Revealing British Euroscepticism in the UK Web Domain and Archive Case Study

Richard Deswarte, University of East Anglia

1.0   Introduction

The emergence of the Web has been one of the most profound and influential phenomena of the last twenty years. Its importance and value as a source of research material, particularly for historians, is only slowly being recognised. One of the dominant features is its constantly changing nature both in terms of content and its technological underpinnings. The content of the Web is an immense resource full of potential for academic researchers both in its current state, but perhaps even more so in its previous forms. Over the last decade or so archives of Web materials, such as the UK Web Archive based at the British Library, have been emerging. These archives are still very much in a nascent form, but do allow researchers access to the previous, and now no longer accessible, content of the live Web. The Big UK Domain Data for the Arts and Humanities Project has been exploring this phenomenon through the example of the UK Web Archive. It is considering the UK Web Archive's potential for historical and social science research, and the tools, functionality and search interfaces needed by scholars and more general users.

This case study focusing on British Euroscepticism in the UK Web Archive was one of a number of funded historical and social science research projects utilising this Web Archive as a means to explore contemporary research topics and to gain practical experience of the peculiar characteristics of Web archives more generally, and the search interface and other prototype tools under development by the British Library for their Web Archive specifically.

The research underlying this case study had a two-pronged basis. First, it was an attempt to gain insight into one of the most controversial and highly debated issues in contemporary British society, namely British Euroscepticism. Second, and equally important, was a desire to explore and unveil from an academic researcher's perspective (as opposed to that of an archivist or computing specialist) the inherent, but still largely unknown, assumptions, strengths and weaknesses of both the data and research interfaces of Web archives in general and the UK Web Archive in detail.

2.0   Revealing British Euroscepticism in the UK Web Domain and Archive

Britain's relationship to and subsequent engagement in the process of European integration is one of the most important political, economic and social developments of the last 50 years. This relationship has always been controversial even before 1973 when the UK joined the EEC, as it then was, and has certainly remained controversial ever since. It will only become more so in the next few years if there is a referendum on Britain's membership as promised by the current Prime Minister. The views and arguments of those individuals and groups who have opposed British membership, commonly referred to over the last 20 years as 'Euroscepticism', have been one of the enduring and vocal elements of British political and media debate. In the previous two decades – exactly the period of the UK Web Domain dataset – much of this debate has been undertaken on the Web with many pro- and anti-European groups setting up webpages and engaging with the public in discussion over the Web via blogs and other postings. Indeed one could argue with reason that this debate has

increased along with the rise of the Web in the last decade, and currently it is certainly vociferous. To date there has been no dedicated research based on these online sites and debates. This case study, which follows on from previous research on the topic as part of the AADDA (Analytical Access to the Domain Dark Archive) project, is an initial attempt to redress that gap. It has and is being undertaken in conjunction with other more traditional forms of research of speeches, debates, printed party propaganda and, all being well, interviews with participants.

3.0   Methodology

Before discussing the specific results of my research into Euroscepticism it is necessary and important to consider research methodology. First and foremost, to my knowledge and in my experience, there are as yet no clearly accepted or commonly used methodological approaches for researching Web archives. As Anat Ben-David and Hugo Huurdeman have indicated in their recent work researchers have either simply searched by URL – the search format used in the Internet Archive's Wayback Machine – or more recently by keyword full-text searching as found in the user interfaces of national Web archives including that of the UK Web Archive.[1] Indeed the UK Web Archive, similar to search engines of the live Web such as Google, provides a keyword search function to find search results which can then be revised using extensive filtering formats. This keyword search design is even more explicit, and indeed is the prime function, in the UK Web Archive's new SHINE prototype search interface.

Thus in common with most, if not all, of the fellow researchers engaged in case studies on this project (and as was the case in the previous AADDA project) I followed uncritically the approach of searching the UK Web Archive using a number of relevant keywords to create a list of website content that I could then analyse both qualitatively and quantitatively. In other words the intention and basic methodological approach were to create a list or database of relevant material which could then be refined, filtered and analysed using a number of the formats and tools available in and beyond the UK Web Archive and SHINE user interfaces, including host domains, links to and from hosts, postcodes and sentiment analysis (available in earlier versions of the search interface). It was hoped that the data would be both reasonably consistently spread over time to allow for quantitative analysis, and be diverse, yet linked enough, in content terms to provide interesting qualitative results.

