others in an online toolkit. The webometric study was therefore designed to test not only the five specific resources that formed the core of the wider study, but to test the usefulness of webometric analysis as an indicator of impact of digitised scholarly resources in general. Factors relating to access were also important, as funders want to ensure the greatest access to target audiences, but those hosting and maintaining digitised resources are bound to consider how to secure their future, and in many cases, had chosen to partner with commercial content providers who have the resources to host data and deliver continued support to users. Several of our target projects had forged such partnerships, with the result that access was largely limited to users from within British HE and FE institutions. One aspect of our webometric, as well as our wider study, was to probe the extent to which issues of access affected the impact, and the ability to trace the impact, of our target resources.

4. Methods and procedures

The methodology was a comparative link analysis, investigating the links to each of our digital resources in comparison with the links to a set of comparator websites (details of comparator websites are given in Table I, and further details can be found in note [3]). In selecting the comparator sites, we tried to include one digital scholarly resource that matched the target project as closely as possible, and to include the URLs of the analogue collections from which the digitised materials were drawn, and locations where the digital materials were duplicated. We also tried to control for open vs closed access, and to look further afield at international resources, but due to the nature of digitisation efforts, many of our resources were exemplars in their fields, and therefore the selection of suitable comparators was a somewhat inexact science. Nevertheless, these comparator sites provided an important context for the results of our target sites. We aimed to judge the extent to which our five target projects, and the selected comparator sites, had generated links to their sites in the period from their launch to the time of the study. Data was collected automatically using a free program (see http://lexiurl.wlv.ac.uk/ for more information). The software is designed to collect search engine data (in this case from Yahoo!) and returns the data in a series of reports. For all target and comparator projects, we excluded inlinks from home sites (e.g. we

Target project	Comparator	Search queries (correct at October 2008)
Histpop	Henry III Fine Rolls Project	link: www.frh3.org.uk site: frh3.org.uk
	Historical Directories	link: www.historicaldirectories.org/hd/index.asp site: historicaldirectories.org
	Contemporary and Historical Census Collections (CHCC)	link: http://hds.essex.ac.uk/history/data/chcc.asp site: hds.essex.ac.uk
	National Archives Census Data	link: www.nationalarchives.gov.uk/census site: nationalarchives.gov.uk
19th Century British Library Newspapers	Library of Congress: Chronicling America	link: www.loc.gov/chroniclingamerica site: loc.gov
	British Periodicals	link: http://britishperiodicals.chadwyck.co.uk/home.do site: chadwyck.co.uk
	British Library Penny Illustrated Paper	link: www.collectbritain.co.uk/collections/pip site: collectbritain.co.uk
British Library Archival Sound Recordings	Imperial War Museum Sound Archive	link: www.iwmcollections.org.uk/qrySound.asp site: iwm.org.uk site: iwmcollections.org.uk
	King's College London, King's Sound Archive	link: www.kcl.ac.uk/kis/schools/hums/music/ksa site: kcl.ac.uk
	Cornell Lab of Ornithology/Macaulay Library Animal Behaviour Sound Archive	link: http://animalbehaviorarchive.org/loginPublic.do site: animalbehaviorarchive.org
18th Century Official Parliamentary Publications	House of Commons Parliamentary Papers	link: http://parlipapers.chadwyck.co.uk/home.do site: chadwyck.co.uk
	Proceedings of the Old Bailey, 1674-1913	link: www.oldbaileyonline.org site: oldbaileyonline.org
	The Diary of Samuel Pepys	link: www.pepysdiary.com/about site: pepysdiary.com
Wellcome Medical Journals Backfiles	Medline	link: http://medline.cos.com site: medline.cos.com
	Internet Library of Early Journals (ILEJ)	link: www.bodley.ox.ac.uk/ilej site: bodley.ox.ac.uk
	Science in the 19th Century Periodical	link: www.sciper.org site: sciper.org
	Nineteenth Century Serials Edition	link: www.ncse.ac.uk site: ncse.ac.uk

Table I.
Comparator sites and queries

excluded all links to the British Library projects from the British Library site), only counting inlinks from other sources. Data was collected twice, in October 2008, and again at the end of the project in April 2009, in order to detect whether any of our results had changed significantly over time. Data presented in the results section below is from October 2008. Differences in the end-of-project data are discussed in the text.

