

Digital barriers and the accessible web: disabled people, information and the internet

Gareth Millward, London School of Hygiene and Tropical Medicine

Summary

- This project initially sought to investigate the information available to disabled people on the early World Wide Web (see section 1.0).
- It did this by searching for key disability organisations to build a corpus of disability websites and pages. A selection of these were then passed through code validation tools to see whether they conformed to accessibility standards as set out by the World Wide Web Consortium (2.2).
- The project encountered a number of issues with the database when trying to answer the initial research questions. The sheer amount of data was overwhelming, and it became clear that traditional 'relevance searching' had created unreasonable expectations about how to ask questions of the available material (2.3, 3.1).
- Further, the method of code analysis employed was far too crude to draw any meaningful conclusions. Somewhat ironically, however, there were some interesting findings with regard to the relationship between the RNIB and web standards over the period covered by the Archive (3.2).
- The focus therefore shifted from a search of the entire database to a focus on specific websites from disability organisations that were deemed to be of importance based on prior knowledge. Given time, a logical avenue of enquiry would be to focus more on the hyperlinks between organisations than on raw string searches of the entire database (3.3, 3.5).
- Despite these failures, useful conclusions were reached; albeit not necessarily in the most 'academically rigorous' way. Rather, a number of historically interesting pieces of information were produced which may pave the way for more focused investigations in the future (3.4, 3.5).
- The overall conclusions from the project were therefore more methodological than concrete. Historians need to be aware of the power and the limits of the database. They also need to shed some of their preconceptions of how to search for information within an unsorted and un-curated archive (3.4, 3.5).
- This may be achieved through better training and more engagement with historical method. However, the way we use web archives will differ according to the sorts of questions we want to ask – and it is at this level that we need to revisit the fundamental assumptions about how we plan and execute historical studies if we are to make best use of the archived web (3.5, 4.0).

1.0 Introduction

My proposal for the 'Big UK Domain Data for the Arts and Humanities' project built on my PhD work and interest in voluntary action history. I wanted to find out more about the activities of disability organisations during the late 1990s and into the twenty-first century. I was not only interested in what these groups did, but how they used the web. What information was made available to disabled people on the web, and how accessible was it? My hope was that by beginning to answer some of these questions I could get a better understanding of how voluntary groups developed in

the digital age; and, crucially, that it would provide me with skills to apply these techniques to future projects.

This report is broken into two core sections beyond this introduction. Section 2 covers the initial plan for the project and the historical background, explaining both what I intended to do and why I felt it was a worthwhile investigation. Section 3 then details the research results, while speculating on how the project could have evolved. It gives an extended discussion on the methodological lessons it provided, and offers some tentative conclusions on how these might be applied to the field of web history.

If you would like to learn more, or would prefer some of these results in a different format, preliminary findings were presented in a seminar at the Institute of Historical Research in November 2014.¹

2.0 Original project plan

2.1 *Historical context*

The rise of the World Wide Web was contemporary with the life of the Disability Discrimination and Equality Acts in the United Kingdom.² The first of these was passed in 1995, while the archive of UK web space available to researchers on this project (hereafter 'the Archive') begins in 1996. As such, this was a particularly interesting period for disability politics. Not only was web technology in a position to deliver information about services to disabled people in a way that had hitherto been impossible, but there was a growing legal framework encouraging (if not always compelling) businesses to provide 'access' to disabled customers and workers.³ These developments were mirrored in the United States with the passing of the Americans with Disabilities Act 1990. Thus, it is fair to hypothesise that 'accessibility' as a concept would also be applied to this relatively new technology of the World Wide Web. Indeed, the negotiations over and subsequent publishing of the World Wide Web Consortium's (W3C) Web Accessibility Initiative 1.0 in 1998 showed that there was an international effort to ensure that disabled people's needs were considered.⁴ The purpose of the guidelines was not simply to help disabled people but to 'make web content available to all users, whatever user agent they are using' (including screen reading peripherals) 'or constraints they may be operating under'.⁵ This was not a trivial matter. There were thousands of software and hardware combinations for accessing the web. The problem of achieving compatibility across browsers just on the Windows platform has plagued developers for decades; but disabled people accessed the internet without using standard screen resolutions, or with adapted mouse and keyboard commands. Screen readers for partially sighted people were also of concern. Having standardised code not only allows users to manipulate the information on a web page to make it accessible (such as by changing the font size, contrast, or using keyboard or voice shortcuts to navigate menus) but also benefits traditional users by allowing cross-compatibility across operating systems and browser software. Although there was no legal obligation to adhere to these rules, W3C was (and remains) the international standards organisation for web programming languages, and as such its recommendations carried weight with professional developers over the course of the period covered by the Archive.⁶