In addition, in terms of methodology, this case study intended to utilise a number of other digital approaches and tools. On the basis of Google-style Ngram key word searching it was hoped that the results could be graphed and that there would be some revealing aggregate patterns relating to the volume, timing, domain type and variety of websites. The use of the proximity search function to search for related terms was also envisaged as way to see, when the results were graphed, if there were any interesting or distinct patterns or trends. In addition the visual elements of the websites would also not be ignored, thus the characteristics and trends regarding the use of images were also going to be examined qualitatively. It was also hoped to extend the preliminary work done in the previous AADDA project using the sentiment analysis function.

However, one immense challenge emerged in the course of this project that forced a reconsideration of the methodology and the scope of my research goals. This challenge was simply the huge amount of research returns once the larger percentages and the majority of the UK Web Archive and Domain were fully indexed and released into the system. As will be discussed in more

detail further on in this report, this issue, while not entirely unexpected, did cause consternation on my part and that of my fellow researchers on this project (a point that may well be a reoccurring theme throughout all the various case studies) and it resulted, at least on my part, in a search for other approaches.

Given the vast amounts of materials – search returns in the tens if not hundreds of thousands – that were being confronted and the need to engage with the data in a logical and methodological manner, I began an exploration of the existing academic literature regarding sampling, particularly as applied by social scientists more recently and by quantitative historians previously.[2] The intention was to reduce the size of my research results to more manageable numbers. However there was one key distinction in the type of data used in sampling by both social scientists and historians. In all of the examples cited in the literature the data was clearly structured and defined. Indeed structured data was considered a prerequisite. Structured data meant it was possible to make clear and academically justified decisions on the size and relevance of representative samples, most notably with regard to the type and characteristics of the material. Unfortunately and problematically the data in Web archives is almost completely unstructured (at least in terms of its content) which renders the choice of and justification for representative sampling much weaker. Nonetheless experimentations with various methods, including something as simple as looking at every 10th or 100th return to create a sample subset, were attempted. This unfortunately was a limited success particularly with regard to qualitative analysis, as will be explained further on in this report.

Alas in summation no satisfactory methodology or solution with regard to dealing with the massive amounts of materials was adopted within the scope of the project, and this is an issue with which I am still wrestling. This is all the more the case given the critique I have received from academic colleagues regarding the futility of attempting to continue to analyse vast amounts of material from the perspective of qualitative analysis and the desire for 'deep reading'. Perhaps this is indeed a wrong-headed approach. Perhaps the methods and lessons learnt through digital textual analysis of big data and other similar methodologies and tools are a better and academically more rigorous pathway, yet they are far from being fully developed or available off-the-shelf. Unfortunately the limited timespan of this project did not allow for any sustained investigation of, training in and experimentation with these methodologies, although the intention is to do so in the future in order to continue the research.

4.0   Research results

As indicated above, a website content list/subset based on a set of keyword searches of the relevant terms was created. These terms included broad thematic terms such as Eurosceptic and Euroscepticism. They also contained more specific terms relating to both entities, such as political parties, and individuals including EU, Referendum Party, UKIP, James Goldsmith and Nigel Farage.

As the size of the UK Web Domain and Archive has grown over the course of this and the previous AADDA project, with larger and larger percentages of the total data being indexed, the search results have similarly increased. Unfortunately they have increased exponentially and by the autumn of 2014, so by the last half of the project, the keyword results had gone from hundreds (in the AADDA project) to thousands to tens of thousands to hundreds of thousands. For example, the search term Eurosceptic yielded only 312 returns when I first searched for it during the AADDA project (when only 0.5% of the domain was indexed and searchable). This went up to 5,604 by the beginning of this

project and now stands at 220,513 (February 2015). This was even more pronounced with regard to the more specific search term of UKIP which now results in returns of over half a million results. Thus unfortunately what began as a manageable number of websites to analyse individually and 'close read' in their entirety became, during the course of the project, an impossible task. While this was not entirely unexpected, this issue, as mentioned previously, challenged all of the researchers on the project to a greater or lesser extent.