4.1 Selection of project URLs

While the selection of project URLs should have been a straightforward aspect of the study, our target project sites presented a number of challenges. Histpop and the British Library Archival Sound Recordings were the least complicated of our projects, as both had a URL independent of their origins (e.g. for Histpop, the University of Essex and the UK Data Archive) and a website where users were able to directly access the digital content. The link data collected for these sites therefore represents as accurate a picture as possible on the impact of these resource based on webometric analysis. Both the 19th Century British Library Newspapers and the Wellcome Medical Journals Backfiles projects had a project page on the parent library site, while the digital content was accessed through established databases elsewhere (Gale Cengage and PubMed Central respectively). It was important in these cases, therefore, to capture link data to the project page and to the digital content where possible. Finally, the 18th Century Official Parliamentary Publications was launched on the University of Southampton website (soton.ac.uk), and continued to operate from this location throughout the project, but was also present as a part of the collaborative British Official Publications Collaborative Reader Information Service (BOPCRIS, bopcris.ac.uk). It was decided that, as many scholars would know the project/resource as "BOPCRIS" and access it through this gateway, we would run queries for both. Our project's digital content was freely available at the original location on the soton.ac.uk site and was also available at the House of Commons Parliamentary Papers, at a restricted (ProQuest) site. The ProQuest version contained an improved version of the eighteenth century material alongside (and cross-searchable with) their nineteenth and twentieth century parliamentary papers, and was available to any user with a (paid) subscription and access was secured by JISC for users of British HE and FE institutions. We therefore added this latter site as a comparator, to assess links to both locations.

Full details of the URLs used can be seen in the Tables I-VI[4], including the comparator sites. The tables also give the search queries submitted to Yahoo!: note that this type of search is likely to be phased out by Yahoo! as part of its takeover by Microsoft.

4.2 Problems and limitations

Having resolved the difficulties surrounding project URLs, we discovered several further complications when running webometric analysis on our target sites. First, two of our projects shifted their URLs during the testing (Wellcome Medical Journals Backfiles and the 19th Century British Library Newspapers). This made our analysis more complicated, as we had to collate data on the old and new URLs. The British Library website was overhauled during the project, which meant that not only did we lose the URL for the 19th Century British Library Newspapers project, but also for a

Project search queries	Comments
link: www.histpop.org	Histpop: short URL
site: histpop.org	
link: www.histpop.org/ohpr/servlet	Histpop: full URL (displayed when the page is
site: histpop.org	reached)

Table II.
Histpop search queries

Project search queries	Comments	
link: www.bl.uk/collections/ britishnewspapers1800to1900.html site: bl.uk	Project page (until October 2008)	
Link: www.bl.uk/reshelp/findhelprestype/news/ newspdigproj/paperdigit.html site: bl.uk	Project page (after October 2008	
link: www.bl.uk/collections/newspapers.html site: bl.uk	URL of original resource (until October 2008)	
Link: www.bl.uk/reshelp/findhelprestype/news/ blnewscoll/index.html site: bl.uk	URL of original resource (after October 2008)	**Table III.** 19th Century British Library Newspapers search queries

Project search queries	Comments	
link: http://sounds.bl.uk site: bl.uk	Project URL	
link: www.bl.uk/collections/sound-archive/nsa.html site: bl.uk	Homepage of original resource (Sound Archive)	
link: www.bl.uk/collections/sound-archive/cat.html site: bl.uk	Catalogue search page of original resource (until October 2008)	**Table IV.** British Library Archival Sounds Recordings search queries
link: http://cadensa.bl.uk/cgi-bin/webcat site: bl.uk	Catalogue search page of original resource (since October 2008)	

Project search queries	Comments	
link: www.parl18c.soton.ac.uk/parl18c/digbib/home site: soton.ac.uk	Link to the C18th resource	
link: www.bopcris.ac.uk/18c site: bopcris.ac.uk	Alternative URL, redirects to the above URL	**Table V.** 18th Century Official Parliamentary
link: www.bopcris.ac.uk site: bopcris.ac.uk	BOPCRIS (main site) home page	Publications Portal search queries

