Digital Archaeology in the Web of Links: Reconstructing a Late-1990s Web Sphere

Peter Webster

Abstract One unit of analysis within the archived Web is the "web sphere", a body of material from different hosts that is related in some meaningful sense. This chapter outlines a method of reconstructing such a web sphere from the late 1990s, that of conservative British Christians as they interacted with each other and with others in the USA in relation to issues of morality, domestic and international politics, law and the prophetic interpretation of world events. Using an iterative method of interrogation of the graph of links for the archived UK Web, it shows the potential for the reconstruction of what I describe as a "soft" web sphere from what is in effect an archive with a finding aid with only classmarks and no descriptions.

1 Introduction

The attention of contemporary historians is now turning to the 1990s, as we reach a sufficient distance from the period to view it in some sort of perspective. The period from the fall of the Berlin Wall until the 9/11 terrorist attacks—what might be called a "long Nineties"—is beginning to come into focus as one possible analytical unit in world cultural and political history. As a result, historians are increasingly turning to the Internet and the Web as another type of primary source for the study of the period.

This attention is being paid to differing aspects of web history. In their differing ways, both the 2018 special issue of the journal *Internet Histories* on "the Web of the 90s" and the *SAGE Handbook of Web History* (Brügger and Milligan 2019) show the interest in the technical history of the medium itself. My own work on Christian sentiment regarding the early Web and its potential effects is a contribution to what might be called a cultural or affective history of technology (Webster 2018). As academic interest in web history has grown, so has the understanding of the

P. Webster (✉)
Webster Research and Consulting, Chichester, West Sussex, UK
e-mail: peter@websterresearchconsulting.com

D. Gomes et al. (eds.), *The Past Web*,
https://doi.org/10.1007/978-3-030-63291-5_12

archived Web as an object of study. As the work of Niels Brügger has shown, though an archived page replayed in a browser appears seductively similar to its original, scholars must grapple with its fundamental otherness and deal with it on those very distinct terms; there is a need for a philology of the archived Web (Brügger 2018, pp. 23–30). But there is more to the Web than individual objects, and indeed individual sites (though much early work in web history was focussed at this level). What follows is an exercise in understanding a particular "web sphere".

2 "Soft" and "Hard" Web Spheres

Brügger defines a web sphere as "web material ... related to a topic, a theme, an event or a geographic area" (Brügger 2018, pp. 34–5). Here I posit a distinction (at least in degree, if not in kind) between "hard" and "soft" web spheres, defined in terms of the ease with which their boundaries may be drawn, and the rate at which those boundaries move over time. Examples of hard web spheres might be those of organisations that have clear forms of membership or association: the websites of the individual members of the European Parliament, for instance, or of the football clubs in the English Premier League in a given season. Given the clarity with which these particular definitions can be drawn, and the stability of those definitions (5 years for Members of the European Parliament, 1 year for the Premier League), the initial work in isolating a web sphere and its archival traces for study is relatively light.

The study of "soft" web spheres tends to present additional difficulties, since the definition of topics or themes (to continue with Brügger's definition) is more difficult if not expressed in institutional terms. Web spheres defined by a concept or theme without an institutional analogue are by definition more fluid and more subjective, since the definition of "European politics" or "English football" may be contested in ways that "membership of the European Parliament" or "Premier League clubs" may not. This chapter is a case study in the reconstruction of just such a soft web sphere, much of which is lost from the live Web and exists only in the Internet Archive.

3 Conservative Christianity, Politics, Law and Morality: The Problem Defined

Scholars of conservative Christianity have known for many years of various correlations between conservatism in doctrine and particular stances on a number of key political, legal and ethical issues, often questions of personal morality that are in some way regulated by the law. The modern history of British evangelicalism shows periodic rises and falls in the levels of general anxiety about the moral state of the UK, although the particular complex of issues at stake was different

at each point (see, for instance, Grimley 2014). In the case of the "long Nineties" in particular, certain issues were prominent, some of which were a continuation of older anxieties. Anglican divisions over sexuality were already visible before exploding in controversy at the 1998 Lambeth Conference, the decennial gathering of the bishops of the worldwide Anglican Communion (Bates 2004); conservative disquiet in the UK was further fuelled by the moves towards recognition of same-sex partnerships in what became the Civil Partnerships Act 2004. Though there was no material change in UK law, public debate continued about the law on abortion and on assisted suicide, in both of which cases there was marked Christian engagement, as there was with the attempt to change the law in relation to therapeutic cloning in 2001. Scholars of right-wing politics in general have also often noted a further correlation between some varieties of conservative Christianity and a generalised sympathy for the politics of the right (Jackson 2010). In the UK, scholars have examined in particular Christian attitudes to the European Union and to concerns about the place of Islam in public life (Smith and Woodhead 2018; Atherstone 2019; Webster 2017).