At the same time, this was a period of change for disability voluntary organisations. The big political campaigns of the previous thirty years – first social security, then equalities legislation – had become less relevant after modest victories on both fronts. The 'Welfare Rights Movement' which had emerged in the 1960s and 1970s secured the first disability benefits for the general population during the Wilson and Heath years.⁷ To varying degrees, they had also managed to protect many of these gains from the neoliberal cuts and reforms under Conservative prime ministers Margaret

Thatcher and John Major.⁸ More fundamental concerns about disability rights and equality had emerged from these discussions. Disabled People's Organisations (DPOs), represented by groups such as the Union of the Impaired Against Segregation (UPIAS) and the British Council of Organisations of Disabled People (BCODP), had become more radical advocates of 'civil rights', drawing inspiration from black and feminist movements of previous decades.⁹ By the time that the Committee on Restrictions Against Disabled People was published in 1982 there was a growing movement within Parliament and without for legal protection on the same grounds as the Race Relations and the Sex Discrimination Acts.¹⁰ While these DPOs were not as well funded as the large, traditional impairment-specific charities such as the Spastics Society (now Scope), the Royal National Institute for the Blind (RNIB) or Mencap, they were able to secure the Disability Discrimination Act in 1995.¹¹ It did not go as far as they had hoped, but it at least provided some legal powers to tackle discrimination in British society. Again, it seems reasonable to hypothesise that tactics would change after this major victory. Voluntary organisations had been unified under the *Rights Now!* banner to campaign for legislation, but by the time the Bill became law the coalition had fractured following fundamental differences of opinion on how to proceed. The more-radical DPOs had recommended rejecting the government's Bill in favour of better legal protections; while others from the Welfare Rights Movement (especially the Royal Association for Disability and Rehabilitation) had favoured a compromise solution in which they negotiated to make the best out of the government's concessions. Campaigns continued to strengthen the Act, and under the New Labour government some demands were met.¹² However, the Archive should provide interesting public pronouncements from the major voluntary organisations on how they intended to proceed in this new political landscape.

2.2 Proposed methods

Building on this background knowledge, I wanted to test the two hypotheses to which I have alluded above. First, what information was being made 'accessible' to disabled people? This question requires an analysis of the breadth of information being published, as well as some discussion as to how accessible this was in practice. Second, what role were the major organisations playing in this? What was being published or commented upon by disability organisations, and how did this compare to other institutions, such as the government or private companies?

To achieve this I planned to engage in some quantitative as well as qualitative research. First, the 'how much' questions would be answered through a series of keyword searches using the Archive's interface. This would include terms such as 'disability' or 'handicap' as well as the names of leading organisations such as 'RNIB' and 'BCODP'. The number of times these appeared in the database would be logged both as an overall total and on a year-by-year basis to attempt some analysis of change over time. This would help to identify the URLs of some of the main websites concerned with disability issues or which directly represented particular organisations. It would also be possible to use the search results to build a corpus of disability-related web pages for further analysis.

This would facilitate the second part of the project. Key websites would be passed through a web validation tool to check the HTML code for errors against the standards set in the WAI. The amount of errors would be tracked over time to discover whether there was a pattern between the development of the web and adherence to such standards. It would also make clear if particular organisations – perhaps due to expertise, resources or other factors – were more aware of these issues than others. I would also take a sample of the corpus of search results and attempt some qualitative analysis of the sort of information that was available. Did it include campaign materials, contact details for advice services, employment and leisure opportunities, etc.? Though this sample would always be far from representative, it would at least provide a window onto the way in which

disability was portrayed on the web and how it changed over time, as well as giving a starting point for further research.