Project search queries	Comments	
link: http://library.wellcome.ac.uk/node280.html site: wellcome.ac.uk	Project (actual) homepage	
link: http://library.wellcome.ac.uk/backfiles site: wellcome.ac.uk	Project (given) homepage	**Table VI.** Wellcome Medical
link: www.pubmedcentral.nih.gov site: pubmedcentral.nih.gov	Digital Archive, held at Pub Med Central	Journals Backfiles search queries

number of our comparator projects which were also located at the British Library website. A second complication was that for several of our projects, the advertised URL is automatically redirected to a different URL. For example, the Wellcome Medical Journals Backfiles URL was advertised as http://library.wellcome.ac.uk/backfiles/ but the redirected page was http://library.wellcome.ac.uk/node280.html. When someone creates a link, they will either right-click a hyperlink and copy the shortcut, or they will simply highlight the URL of the page they are visiting as displayed in the address bar, copy and save this link. If these URLs differ, it is important to analyse links to both.

Finally, search engines return a maximum of 1,000 results. In our results, then, any URL count of over 900 was likely to be a significant underestimate. It is important to note that when a result of over 900 is returned, we can assume that the search engines know of more results, but are not returning them all.

4.3 Definition of terms
Note that domains is the most reliable of these measures, and is highlighted in each of the tables below:

- *Domain*. The domain name of an URL – typically the segment of the URL after http:// and before any subsequent slash.
- *Site*. The distinguishing end of the domain name of an URL (e.g. microsoft.com, ox.ac.uk).
- *STLD*. The second level domain (when existing, otherwise the top level domain:, e.g. co.uk is a STLD) of an URL.
- *TLD*. The top level domain of an URL.
- "Inlinks" refers to the links to a site or page from different web sites.

5. Results
Successful data analysis depended to a large extent on the quality of the webometric data collected. The two projects with the most straightforward URLs (for the purposes of this study) provide the clearest results. Histpop performed well in comparison to its closest comparator, Historical Directories. This is particularly significant given that Historical Directories was launched in 2005, more than two years earlier than Histpop, giving it substantially longer to have attracted links and users. Histpop also performed well in comparison to the National Archives Census Data site, a popular resource for family historians seeking personal information about their ancestors, although as a caveat, we would perhaps have expected the latter resource to have attracted more inlinks.

Analysis of the secondary and top-level domains of pages linking to Histpop show that 30 per cent came from the.com domain, 18 per cent from the ac.uk domain (academic pages based in the UK). End of project data shows that the number of links to the main Histpop page (www.histpop.org) dropped, from 122 domains to 99. The number of links to the full URL as displayed in the browser window, however, continued to rise slowly, from 92 to 96 domains. This suggests that the number of links created by users continues to grow. The comparator sites showed a similar steady rise in numbers of links by domains (Table VII).

The British Library Archival Sound Recordings project also performed strongly, registering significantly more links that selected comparator sound archive digital archives at the Imperial War Museum, King's College London and Cornell University.

However, it is important to point out that finding comparable resources to the highly eclectic collection at the British Library Archival Sound Recordings was extremely difficult. Comparing links to the digital collection to those accumulated by the catalogue and home pages of the original (analogue) collection produced a more meaningful result. The digital resource had attracted far fewer links than the pages relating to the analogue materials, a result that should perhaps not surprise us given the short time that the digital resource had been available compared to the more established analogue resource. By comparing links to the digital resource with links to pages about the original British Library Sound Archive webpages, we were able to see that one of the pages with the most inlinks was the British Library Sound Archive catalogue. While there was a link to the digitised material on the Sound Archive home page, there was no such link or mention of the digital content on the Sound Archive catalogue page. Users who had created the links directly to the Sound Archive catalogue were therefore not seeing links to the digitised material, potentially lessening its impact.