Less obvious both in the media and in popular consciousness was the engagement of conservative Protestants in particular with issues of biblical interpretation that have cultural consequences: creationism and its place (if any) in science education in schools and in public broadcasting (Numbers 1992, pp. 323–30), and the interpretation of biblical prophecy in relation to current politics, including the Middle East, climate change and the end of the world (Sweetnam 2019). The period from the mid-1990s—the beginning of the web age—was, however, given a particular sharpness—what quizzical journalists dubbed "pre-millennial tension"[1]—by the approach of the year 2000. The projected catastrophe of the Y2K bug was itself a subject of prophetic speculation (McMinn 2001). But there was no shortage of material in world politics to fuel the reflection of Christian Zionists who were already "semiotically aroused" (Clark 2007, pp. 149–255; the phrase is of Landes 2007, p. 4). In Israel, the Oslo Accords of 1993 and 1995 were followed by the assassination of Prime Minister Rabin, the failed talks at Camp David in 2000 and eventually the outbreak of the Second Intifada in 2001. The notion of a "clash of civilisations"—between "the West" and "Islam"—already current in the late 1990s, was brought into new relief by the terrorist attacks of 11 September 2001.

Before the widespread adoption of the Web, conservative Christian anxiety in the UK about issues such as these was often expressed by and through various parachurch organisations—Christian in ethos and leadership but functionally independent of the churches themselves. Some of these had long histories, such as the Society for the Protection of the Unborn Child, set up in reaction to what became the Abortion Act 1967, the decisive legalisation of abortion in most of the UK; some were rather newer. Before the Internet, these organisations tended to communicate

[1] 'Have you got pre-millennial tension?', *The Independent*, 29 December 1996, retrieved 16 May 2019 from https://www.independent.co.uk/life-style/pmt-have-you-got-pre-millennial-tension-1316310.html

by means of events and public speaking, video and audio cassette tape distribution, and in a panoply of ephemeral literature; all of these were poorly archived and are now hard to recover. The Web, as it lowered the barriers to entry to public discussion, not only gave existing and newly formed groups alike a novel means of communication but also allowed individuals to add their voices to the chorus.

The archived Web affords an opportunity to address two particular questions that were previously difficult if not impossible to approach. The first of these is the degree to which, though each of these disputes has been observed singly, the same individuals or groups engaged with several or all of them as a package. Was engagement with this particular cluster of issues a function of something more fundamental in theology and practice? Or, alternatively, did conservative British Christians engage with each issue on its own merits, as a singular interest, rather than as part of a wider set of common concerns? The question has not been addressed directly by scholars so far, although the literature when read in the round (and anecdotal experience of the constituency) suggests the former. An examination of the link relationships in the archived Web affords one way of approaching the question. To what degree were the parts of the Web that address each individual issue connected with each other?

Secondly, the same data can also be used to investigate the extent to which the Web allowed an internationalisation of religious interchange. Christians in the UK had for many decades been in dialogue with believers in the USA, but were limited by the cost of travel and of importing print and other material; the well-studied television and radio output of American Christians was largely unavailable in the UK. The Web allowed readers and publishers alike in the UK to access material from other countries that previously would have been difficult if not impossible to obtain. How (if at all) is such an internationalisation to be observed in the link relationships of an Anglo-American conservative Christian web sphere?

How, then, to recover the sound of this chorus of voices, in order to address these two questions? Internet Studies scholars have made creative use of search engine results as a means of identifying soft web spheres on the live Web and of creating archival collections of that content (see, for instance, Ackland and Evans 2017). Such keyword searching, however, presupposes (first) that there are useful keywords to employ, that are distinctive (in the way that "premillennialism" is, but "apocalypse" is not), and stable in meaning over time. It also supposes that one's initial knowledge of the field is sufficient to create such a list in the first place, without omitting terms that are in fact of an importance about which one did not know. Keyword searching also recovers results of the use of particular terms in content from organisations and individuals which might in fact oppose the agenda of those organisations under examination; a useful line of enquiry in its own right, but quite distinct.