2.3 *Foreseen issues*

Before the core research began, a number of issues were apparent. First, since the search engine relies upon strings and cannot make decisions on relevance, it was clear that certain searches would produce results that had little semantic relationship to the questions being asked. A core example would be the Royal Association for Disability and Rehabilitation, known primarily as RADAR. This organisation was active until c.2010, and was one of the most important 'Welfare Rights Movement' groups in the late twentieth and early twenty-first centuries. While a search for the full name (in quotation marks) was likely to produce results of direct relevance to the organisation, 'RADAR' would include web pages on military hardware, multiple references to colloquialisms such as 'on my radar' and so forth. This meant that mentions of the organisation which did not use the full name would be missed in any search, skewing quantitative results. This sort of issue was more acute with charities such as Mind or Scope, which did not (during the period covered by the Archive) have a full name which could be used as a proxy. Thus, even a search term such as "mind AND disability" would not produce a corpus directly related to the specific charity, and could include a range of references to mental health and learning difficulties. I accepted this as inevitable, and hoped that through experimentation with search terms it would be possible at least to produce a workable-yet-flawed corpus of web pages upon which I could conduct further analysis.

Another potential issue was that not all UK-related content was contained within the dataset based on the .uk domain (the only data held in the Archive). For example, the Disability Benefits Consortium hosted a blog with the suffix blogspot.com. Similarly, the United Kingdom Disabled People's Council used a .net suffix, while the Disability Alliance used .org. This would inevitably affect results, as other organisations would have their own web pages included in the search results. It was also clear that the sheer amount of data would be a barrier to a comprehensive analysis of any corpora that were produced. With over two billion entries in the database, it was to be expected that certain searches would produce hundreds of thousands of results. While this would be an issue, I hoped to be able to 'weed out' a number of irrelevant pages through the use of facet searching – and to focus on a sample of the results for code analysis and qualitative reading of the pages themselves.

3.0 Results

3.1 *Initial search terms*

The initial quantitative part of the project was relatively easy to produce. I focused primarily on the names of key disability organisations, cross-referenced against 'disability' and a *.¹³ search of the entire database. A selection of search terms produced clear patterns within the data on a year-by-year basis and in gross terms. It became clear early on, for example, that the RNIB had a relatively high number of references compared to its peers, an advantage enjoyed throughout the period.

The choice of which organisations to search for was not random. It was based on two main criteria: first, I chose groups that had been prominent in my research for my PhD thesis; and second, I specifically used organisations which had a 'unique' name. That is to say, I wanted to be reasonably confident that any page using these strings would be discussing the specific organisation.¹⁴ As Table 1 shows, there were ten key strings:

1. RNIB – The Royal National Institute for the Blind (now the Royal National Institute of Blind People), the leading sight loss charity across the period.¹⁵

2. MENCAP – The Royal Mencap Society, one of the leading charities for mental handicap in the UK. Unlike Mind, Mencap is a unique name, and I assumed that references to this string would be directly related to the charity.¹⁶
3. RNID – The Royal National Institute for the Deaf, and since 2011 Action on Hearing Loss, the leading organisation for hearing loss.¹⁷
4. “disability alliance” – The Disability Alliance was a small, but highly-respected organisation focused on welfare rights.¹⁸
5. “royal association for disability and rehabilitation” – RADAR was the main welfare rights organisation over the period. The full name was considered more reliable than its acronym, though this may have neglected results which only used the abbreviation.¹⁹
6. BCODP – The British Council of Organisations of Disabled People was a historically important disability rights organisation, though with limited resources.²⁰
7. UKDPC – BCODP changed its name in 2007 to the United Kingdom Disabled People’s Council.²¹
8. “spinal injuries association” – The SIA was historically significant, with many of its members in the 1970s and 1980s involved in the campaigns for greater recognition of disabled people’s right. It was both reasonably well-funded, and a ‘unique’ search term.²²
9. “centre for independent living” – A small-but-significant organisation concerned with independent living for disabled people. There was a national group as well as a number of local projects. With the Disability Alliance and RADAR, the national organisation merged to form Disability Rights UK in 2010.²³
10. “disability benefits consortium” – A federal campaigning organisation for better disability benefits, its members included BCODP and RADAR.²⁴

These results were compared against a search for ‘disability’ and the total number of entries per year held in the Archive.

There were spikes in mentions of the RNIB in 1998 and 2006 relative to the entire database, though further analysis would be required to explain why this was so.²⁵ It could have been due to specific events – or it could have been due to the quirks in the way that data was collected and the crawling process that created the database (on which more later). Although all three organisations in Figure 1 had more mentions relative to the whole in 2010 than in 1996, such growth was not even or sustained. More generally, it suggested that the RNIB was better than its peers at getting recognition on the web, with over half a million entries (more than the next two, Mencap and the RNID combined). It also showed that while groups such as RADAR, BCODP and the Disability Alliance had historical significance as lobbying organisations, their presence on the web was much smaller than charities that provided specific services. A more in-depth analysis of those groups’ activities and the ways in which they promoted themselves may yield some answers as to why this was the case. A starting hypothesis, however, is that these lobbying organisations still largely operated by discussing issues with people in positions of power (such as ministers and civil servants),²⁶ while the charities needed both to advertise their services and more openly solicit donations from the public.