The British Library Sound Archive catalogue page was moved in a large restructuring of the British Library website in October 2008. As the table above shows, the number of links to the original page was very large, compared to the small number of links registered for the new page in April 2009, just 3 per cent of the number of links to the original page. In contrast, the British Library Archival Sound Recordings page showed a significant increase in the number of links, rising from 131 domains to 229. The project was at the time of the study releasing new material as part of a renewed investment from JISC, which may account for its increased impact. Of the comparator projects, both King's College London's Sound Archive and the Imperial War Museum Sound Archive retained the links that they had built up, but did not increase the number. The Cornell Lab of Ornithology/Macaulay Library Animal Behaviour Sound Archive showed an increased number of links, rising from 75 domains to 102 (Table VIII).

	URLs	Domains	Sites	STLDs	TLDs
Histpop	142	122	106	21	17
Histpop (2)	114	92	87	18	13
Henry III Fine Rolls	13	12	11	6	6
Historical Directories	170	142	138	19	13
CHCC	4	3	3	1	1
National Archives CD	288	237	226	24	15

Table VII.
Overview of results:
Histpop

	URLs	Domains	Sites	STLDs	TLDs
BL Archival Sound Recordings	162	131	123	18	14
British Library SA home	542	450	421	52	43
British Library SA Catalogue (before October 2008)	814	733	694	66	49
British Library SA Catalogue (October 2008-April 2009)	47	37	37	17	14
Imperial War Museum SA	13	11	11	5	3
King's College London SA	8	8	8	6	5
Cornell/Macaulay SA	99	75	69	15	12

Table VIII.
Overview of results:
British Library Archival
Sound Recordings

For two of our project sites, 19th Century British Library Newspapers, and the Wellcome Medical Journals Backfiles, we counted links to both the project pages hosted on the parent library website, and the location of the digital content, where possible, as this was hosted elsewhere. It was not possible to collect link data to the digital content in the case of the 19th Century British Library Newspapers, which at that time operated only secured access via individual institutional library portals, or free access on site at the British Library. The project page for the 19th Century British Library Newspapers project had a surprisingly high number of links, given that this page did not contain a link to the digitised material, simply displaying information about the project and details of the content. This perhaps points to a wider impact than log files or analytics might show, as it reveals a community of interested parties who may or may not have regular access to the resource. At the time the data was collected, the project was in receipt of a second phase of funding from JISC to produce a further 1 million pages of content, which may explain high levels of interest in the project itself. Analysis of the URLs linking to the 19th Century British Newspapers project page revealed a number of blog posts (13 of 187, 7 per cent) to this page and to the Penny Illustrated Paper at Collect Britain (5 of 105, 5 per cent), which suggests that these projects and their resources were well known in the blogosphere. As with our other British Library project, link counts to information pages about the original (analogue) materials were high, indicating that links to information about these resources were firmly embedded in the community. The link count for the (analogue) British Library Newspapers information page was higher than that for one of our comparator digitised projects, Chronicling America, an American digital newspaper resource hosted by the Library of Congress which at the time of our study, allowed you to search and view newspaper pages from 1890-1910 and find information about American newspapers published between 1690-present. Projects may in future consider using webometrics to find out how well established their existing collections are, and use this information to decide where to situate links to the (new) digital collections, therefore maximising potential impact.

The 19th Century British Library Newspapers project was one of two resources that created content that was largely only accessibly by British HE and FE institutions. Compared to a single open access digitised 19th Century British Library newspaper, the *Penny Illustrated Paper*, our project site performs less well, with fewer than twice as many links, perhaps indicating the impact of restricted access to content on the number of pages linking to a resource.

A striking result from the link analysis for the 19th Century British Library Newspapers is that the variety of top and second-level domains linking to our project is considerably higher than other British projects, pointing to a more varied and international network of links. Of the links that can be attributed to particular countries (leaving out.com,.org,.net as these cannot be definitely attributed to a particular country), 25 per cent came from the UK, but 26 per cent came from outside the UK, an important result when one considers that this resource was not available outside the UK, and within the UK was only available to users within the HE/FE communities.

The end-of-project data showed that there were a very small number of links to the new project page at the British Library, with only four domains linking to the page, two of which originated from our research project. This shows that the impact that this

project had created was subsequently lost through the relocation of the pages. The *Penny Illustrated Paper* was amalgamated into the 19th Century British Library Newspaper digital resource, after which we were unable to trace its individual impact. The two remaining comparator sites showed a steady increase in the number of links as measured through domains, revealing a steady increase in web impact as expected (Table IX).