Given this situation in the last half of the project I began exploring different ways and means of limiting the data lists/subsets by refining and reducing my searching to more specific terms, or to linked terms such as UKIP or Nigel Farage (also using the useful new proximity filter). This, however, still resulted in massive returns in the tens of thousands, if not hundreds of thousands. For example, a combined search of UKIP and Farage currently brings up 51,262 results. Even when analysing only specific years and removing links to newspaper domains the results were still in the thousands. That being the case I found myself in a quandary, but thankfully not quite in despair as I was attempting to do representative sampling to reduce the material to manageable amounts. While 'close reading' of random individual pages was yielding interesting and useful qualitative results (see below), these useful specific bits of information were unfortunately not sufficiently linked or coherent to produce solid research trends and conclusions. This was also proving to be an incredibly time-consuming process that, even when engaged in sampling, could not be sustained. Therefore by the end of the project unfortunately I had only limited clear and useful research results with regards to the phenomenon of Euroscepticism.

In terms of qualitative results, as indicated above, I found numerous interesting and valuable individual pieces of information. For example, I was able to unearth the full text of an early speech by Nigel Farage and discover references to a revealing cross-European documentary series entitled 'Eutopia'. In addition I found a number of historical websites/URLs that appear to be offshoots of the main ukip.org URL of which I was unaware and which do appear to be linked to or from the main website. Examples include eukip.org and independenceuk.org.uk. These sites provided much valuable material including the speech mentioned above, but their discovery was serendipitous rather than based on a sound methodological approach to analysing the increasing mountains of materials. However, while I was able to find some evidence for local or regional UKIP websites it was not possible to identify a sufficient number to be able to make any firm conclusions about the connection and interaction of the local versus the central with regard to formal Eurosceptic political movements such as UKIP. Unfortunately this was one of the specific research points for which it was hoped to find material and it was disappointing not to be able to pursue this aspect of my study in any meaningful detail. In addition, numerous intriguing images and designs were found on the various webpages. However it was not possible to develop a sufficiently detailed or large enough database of them because unfortunately one cannot search under images so you have to discover them on your own by examining each webpage – alas there is no Google Images for Web archives. Thus in terms of images, no interesting or revealing research results were obtained other than the rather expected improvements and trends in Web design over the years. In order to pursue this further a massive amount of time would have had to be spent creating a subset or database. This was not feasible as an ongoing or future research track given that it would likely provide only marginal research insights.

Quantitatively it was hoped that the mapping of various search terms using the Ngram search display or its subsequent variation in the SHINE interface would produce some revealing patterns and trends. Therefore all of the keyword search terms were examined in graphs produced through the Ngram or 'Trends' search display. The analysis of these results proved problematic as the upward or downward trends did not match that closely or consistently what was to be expected given key events such as national or European elections where Euroscepticism was an issue. Nor did they map well with regard to other issues such as the creation and inauguration of the Euro, the decision whether Britain should join the Euro based on the so-called five economic tests, or the debates over enlargement or constitutional change. Investigating the peaks and troughs to see if these graphs were pointing towards something unexpected and therefore revealing did not yield any obvious answers or lead to any insights when compared to the historical events of those specific dates. Thus, for example, the increased profile and electoral success of UKIP in 2009 and 2010 is represented by the steep rise in the graph from 2007 onwards. However, the slight decline from 2009 onwards does not make sense given the continued successes and media profile of UKIP from 2009 to 2010. Similarly UKIP's initial breakthrough in the various elections of 2004 and 2005 is not anywhere near as highly represented in the rise shown in the graph. Similar inconsistencies in the graphical representations of other key search words, notably Euroscepticism, are also evident. My suspicion is that the reasons for these unexpected trends may be less to do with real trends related to historical events than to do with the underlying state of the data, resulting from issues such as duplicate data and/or the variability in crawl frequency (see below for further discussion of these issues). Therefore in summary, the mapping of the data and quantitative analysis in general did not produce any revealing or valuable results in terms of furthering the understanding of the phenomenon of Euroscepticism.

Finally with regards to the research question of the growth of Euroscepticism, a number of concluding points can be affirmed. First, there is certainly a strong Web presence for the phenomenon of Euroscepticism, in all its various forms. This phenomenon was without doubt expanding its web presence (as was to be expected) over the years. However, as a second point, it remains to be proven whether quantitative analysis and graphs can illustrate the material effectively and/or point to new or unexpected trends or issues. Third, given the limited size of the domain and archive for the early years of the internet in the 1990s, the resource is not able to answer some of the expected factual queries, such as when did the first UKIP website originate (it must have been at least in 1999 although the earliest one in the domain and in the Internet Archive dates to 2000). Fourth, while it is evident that there are differences between local UKIP websites and the main national one no firm results, patterns or conclusions can be identified. Clear insight remains to be found and will have to be done by further searching of the database or by traditional research approaches, such as through interviews with local UKIP members. Fifth, in qualitative terms it is indisputable that specific and unexpected research 'gems' can be found. However, that is very much by serendipity. Finally, and perhaps most important, the immense size of the search returns, even following attempts at sampling, made it impossible to present any firm conclusions on the phenomenon of Euroscepticism as it was impossible to 'close read' all that material, and qualitative analysis proved equally unrewarding.