The problem with this data, however, is that it was not particularly useful *per se*. While the figures produced were interesting, the inevitable question is ‘so what?’ Do we actually gain anything from knowing that Mencap is mentioned almost ten times as much as the Disability Alliance? Only through disaggregating these results can we gain any real meaning that might be of use to historians. This requires the building of corpora and qualitative analysis of those web pages mentioning these organisations. There is also a nagging question of whether we are genuinely comparing like with like, or if we are missing key information. For example, we cannot reliably

compare the RNIB with Scope or Mind (as mentioned above); nor can we be sure that the searches for the “Royal Association for Disability and Rehabilitation” are capturing everything to do with that organisation. While we can be reasonably sure that the data pertaining to the RNIB is significant – it are mentioned substantially more than the others in this sample – putting a reliable weight on how important it was relative to its peers is difficult, if not impossible.

A way of approaching this would be to perform link analysis. This would focus on direct hyperlinks to the major disability organisations, as these would represent a concrete reference. While ‘Mind’ is difficult to search for, links to ‘mind.org.uk’ are unambiguous.²⁷ It would also give an indication as to the sorts of relationships these groups had developed. Unfortunately, the Archive only contains links in .uk web space, meaning that a number of incoming links would be unavailable. Moreover, not every reference to an organisation is accompanied by a link, so the results would be partial at best. Given more time, I would have begun this analysis to establish how reliable and useful these results would be. To a limited extent, the use of facet searching by link destination helped to narrow down a corpus of results, as is discussed later in this report.

3.2 Code validation

Validating whether code adhered to WAI standards was relatively easy. The *EvalAccess* tool was good at producing reports that showed precisely which parts of the code were problematic, prioritising the errors according to three levels of severity.²⁸ However, this did not produce particularly useful data. While it was clear where errors were on a particular web page, this was not commensurate across websites or across time. For example, the number of errors would inevitably increase as the code for web pages became longer and more information was provided (or, indeed, vice-versa). There was no reliable way to judge ‘errors per line’ as this did not have much bearing on whether the page was accessible or not; nor would it be possible to compare like-with-like. For the RNIB home page, for example, there was no real relationship between the number of errors and time. The number of ‘priority 2’ errors ranged from five to 91, peaking in January 2004, while jumping up and down over the period from 1996 to 2008.

I did not have enough deep knowledge of HTML or web design theory to be able to make judgements on the ‘quality’ of the code being used. Simply adhering to standards is also not a guarantee of compatibility with certain types of screen reader; nor does it ensure that the page is subjectively ‘good’ at displaying useful information in an intelligible way.²⁹ I believe this would be a useful exercise, as accessibility was a constant concern for disability organisations. The RNIB’s lawsuit with *bmbibaby*, referenced earlier, is a good example. The evolution of the quality of code and its compatibility with browsers and other peripherals would be interesting; as would some assessment on the affordability and spread of internet technologies at the time. Given the timeframe for this project and the difficulty in assessing a representative sample of sites, however, this was abandoned (see below).

3.3 Corpus creation

While the size of database was always a potential issue for the ‘readability’ of the Archive, it was not until I actually began to create corpora that I began to appreciate the real extent of this problem. Simply put, there was far too much information to answer the research questions I had originally formulated. Throughout the project, this was the most ‘useful failure’, as it emphasised the importance of developing clear and achievable research questions. It became obvious that the techniques used by traditional documentary historians *rely* on a lack of available evidence. That is, we tend to identify a question and source base, go to the archive, and then mine what we can until that vein is exhausted. This is possible because we have a relatively small amount of evidence which

has survived. With the Archive, however, it is virtually impossible to create a corpus that is both small enough to be human-readable *and* provides a useful, relevant and representative sample across time. But it was important at least to try. In doing so, I had to make certain choices which had a profound impact on the corpora that I created.