In the case of the Wellcome Medical Journals Backfiles, we were able to collect link data for the site where the digitised material was deposited, but this site was Pub Med Central (www.ncbi.nlm.nih.gov/pmc/), a large and heavily used digital archive of biological and life science journal literature, where the backfiles simply formed part of the content available. As the material could not be isolated from this existing archive, link data for Pub Med Central was of limited use (Table X).

The Wellcome project page attracted a high number of inlinks considering that the digital material was housed elsewhere. However, in contrast to the 19th Century British Library Newspapers page, the Wellcome Medical Journals Backfiles project page does contain a link to the digitised material, and the digitised material is freely available to download at this link. When compared to a stand-alone site containing digitised backfiles of similar journals, such as the Internet Library of Early Journals (ILEJ), our project page performs comparatively poorly, suggesting that sites which house information and digital content are more likely to be linked to and show greater web impact when measured in this way. The results for Pub Med Central are perhaps irrelevant to this study, as it inevitably attracts large numbers of inlinks due to the wide range of its coverage and its relevance to the medical sciences. These results reinforce the fact that projects inserting digital content into existing (and particularly high profile, heavily used) digital repositories are much harder to track using webometric methods.

The results for the comparator site the Internet Library of Early Journals indicate the kind of results we can see when the digital content is held on a URL that could be easily analysed for inlinks. This content was associated with the Bodleian Library,

	URLs	Domains	Sites	STLDs	TLDs
19th Century BL Newspapers	187	147	141	31	24
British Library Newspapers	897	755	711	60	44
Chronicling America	860	662	594	74	34
British Periodicals	13	11	10	3	3
BL Penny Illustrated	105	83	80	16	10

Table IX.
Overview of results: 19th Century British Library Newspapers

	URLs	Domains	Sites	STLDs	TLDs
Wellcome (actual) project page	62	45	43	16	14
Wellcome (given) project page	23	18	15	8	8
Pub Med Central	989	783	687	56	41
Medline	718	642	597	80	60
ILEJ	903	725	641	64	42
Sciper	114	91	79	15	14
NCSE	64	47	43	12	8

Table X.
Overview of results: Wellcome Medical Journals Backfiles

making it a useful comparator site for the material created by the Wellcome Library, a similarly prestigious and well known library. The ILEJ records a very large number of inlinks, from a widely diverse set of secondary and top level domains, suggesting that this project has made a considerable impact on a large research community.

End-of-project data from the Wellcome project page indicates the impact of moving URLs of project pages. The given project page registers a small increase in the number of links, rising from 18 domains to 24. The (new) actual project page registers only two links, one from JISC, and one from a library blog at Llandrillo (FE) College in North Wales. The comparator projects registered a small increase in the number of links, reflecting a growing web presence[5].

Finally, the 18th Century Official Parliamentary Publications delivered content in two separate locations; free of charge as part of the British Official Publications Collaborative Reader Information Service (BOPCRIS), University of Southampton, and as part of an enhanced digital resource at ProQuest where access was restricted to members of the British HE/FE community (Table XI).

The number of links to our project's page at BOPCRIS was relatively low. Analysis of the secondary and top-level domains linking to the project page reveals that the majority of these inlinks came from the UK academic community, with 75 per cent registered to the ac.uk domain. The number of inlinks to the alternative project page (NB: not the URL that appears in the browser window) was considerably higher. When we analysed the links to the alternative URL, we found that of the 56 URLs returned by the search software, 46 (82 per cent) were from libraries, academic resource pages, or JISC related sources, suggesting that these links originated from the launch of the project, rather than from users creating links (as we hypothesise that users would be more likely to copy the URL displayed in the browser window). This result suggests that the impact of the project (as measured through this URL alone) is directly related to its promotion by JISC and through university library pages.