5.0   Nature of the data and researching Web archives

This case study, as indicated in the introduction, was not simply focused on Euroscepticism, but was just as much concerned with the experience of scholarly research using Web archives. In that regard

it is appropriate to begin with a discussion of the nature of the 'primary sources', or to be more precise the 'data', found in Web archives.

Web archives, and this is certainly true of the UK Web Archive, are not like the traditional archives that historians use. In many ways the term 'archive' is a misnomer. Web archives do provide access to no-longer-active webpages or previous versions of webpages. This is usually done through a keyword or URL search interface, and access to the webpage is invariably enabled through the Internet Archive's Wayback Machine.[3] Unfortunately the Wayback Machine is unable to present those pages as they first appeared on the screen. Many of the graphics, links or sub-pages are lost and not retrievable. These webpages are thus only partially reconstructed. They are not 'original' artefacts or documents in the sense traditionally recognised by historians. The Wayback Machine does not present an extant artefact or document, it presents a partial simulacrum of the original webpage. Further, the material stored in and accessed from traditional archives is catalogued (or to use the digital term, structured). It has usually been deposited and/or catalogued according to particular themes or subjects, such as government departments or individual papers. This is not the case for Web archives, which consist of almost entirely unstructured material or data (except for curated special collections). The material is 'stored' on the basis of crawl dates, that is, when the data was first scraped from the live Web. The only consistent aspect is the URL. While there is information or metadata on these technical aspects there is no structured information on the content.

Even more important is the fact that the amount of 'data' retrievable from Web archives is unbelievably immense. It consists of tens of thousands of times the size of material contained in traditional national archives. For example, currently the Internet Archive boasts of having saved 455 billion webpages.[4] The UK Web Archive states that its size as of February 2015 was 25.58TB.[5] As described above, the search returns made in the course of this case study often resulted in tens of thousands, if not hundreds of thousands, of links to webpages. For the humanities and even social sciences this is truly 'big data'. This 'big data' is of a size far beyond the usual amounts of source material analysed in traditional historical research projects, even quantitative ones. As shown above, attempts at sampling did not reduce the number of potential sources to the more typical several hundreds or thousands of primary sources that form the basis of most historical scholarship. While not entirely unexpected, the sheer volume of the data did render ineffective the traditional approach of 'close reading', examining in detail each and every source.

Unfortunately there are also further issues with the data beyond its immensity. The data is often 'dirty'. There are many duplicates of websites where the webpages have not changed or been updated. More problematically there are numerous 'false' returns where the keyword search has brought up results that are incorrect. These 'false' returns can take many forms including multiple meanings of a search term or acronym. For example UKIP is not only the acronym of the United Kingdom Independence Party, but was also a computing software company. Another typical type of 'false' return involves the banner or side menu where a reference or even worse an advertisement contains the keyword. Thus I was perturbed when analysing the results of a search of the keyword 'Eurosceptic' to discover hundreds of references to the sports pages of *The Yorkshire Post*. It was only through time-consuming detailed analysis that I realised the webpage had a rotating news-banner across the top which included a reference to the election of a Eurosceptic politician. Thus my search results were contaminated by dozens if not hundreds of 'invalid' webpages. This situation

certainly has the potential to skew seriously any attempts at quantitative analysis. Therefore there were many links to website pages that had nothing at all to do with Euroscepticism. This tainted data was very difficult to rectify other than by manually checking each webpage. I was unable to discover any way to automate the removal of these 'false' pages. These were problems encountered by the other researchers on the project as well, and as far as I know no one was able to find a simple way to rectify these or similar issues.