The fundamental difficulty with the Archive is that the search results are not necessarily ‘useful’ for linguistic or thematic analysis through traditional ‘big data’ techniques. To explain this, take the example of election pamphlets. It is reasonable to make a decision on what constitutes an ‘election pamphlet’, and moreover it is reasonable to assume where these might be found. They can be taken, their text extracted, and the language analysed to make historical conclusions.³⁰ This works because it is possible to take a representative sample. There will not be duplicates; the collection method will be the same for all; the publication date can be reasonably ascertained; we can make reasonable assumptions about both the intended and actual audience; and each artefact will directly represent the material described in the research question. None of these (relative) certainties exist within the Archive. There are thousands of duplicated pages, or pages which appear unique but actually only contain very slight differences in the underlying code. We have no publication dates. Instead we have ‘capture’ dates, which may either obscure the original date of publication or hide previous drafts of the same page. The ‘web crawl’ as a method of data collection is also far from systematic. It follows links on web pages, and thus may take multiple snapshots of one site and none of another. We cannot, for example, take every page that was live during a particular moment in the same way we can collect all the manifestos for a particular general election.³¹ As a result, we don’t have an even or curated archive, and we are presented with a lot of ‘noise’. If we want to use linguistic analysis to determine how a particular organisation was represented on the internet, is a captured page of search results from an obscure search engine really of any use? A search for, say, ‘Mencap’ will return every mention of Mencap – this does not necessarily mean that people were talking about or analysing the organisation in the same way that an election pamphlet containing the string ‘Ireland’ is probably talking about political issues related to Ireland.

Because of this, we have to try to make blanket decisions on what we include or exclude from our corpora. In turn, any choices that we make in weeding out ‘irrelevant’ information have a circular effect on our research questions. Our interests will affect what we determine is irrelevant; and with each omission, the corpus evolves into a new set of data, representing a slightly different research question and providing slightly different answers. To show this, I followed through the raw search ‘RNIB’ and began applying facet search filters to the results.³² For example, having clicked through a number of pages that, based on the titles, appeared to be completely irrelevant to the RNIB organisation, it was clear that British websites had used ‘RNIB’ as an acknowledgement of accessibility standards. Often, the footers of websites would declare that their code met the standards set out by W3C *and* the RNIB. It became clear, therefore, that any page linking out to W3C *probably* contained no analysis or opinion on the RNIB (or even disability in general). Excluding pages with links out to w3.org removed around 70,000 results. This still left far too large a corpus (around 450,000), so I modified my query slightly. Instead of looking at all mentions of the RNIB, I chose to focus on mentions of the RNIB on external websites – i.e., on websites that were not owned by or affiliated with the RNIB. Thus, I excluded all results from rnib.org.uk and rnib.co.uk. In doing so, I discovered that that RNIB also owned big-print.co.uk, and excluded that too.

That represented a large chunk (around 150,000), but the corpus was still far from readable. I therefore looked again at the domains to see if certain types of website could be cut. mla.org.uk (the Museums and Library Association) and guardian.co.uk (*The Guardian* newspaper) contained 32,000 and 17,000 mentions of the RNIB respectively. I looked at their pages and found that both sites could

probably be removed from the sample. The Museums and Library Association advertised RNIB-approved tours and audio books. While this was useful information in itself, there was so much from this organisation that it bloated the corpus and obscured other types of site. In *The Guardian*, most of the entries related to an advert for an RNIB-approved talking watch. Thus, it was not talking about the organisation directly; and had multiple entries because the advertising scripts had placed the advert on thousands of pages that had little-to-nothing to do with disability. Both sites were excluded.

In an attempt to reduce the sample further, I also excluded any results after 2005 to give a 10-year sample. This still left me with a corpus of 39,270 results. Were I working on this full time, I could have potentially taken one in every 100 entries and performed some sort of qualitative analysis. The conclusions may have been interesting, showing a series of case studies of websites and how they talked about the RNIB. However, it was very clear that these results would be a curiosity. There was very little academic validity to the corpus, and it was difficult to defend the results as representative or in any way objective. The removal of *The Guardian*, for example, was problematic. Yes, it removed the 'duplicate' of the watch advert; but it also inevitably removed a number of national news stories about the RNIB. Historically a left-wing publication, *The Guardian* often gave greater coverage of voluntary organisation campaigning and activities than its competitors.³³ (Indeed, this may have been a core reason for advertising an RNIB-approved watch on that particular website.) In any case, the watch advert was, in itself, an interesting finding. By removing it from the corpus, was I now forbidden from commenting upon it? The same was true of the W3C links – was it not in itself interesting that so many sites used the RNIB as a form of self-validation?