The number of links to the subscription access website (ProQuest's House of Commons Parliamentary Papers) was higher than the number of links to our project's BOPCRIS page. Of the links that we could identify by country (again omitting .com, .org and .net), 66 per cent came from the UK, and 25 per cent were from outside the UK. In comparison, the BOPCRIS homepage received a far larger number of inlinks from a wide variety of secondary and top-level domains, including 25 per cent from the .ac.uk domain, showing strong coverage from the UK academic community. The large number of inlinks to the BOPCRIS homepage shows the web impact gained by a project when linking to existing well known and well linked-to digitised resources. Creators of digitised resources would be well advised to use webometric data to identify such resources when deciding where to situate their content.

	URLs	Domains	Sites	STLDs	TLDs
BOPCRIS: 18th Century PP	16	14	14	3	2
Alternative URL at BOPCRIS	56	44	40	11	9
BOPCRIS Home	678	545	493	55	42
House of Commons PP	62	38	34	7	6
Old Bailey Online	991	814	743	43	25
Diary of Samuel Pepys	72	53	49	9	9

Table XI.
Overview of results: 18th Century Official Parliamentary Publications

An interesting result from one of the comparator sites, the Old Bailey Online, showed that this site attracted 20 per cent of its links from the.edu domain, showing that this resource attracted a high number of links from the educational sphere in the US. This is striking, and shows that the Old Bailey Online is both considerably more linked-to overall, and has successfully embedded itself into the American educational sphere. The Old Bailey Online is a free to access resource.

The end-of-project data shows that the 18th Century Parliamentary Publications at BOPCRIS registered an increased number of links during the project, showing that the impact of this project, though small when measured using webometric data, continues to grow. The commercial version of this resource, the House of Commons Parliamentary Papers at ProQuest also registered a small increase, from 38 domains to 55. The BOPCRIS home page registered a large increase in the number of links, rising from 454 domains to 545, reinforcing the added web impact of our project by linking to this resource. Both remaining comparator projects, The Diary of Samuel Pepys and the Old Bailey Online, registered a small decrease in the number of inlinks.

6. Conclusions
The case studies presented here demonstrate that it is possible to conduct a webometric analysis of digital resources and that the comparative link analysis approach used is both practical and useful. In all cases the link analysis gave information about the impact of the main resource investigated as well as additional contextual information about where the resource's main impact occurred. Impact was judged to be represented by the spread of information discovered about these resources, the intellectual context thus represented, and the community in which the resources were found. The usefulness of webometric analysis for these projects depended to a large extent on the location of the digital content and the permanence of those URLs. A project such as Histpop, where the project has a single, stable, unique URL, from which full access to the digital content is freely available, is an ideal candidate for webometric analysis. It is possible to learn a great deal about the project's impact from webometric data. Data collected on the comparator site Old Bailey Online, a project which also has a single, stable, unique URL, from which full access to the digital content is freely available, reinforces this view, showing that educational and international impact can clearly be discerned from the data collected. Digital resources with more complex web presence are more challenging candidates for webometric data collection and analysis. A project such as the 18th Century Parliamentary Publications, with data held in multiple locations with different conditions of access and varieties in content, requires a much more wide-ranging mode of data collection, and more complex analysis to ensure that the full extent of the project's web impact is captured.

There were several practical issues in carrying out the analyses. Multiple URLs for projects mean that data has to be collected on each URL, and then collated. Project URLs that shifted during the period of study also created difficulties as, despite a high level of co-operation from the projects, we were not always kept informed of these changes. Key project content and information was moved elsewhere, as in the case of 19th Century British Library Newspapers and the Wellcome Medical Journals Backfiles, creating challenges to the collection and interpretation of webometric data. Changes to project websites also wiped out digital resources of interest (such as Collect

Britain, the site of the *Penny Illustrated Paper*). As a result of these problems and other issues, such as different ages of the comparator websites, it is important that these quantitative results should be interpreted cautiously.

Careful interpretation of the results, however, can yield important insights, and also suggests new uses for webometric analysis in the planning and marketing of resources, in addition to the impact assessment method presented in this paper. Comparative analysis of the analogue resources in the British Library projects revealed that certain information pages relating to the Newspaper holdings and the Sound Archive were well linked-to, only some of which gave information about the new digital resources studied here. This suggests that webometric impact analysis of pages relating to existing analogue resources would be useful when planning the location and marketing of new digital resources. This would also have the advantage of building strong links between analogue and digital resources, and within communities of digital resources. In addition, webometric analysis has been shown to produce more than quantitative data. Using these data as a starting point for qualitative analysis can also be fruitful, giving insights into the positioning of the resource on educational and other websites, (such as in the clustering of resources on library and information pages), or the extent to which the resource has permeated particular communities (such as coverage in the blogosphere).