One final problematic aspect of the data relates to the way it has been acquired. The archived websites found in Web archives are taken by crawling the Web and scrapping the data from it. This is done at regular or irregular intervals, with some websites being crawled more frequently than others. This has resulted in some inconsistencies or peculiarities within the data depending on the frequency of the crawls. The Internet Archive and its Wayback Machine use the URL as a way to access archived websites. It provides a calendar showing the dates when particular archived websites are available. This approach has a number of disadvantages, notably that you cannot do full-text searching for keywords. However, it does have the advantage that users can easily see when and how frequently the specific websites have been crawled and then access some statistical information on that. This openness allows for quicker analysis, particularly when formulating quantitative questions, where it is essential to have a clear understanding of the nature and characteristics of the data as a whole. Unfortunately while this information is usually available in other Web archives, such as the UK Web Archive, it is presented in a much less accessible form. Thus it is much more difficult to get a broader overview and to utilise that broader understanding in formulating appropriate research questions and interrogating the responses. In summation it is crucial that historians and other scholars are aware of the peculiar characteristics and limitations of the data and Web archives. In doing so they will therefore approach these reconstituted 'webpages' differently from either the live web or more traditional primary sources and archives.

A second area worth consideration is the emphasis on searching as the means to access and interrogate web archives. As Ben-David and Huurdeman have shown, keyword full-text searching has become the standard approach offered by Web archives to retrieve archived versions of webpages.[6] Almost all Web archives, with the notable exception of the Internet Archive (which as indicated above uses a URL-focused retrieval method), offer this method of accessing results. This approach is obviously an offshoot of how the live Web is used and the predominance of search engines, notably Google. However, unlike Google with its complex algorithms designed to provide the most relevant links to queries, the listing of websites in Web archives is not based on relevancy, but appears in no particular order (unless directed by filters to prioritise results based on the date of crawl, domain or other facets). From the user perspective this results in certain misconceptions, notably that the user utilises the Web archive search interface with the conditioned expectations of successful Google searching. In other words, either implicitly or explicitly, the researcher assumes similar successful returns to that invariably provided in Google searches. In Google searches users expect to find what they are looking for in the first dozen or so links. If not they make a refined search. Unfortunately this 'Google mentality' is utterly misleading when researchers use keyword searching in Web archives. There are no best results that appear within the first dozen. All of the results will potentially be of interest, not least because there is no order or relevancy. This creates for a researcher the methodological problem of which website links to analyse or not, because there are invariably too many to read all of them. This situation is connected to the desire to engage in qualitative 'close reading', particularly on the part of historians. As has been shown above, the

possibility of doing 'close reading' using Web archives is almost certainly wrongheaded unless smaller sub-sets can be created out of the masses of search results. This view is strengthened by the fact there is no ranking of search results in Web archives.

In conclusion the 'data' found in Web archives differs considerably from that in traditional archives. This data also has certain peculiarities and challenges, not least its immensity, that need to be overcome before it can used for robust academic research. The emphasis on full-text keyword searching, while certainly useful for researchers, is also not without its issues. Yet the tools and interfaces offered by Web archives are very much in an early stage of development and Web archivists are only beginning to grapple with the strengths and weaknesses of both their data and search interfaces. They can only be encouraged to listen to the experiences of researchers and remain flexible in their approaches and interfaces.

6.0   Overall conclusions and recommendations

Starting from the perspective of focused historical research into the phenomenon of Euroscepticism this case study has been far from successful. As indicated above, the research findings have been limited and insufficient to realise the initial research questions. However, that inability to achieve any detailed useful research results in this case study must be seen within the context of the experimental nature of research using Web archives as the main primary source. If anything, this case study shows the need to encourage such speculative and experimental research to discover the unexpected pitfalls and challenges inherent in using Web archives and digital 'big data'.

In addition to that broad conclusion there are a number of other more specific conclusions and recommendations:

(1) Euroscepticism as a research topic for Web archives is possible and certainly has potential. However, in this case study it has proven not to be a viable subject of study, given an overload of research results, the lack of success with sampling, 'dirty' data, and my own lack of expertise in digital methodologies, such as textual analysis, corpus linguistics, data manipulation and clustering algorithms. Unfortunately given the limited time span of this project there proved to be insufficient opportunity to develop such methodological skills or to find research partners with those skills. Perhaps such skills are a necessary requisite for successful research based primarily on Web archives? If so that would limit the use of Web archives to a select few. Given the expectation that this research will continue beyond the project, these aforementioned points will be a personal priority for development. In addition there is a clear recognition that research using Web archives needs to be done in conjunction with more traditional forms of research such as, in this case, interviews, and the examination of printed party propaganda, speeches and debates. Finally, there is certainly an advantage to considering a focused and limited research topic when using Web archives.