Corpus creation, then, is a very tricky area. The nature of the web crawl means that the data is in itself not entirely reliable, and this is exacerbated when researchers do not ask appropriate questions. Our inability to call on relevance-based searches (*à la* Google) means we are given a mass of un-curated data across a reasonably large time period. But even if we could do a relevance search, do we necessarily want to find exactly what we are looking for? When we use a search engine on the live web, we are usually seeking a specific answer to a specific question. With an archive, we have a general question which (supposedly) allows us to discover the historical story contained within. Finding exactly what we expect to find ought to set off alarm bells for historians. At the same time, with no relevance searching at all, we run the risk of being unable to make any sense of the mass of data we have archived.

3.4 General conclusions

As should be clear from the above, I was unable to answer my initial research questions. But this does not mean there were no useful conclusions. The impact of the RNIB on the early web came out very strongly from the data I collected. Not only did the organisation have significantly more mentions in the database than its peers (see Table 1), its connection with W3C showed its commitment to accessibility standards. Third-party websites in the .uk domain used the RNIB's name as a mark of quality, suggesting not only that the organisation had successfully advertised WAI-like standards, but that it was also a highly respected charity that gave legitimacy to these endeavours. It extended beyond code to other disability aids, such as the talking watch advertised in *The Guardian*. The charity was originally founded as The British and Foreign Society for Improving Embossed Literature for the Blind. We can therefore see a continuity from braille through to web access as a core part of the charity's remit. It would be interesting to follow these findings up with other historical approaches such as analysis of the organisation's administrative documents and oral histories with people involved with the RNIB during the 1990s and 2000s. Was there a specific web

strategy, for example, and how was this effected? This could be combined with a deep qualitative analysis of rnib.org.uk to look at public announcements and campaigns related to web accessibility.

3.5 Reflections on what would be done differently

This project was highly experimental, and I entered it without an appreciation for how vast the archive of UK web space is. This was an advantage, as I could approach it without preconceptions and attempt to use knowledge from my previous research; but it also meant that a number of my initial questions were shown to be practically unanswerable.

Although not presented here, I began to qualitatively analyse the home pages of key disability organisations to see how their sites evolved over time. This included the specific text that was presented, as well as the use of images and the positioning of elements on the screen. If I were to begin this project from scratch, I would probably use this as my starting point. As explained in the opening section, my interests centred on how disability organisations campaigned and how they adapted during the years after the Disability Discrimination Act. Taking a sample of traditional charities, welfare rights organisations and disabled people's organisations would give an insight into whether they adopted different approaches to the web. Were they constrained by budgets? Did certain organisations see the web as an opportunity for campaigning, or did it remain quite niche? Did they link to each other, and if so were they more disposed to create web relationships with certain types of institution? Knowing now what information can be gleaned from the archive, I believe these sorts of questions are both answerable and would relate more directly to my historical interests. This would also have the advantage of creating a more representative corpus of web pages, opening up possibilities for corpus linguistics and other 'big data' analytical techniques.

I would, however, be keen to exploit the power of the data we have available. The project I outlined in the above paragraph does not really require any input from the British Library's data or search engines. Since I am taking a known corpus of websites, I could simply type the URLs into the Wayback Machine and circumvent the raw data entirely. While this would be legitimate and produce useful conclusions, it also seems to be missing an opportunity. I would be interested, therefore, in going a stage further and performing link analysis using the .uk data. Both the links *to* the websites in my corpus and *from* are very important for understanding how the web – by definition a network, not a series of isolated data points – was negotiated by disability organisations. There are limits to what this analysis could achieve in terms of *in* links, given that we only have access to the .uk domain; but for the most part the hyperlinks *from* my corpus sites would be complete. A way around this might be to make use of the demarcation in the Archive between general websites (.org.uk; .co.uk) and those from government or academic institutions (.gov.uk and .ac.uk). This would give some indication about the relationship between government (local and national) and disability organisations. That could include lobbying, consultation, provision of services, exchanges of information, and so on. Certainly this was a key feature of the voluntary sector before 1996, and something I would expect to see in the data.³⁴ Analysis of these relationships may reveal some of the specifics of how and to what extent the internet was used for such purposes.