While these methods are not definitive statements of the extent or the nature of online impact of the digital resources studied, they are indicative of impact and scope of impact. Webometric methods are therefore successful in judging one aspect of the impact of digitised scholarly resources.

Notes

1. The five projects reported here represent all the projects funded in the first round of JISC digitisation (Phase I: 2004-2006) which agreed to be included in the usage and impact study. One project from Phase I declined to be included.

2. The methods for the wider study formed the based for the creation of a toolkit (TIDSR: http://microsites.oii.ox.ac.uk/tidsr/) designed to use a variety of methods to understand the kinds of impacts digitised scholarly resources can be demonstrated to have.

3. Further details of the comparator sites can be found in the project report, Meyer *et al.* (2009) "Final Report to JISC on the Usage and Impact Study of JISC-funded Phase 1 Digitisation Projects and the Toolkit for the Impact of Digitised Scholarly Resources (TIDSR)", available at: www.jisc.ac.uk/media/documents/programmes/digitisation/tidsr_finalreport.pdf

4. Since our project finished in April 2009, the 19th Century British Library Newspapers project has been updated with a second phase of content totalling a further 1 million pages, and has also moved to a new URL where the public can access a pay per view service: http://newspapers.bl.uk/blcs/. The British Library Archival Sound Recordings project also received a second phase of JISC funding, increasing the number of recordings available from 12,000 to over 44,000 and, where rights permitted, extending free access: http://sounds.bl.uk. The 18th Century Official Parliamentary Publications digitised material is now no longer available through BOPCRIS, and can only be accessed through ProQuest: http://parlipapers.chadwyck.co.uk/home.do.

5. Pub Med Central showed a small drop, but as the results are close to 900, we must discount this data.

References

Atkins, D., Brown, J.S. and Hammond, A.L. (2007), "A review of the open educational resources (OER) movement: achievements, challenges and new opportunities: report to the William and Flora Hewlett Foundation", available at: www.hewlett.org/Programs/Education/OER/OpenContent/Hewlett+OER+Report.htm

Bar-Ilan, J. (2004), "A microscopic link analysis of academic institutions within a country – the case of Israel", *Scientometrics*, Vol. 59 No. 3, pp. 391-403.

Carson, S. (2006), *2005 Program Evaluation Findings Report: MIT OpenCourseWare*, Massachusetts Institute of Technology, Cambridge, MA.

Harley, D. (2007a), "Use and users of digital resources: a survey explored scholars attitudes about educational technology environments in the humanities", *Educause Quarterly*, Vol. 30 No. 4, pp. 12-20.

Harley, D. (2007b), "Why study users? An environmental scan of use and users of digital resources in humanities and social sciences undergraduate education", *First Monday*, Vol. 12 No. 1.

Harley, D., Henke, J., Lawrence, S., Miller, I., Perciali, I., Nasatir, D., Kaskiris, C. and Bautista, C. (2006), *Use and Users and Digital Resources: A Focus on Undergraduate Education in the Humanities and Social Sciences*, Center for Studies in Higher Education, University of California Berkeley, Berkeley, CA.

Harzing, A. and van der Wal, R. (2009), "A Google Scholar h-index for journals: an alternative metric to measure journal impact in economics and business", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 1, pp. 41-6.

Kousha, K. and Thelwall, M. (2007a), "The web impact of open access social science research", *Library and information Science Research*, Vol. 29 No. 4, pp. 495-507.

Kousha, K. and Thelwall, M. (2007b), "Google Scholar citations and Google Web/URL citations: a multi-discipline exploratory analysis", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 7, pp. 1055-65.

Kousha, K. and Thelwall, M. (2007c), "Google Scholar citations and Google Web/URL citations: a multi-discipline exploratory analysis", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 7, pp. 1055-65.