(2) Many of the best and most useful results of this project in terms of Euroscepticism came as a result of 'close reading'. However, these were serendipitous findings that would not even be guaranteed by sampling given the unstructured nature of the data. Such 'close reading' can be successful in more limited and structured digital resources, for example in the curated 'Special Collections' of the UK Web Archive. Indeed, the creation of such structured thematic sub-sets by researchers, as was done in the case study of Saskia Huc-Hepher, certainly do have their place and value in Web archives. However, Adam Crymble in his blog response to a presentation on my case study at the IHR Digital History seminar argued that historians should abandon the desire for 'close

reading' and adopt tools and approaches such as corpus linguistics, data manipulation, clustering algorithms and distant reading.[7] He may very well be right, and that is certainly one way forward, but as he wryly observes very few historians currently have those skills, or connections to the people who do.

(3) I, alongside most of the other researchers on the project, struggled with the huge number of returns and how to limit search results and make them meaningful. This case study grappled with different methods of sampling achieving no successful results. Overall, no clear or easy answers emerged from the project in this area, notwithstanding the welcomed and successful development of an online tool by the UK Web Archive team which allowed for the creation of downloadable and saveable subsets that could be edited online to clean and refine the results. Despite this, the creation of viable subsets in this case study proved to be too labour intensive and time-consuming to be worthwhile.

(4) Keyword full-text searching as the standard methodology needs to be critically reconsidered. It patently has value, but its pre-eminence as the main approach to accessing Web archives cannot remain unquestioned. It is surely worth analysing the methodologies that social science scholars using the live web have adopted. Such analysis was beyond the scope of this project, but is certainly something to be encourage (and this scholar will definitely considering doing so).

(5) More broadly this case study has most definitely found that there is a need for continued research and investigation into searching methodologies and rankings in search returns, as has been started by Ben-David and Huurdeman, and by Richard Rogers.[8] Answers to questions such as the following remain to be considered. Do research returns need to be ranked on the basis of algorithms, like Google? Is there an optimal order for the returns? Should the returns always be filtered? Can they be randomised so that you could literally just take the first hundred or thousand hits for a sample? Is it possible for the archived data to be semantically structured to allow for search return rankings?

(6) Following on from the above point it is important to stress to researchers that search tools in Web archives are not the same as in Google and do not return results in a similar fashion. Consequently, researchers need to avoid the 'Google mindset' and recognise the characteristics of Web archive search results, and that they can be problematic. Web archive search returns will be huge, not listed with regard to 'importance', and full of irrelevant material and 'false' returns.

Despite the above issues, Web archives are inordinately rich, rewarding and immense sources of information, but they are also something new and unique. The old methods and mindsets of both historians and archivists will have to be abandoned or at least revised. Grand research goals and hopes need to be tempered by the realities and challenges of the material and available tools. Historians and archivists will have to find new ways and tools to make use of this 'big data', either by developing relevant approaches based on the experiences of corpus linguistics and others, by adopting distant reading or by finding more robust ways to access and limit the rich resources needed for close reading.

[1] Anat Ben-David and Hugo Huurdeman, 'Web archive search as research: methodological and theoretical implications', *Alexandria*, 25 (2014), 93–111, at pp. 98–100.

[2] Roderick Floud, *An Introduction to Quantitative Methods for Historians* (2nd edn., London, 1979); Peter M. Nardi, *Doing Survey Research: a Guide to Quantitative Methods* (Boston, Mass., 2003); Pat Hudson, *History by Numbers: an Introduction to Quantitative Approaches* (London, 2000).

[3] 'Internet Archive Wayback Machine', Internet Archive, https://archive.org/web/ [accessed 27 February 2015].

[4] *Ibid*.

[5] 'UK Web Archive Statistics', UK Web Archive, http://www.webarchive.org.uk/ukwa/statistics [accessed 25 February 2015].

[6] Ben-David and Huurdeman, 'Web archive search as research', pp. 98–9.

[7] Adam Crymble, 'Interrogating the archived UK Web – postscript', IHR Digital History Blog, 6 November 2014 http://ihrdighist.blogs.sas.ac.uk/ [accessed 28 February 2015].

[8] Ben-David and Huurdeman, 'Web Archive search as research', pp. 93–111; Richard Rogers, *Digital Methods* (Cambridge, Mass., 2013) (see, in particular, 'The website as archived object', pp. 61–82).