I am still interested in 'presence' or 'reach' on the web, and the relative size of the different organisations. However, I feel that raw searches for the names of disability organisations were flawed for this purpose. Direct links *in* to key sites are probably more useful, although not without their problems. The data I did create in this regard was helpful for developing new questions (such as 'why does the RNIB get so many more mentions than its peers?'), but was largely irrelevant for drawing conclusions in its own right. By developing an idea of which sites were linking to disability organisations, a corpus could be created in which a sample could be investigated. This would, I

believe, still have discovered *The Guardian* watch advert and the relationship between the RNIB and W3C without having to spend so long worrying about how to whittle down my results to a human-readable amount. However, such concepts of ‘reach’ require a proper conceptualisation before they can be measured.

4.0 Conclusions

Historians need to think more about how we are going to make use of web archives. The history of the early twenty-first century is going to be very difficult to write without them. Studies will have to be more carefully planned and executed than traditional documentary investigations can sometimes be. The art of diving into an archive with a rough idea of what we want to look at, and using those initial findings to construct a documentary analysis of the past is nigh-on impossible given the vast amounts of data the web has produced. The data is unsorted, contains plenty of ‘useless’ information, and is unnavigable without at least some plan as to how to get from our initial interests to a usable corpus of intelligible data.

This presents the discipline with some fundamental challenges. We need to have a clear enough idea of what we want to look for without predetermining what we will find. For example, building a corpus of websites that have already been declared ‘important’ may exclude the possibility of discovering information that challenges our preconceptions. Similarly, if we limit our searches solely to the .gov.uk suffix (because it’s a relatively self-contained group) we limit the questions we can ask about British society and the lens through which we will be able to access it. But even if we did want to plan an investigation thoroughly, we do not have enough case studies or methodological literature to help us design this research.

In many ways these dilemmas of historical method are not new – the difference is the scale. Mixing old and new methods may, therefore, may provide us with some answers. It is important that more people use this source and continue to make mistakes. In doing so, scholars working with the Archive can pass their experiences on to others, and we can begin to develop the methodological tools necessary for making sense of all this information.

¹ Institute of Historical Research, ‘Interrogating the archived UK web: historians’ and social scientists’ research experiences’ (4 Nov. 2014) <http://www.history.ac.uk/podcasts/digital-history/interrogating-archived-uk-web-historians-and-social-scientists-research> [accessed 19 January 2015].

² Including, but not limited to: Disability Discrimination Act 1995; Disability Rights Commission Act 1999; Disability Discrimination Act 2005; Equalities Act 2006; Equalities Act 2010.

³ *De jure*, the Equalities Act 2010 made it compulsory for companies to make their websites accessible for disabled people. As yet nobody has been convicted, but the RNIB has initiated some civil cases (see A. Savvas, ‘Bmibaby sued over “inaccessible” website’, *PC Advisor* (27 January 2012) <http://www.pcadvisor.co.uk/news/tech-industry/3333281/bmibaby-sued-over-inaccessible-website/> [accessed 8 January 2015]).

⁴ As a working draft (W3C, *Web Accessibility Initiative HTML Author Guidelines – version 1.0* (23 January 1998) <http://www.w3.org/WAI/GL/WD-WAI-HAG-19980123.html> [accessed 8 January 2015]).

⁵ 'Abstract' of W3C, *Web Content Accessibility Guidelines 1.0* (5 May 1999) <http://www.w3.org/TR/WCAG10/> [accessed 8 January 2015].

⁶ See, e.g., W3C, *HTML 4.01 Specification – W3C Recommendation* (24 December 1999) <http://www.w3.org/TR/html401/> [accessed 8 January 2015].

⁷ G. Millward, 'Social security policy and the early disability movement – expertise, disability and the government, 1965-1977', *Twentieth Century British History* (2014).

⁸ G. Millward, 'Invalid definitions, invalid responses: disability and the welfare state, 1965–95' (unpublished University of London PhD thesis, 2013).

⁹ J. Campbell and M. Oliver, *Disability Politics: Understanding our Past, Changing our Future* (London, 1996); M. Oliver and C. Barnes, *Disabled People and Social Policy: from Inclusion to Exclusion* (London, 1998).

¹⁰ Committee on Restrictions Against Disabled People, *Report* (London, 1983); Race Relations Act 1965; Sex Discrimination Act 1975. Also important in these two areas was the creation of the Commission for Racial Equality and the Equal Opportunities Commission. Both bodies promoted and enforced their respective acts.