Kousha, K. and Thelwall, M. (2008), "Sources of Google Scholar citations outside the Science Citation Index: a comparison between four science disciplines", *Scientometrics*, Vol. 74 No. 2, pp. 273-94.

Kousha, K. and Thelwall, M. (2009), "Google Book search: citation analysis for social science and the humanities", *Journal of the American Society of Information Science and Technology*, Vol. 60 No. 8, pp. 1537-49.

Kousha, K., Thelwall, M. and Rezaie, S. (2010), "Can the impact of scholarly images be assessed online? An exploratory study using image identification technology", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 9, pp. 1734-44.

Li, X., Thelwall, M., Wilkinson, D. and Musgrove, P.B. (2005), "National and international university departmental web site interlinking, part 2: link patterns", *Scientometrics*, Vol. 64 No. 2, pp. 187-208.

Meho, L. and Yang, K. (2007), "Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs Scopus and Google Scholar", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 13, pp. 2105-25.

Meyer, E.T., Eccles, K., Thelwall, M. and Madsen, C. (2009), "Final report to JISC on the usage and impact study of JISC-funded Phase 1 digitisation projects and the Toolkit for the

Impact of Digitised Scholarly Resources (TIDSR)", available at: www.jisc.ac.uk/media/documents/programmes/digitisation/tidsr_finalreport.pdf

Smith, A.G. and Thelwall, M. (2002), "Web Impact Factors for Australasian universities", *Scientometrics*, Vol. 54 Nos 1/2, pp. 363-80.

Spiro, L. and Segal, J. (2007), *The Impact of Digital Resources on Humanities Research*, Rice University, Houston, TX.

Thelwall, M. (2002), "Conceptualizing documentation on the web: an evaluation of different heuristic-based models for counting links between university web sites", *Journal of American Society for Information Science and Technology*, Vol. 53 No. 12, pp. 995-1005.

Thelwall, M. (2003), "What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation", *Information Research*, Vol. 8 No. 3, available at: http://informationr.net/ir/8-3/paper151.html

Thelwall, M. and Harries, G. (2004), "Do the web sites of higher rated scholars have significantly more online impact?", *Journal of the American Society for Information Science and Technology*, Vol. 55 No. 2, pp. 149-59.

Thomas, O. and Willet, P. (2000), "Webometric analysis of departments of librarianship and information science", *Journal of Information Science*, Vol. 26 No. 6, pp. 421-8.

Vaughan, L. and Hysen, K. (2002), *Relationship between links to journal web sites and impact factors, ASLIB Proceedings*, Vol. 54 No. 6, pp. 356-61.

Vaughan, L. and Shaw, D. (2003), "Bibliographic and web citations: what is the difference?", *Journal of the American Society for Information Science and Technology*, Vol. 54 No. 14, pp. 1313-24.

Vaughan, L. and Shaw, D. (2005), "Web citation data for impact assessment: a comparison of four science disciplines", *Journal of the American Society for Information Science and Technology*, Vol. 56 No. 10, pp. 1075-87.

Warwick, C., Terras, M., Huntington, P. and Pappa, N. (2008), "If you build it will they come? The Lairah Study: quantifying the use of online resources in the arts and humanities through statistical analysis of user log data", *Literary and Linguistic Computing*, Vol. 23 No. 1, pp. 85-102.

Warwick, C., Terras, M., Huntington, P., Pappa, N. and Galina, I. (2006), "The LAIRAH Project: log analysis of digital resources in the arts and humanities. Final Report to the Arts and Humanities Research Council", available at: www.ucl.ac.uk/infostudies/claire-warwick/publications/LAIRAHreport.pdf

Wilkinson, D., Harries, G., Thelwall, M. and Price, E. (2003), "Motivations for academic web site interlinking: evidence for the web as a novel source of information on informal scholarly communication", *Journal of Information Science*, Vol. 29 No. 1, pp. 49-56.

Zuccala, A., Thelwall, M., Oppenheim, C. and Dhiensa, R. (2007), "Web intelligence analyses of digital libraries", *Journal of Documentation*, Vol. 63 No. 4, pp. 558-89.

Corresponding author
Kathryn E. Eccles can be contacted at: kathryn.eccles@oii.ox.ac.uk