¹¹ C. Barnes and M. Oliver, 'Disability rights: rhetoric and reality in the UK', *Disability & Society*, 10 (1995), 111–16; Rights Now, *Accounting for Discrimination: Estimating the Costs of Excluding Disabled People from the Workforce* (London, 1994); Leeds Disability Studies Archive, M. Oliver and C. Barnes, 'Disability politics and the disability movement in Britain: where did it all go wrong?' (2006) <http://disability-studies.leeds.ac.uk/files/library/Barnes-Coalition-disability-politics-paper.pdf> [accessed 19 January 2015].

¹² E.g., the creation of the Disability Rights Commission which the Conservative government refused to allow (see Disability Rights Commission Act 1999).

¹³ Shorthand for 'all'. This may be familiar to users of command line arguments in MS-DOS and Windows.

¹⁴ See the distinction between 'RADAR' and 'Royal Association for Disability and Rehabilitation' discussed above.

¹⁵ Each organisation's current home page is available either through the live web or through the Internet Archive's Wayback Machine (see RNIB <http://www.rnib.org.uk> [accessed 19 January 2015]).

¹⁶ Mencap <http://www.mencap.org.uk> [accessed 19 January 2015].

¹⁷ Now Action on Hearing Loss <http://www.actiononhearingloss.org.uk> [accessed 19 January 2015]; previously Royal National Institute for Deaf People <http://web.archive.org/web/19981212032354/http://www.rnid.org.uk/> [Internet Archive, site captured 14 November 1999].

¹⁸ Now Disability Rights UK <http://www.disabilityrightsuk.org> [accessed 19 January 2015]; previously Disability Alliance <http://web.archive.org/web/20070116091524/http://www.disabilityalliance.org/> [Internet Archive, site captured 16 January 2007].

-
- ¹⁹ Now Disability Rights UK; previously RADAR <http://web.archive.org/web/20051227002512/http://www.radar.org.uk/RANE/Templates/frontpage.asp?HeaderID=227> [Internet Archive, site captured 27 December 2005].
- ²⁰ Now United Kingdom Disabled People's Council <http://www.ukdpc.net> [accessed 19 January 2015]; previously British Council of Disabled People and British Council of Organisations of Disabled People <http://web.archive.org/web/20060131094342/http://www.bcodp.org.uk/> [Internet Archive, site captured 31 January 2006].
- ²¹ *Ibid.*
- ²² Spinal Injuries Association <http://www.spinal.co.uk> [accessed 19 January 2015].
- ²³ Now Disability Rights UK; previously National Centre for Independent Living <http://web.archive.org/web/20051025162738/http://www.ncil.org.uk/> [Internet Archive, site captured 25 October 2005].
- ²⁴ Disability Benefits Consortium <https://disabilitybenefitsconsortium.wordpress.com/> [accessed 19 January 2015].
- ²⁵ Or even whether comparing against *.* was useful or necessary.
- ²⁶ See P. Whitely and S. Winyard, *Pressure for the Poor: the Poverty Lobby and Policy Making* (London, 1987).
- ²⁷ As unambiguous as they can be within this specific context.
- ²⁸ Universidad del Pais Vasco, *EvalAccess 2.0* <http://sipt07.si.ehu.es/evalaccess2/index.html> [accessed 17 July 2014].
- ²⁹ W3C, 'Selecting web accessibility evaluation tools' <http://www.w3.org/WAI/eval/selectingtools.html> [accessed 12 January 2015].
- ³⁰ See L. Blaxill, 'Quantifying the language of British politics, 1880–1914', seminar paper delivered at the Institute of Historical Research, 23 Oct. 2012 <http://www.history.ac.uk/podcasts/digital-history/quantifying-language-british-politics-1880-1914> [accessed 13 Jan. 2015]; L. Blaxill, 'Quantifying the language of British politics, 1880–1914', *Historical Research*, 86 (2013), 313–41.
- ³¹ With the obvious caveats of having to rely on what survives, which organisations and individuals maintain their archives and the usual quirks of historical research.
- ³² At the time of this project, the *Shine* interface allowed the user to include/exclude results by: content type; public suffix; domain; crawl year; postcode; author; link domain; language; link to public suffixes.
- ³³ See, e.g., analyses in M. Hilton and others, *A Historical Guide to NGOs in Britain: Charities, Civil Society and the Voluntary Sector since 1945* (Basingstoke, 2012).
- ³⁴ M. Hilton and others, *The Politics of Expertise: How NGOs Shaped Modern Britain* (Oxford, 2013).