The would-be historian of the late 1990s must deal with a more acute problem still—a problem with several layers. First, the vast bulk of content from the period is no longer on the live Web; second, as scholars, we have few, if any, indications of what has been lost—no inventory of the 1990s Web against which to check; third, of the content that was captured by the Internet Archive (more or less the only archive

of the Anglophone Web of the period), only a superficial layer is exposed to keyword search, and the bulk may only be retrieved by a search for the URL. Not only do we not know what was never archived, among the content that was archived it is difficult to find what we might want, since there is no means of knowing the URL of a lost resource. To use an analogy from an archival context, there is a finding aid, but one with only classmarks, and no descriptions, or (alternatively) a library catalogue containing only ISBN numbers, when the ISBN registry is lost.

4 The First Stage of the Method: Identifying an Initial List of Domains

Scholars require, then, some means of understanding the archived Web using only the technical data about it that it itself can be made to disclose. In the remainder of this chapter, I outline a method of web sphere reconstruction based not on page content but on the relationships between sites, i.e. the Web of hyperlinks that itself forms part of the archived Web. The method is iterative in nature and involves both the computational interrogation of large datasets and the close examination of individual archived pages, along with the use of printed and other non-digital sources. It builds upon several recent studies, all of which in their different ways explore the available primary sources from outside the Web from which it may be reconstructed, both on small scales (Nanni 2017; Teszelszky 2019) and at the scale of whole national domains (Ben-David 2016; Ben-David 2019). It is an elaboration of two studies of my own in which the particular method was essayed in relation to smaller and less complex web spheres (Webster 2017; Webster 2019).

The first stage was to establish an initial list of the domains of the conservative Christian organisations concerned. This was derived from several sources. The first was the live Web, since many of the campaigning organisations that were active in the late 1990s still exist. As such it was possible to compile the current domains of all those organisations which still existed in 2019. (The assumptions involved in this are examined below.)

The second source was the body of published scholarly work on the subject published during the period under examination. Contemporary historians are well accustomed to relying on the work of scholars in other disciplines carried out in real time. Economists, sociologists and others studying the social phenomena of their own time have left a rich legacy of scholarship that both documents and explains those phenomena. Quite often those discursive insights are accompanied and supported by data, created at the time, which when shared have now become primary sources in their own right. In the case of the religious Web, there is a significant body of pioneering early work by Internet Studies scholars (on the apocalyptic Web see, for instance, Campbell 2004). However, it is now cut adrift from the matters it describes, as the materials it analysed no longer exist. The analogy with economics or sociology thus breaks down; since this research created

no research data as such, the work described stands in the same relation to the present as does the embedded observational work of anthropologists, from which data is not usually derived. *Give me that online religion* by the American scholar Brenda E. Brasher was one of the earliest examinations of religious life online (Brasher 2001). It carried the disclaimer that "we know that many of the Websites listed in this book no longer exist. That is the nature of the Web". It referred readers to Brasher's own website for a list of current sites. Nearly 20 years on, that site also no longer exists, and itself was only partially archived.[2] It was possible, then, to derive additional relevant domains from a review of early published scholarship on online religion. Where the scholar gave the domain, the task was straightforward; where they had not, it was possible in some cases to identify the domains using the keyword homepage search in the Internet Archive.

Readers over a certain age will remember a time when publishers rushed to issue printed books that guided surfers to the "best" of the Web. One such book was *God on the Internet* (Kellner 1996), which gave short accounts of several hundred religious sites and their URLs. Some domains were derived from this and other similar publications.

Even at the time they were published, it was clear that such printed accounts would very soon fall out of date, and readers could also turn to sites such as Yahoo, which was listing thousands of religious websites as early as 1996, and the Open Directory Project (dmoz). (On the history of finding things on the Web before the dominance of search engines, see Halavais 2019.) The Internet Archive holds archived versions of these lists from 1996 for Yahoo and 1999 for Dmoz, and these were used to derive further domains of relevance. (An example of the use of historic web directories is Musso and Merletti 2016.) There were also examples of Christians creating directories of websites on a smaller scale, such as the Organization of Christians with Internet Home Pages (ocih.org, created in 1995 in the USA). Archived versions of the OCIH's listing survive from 1999 until 2008, after which the site disappears from view. (The domain passed to Orange County Implant Health, a plastic surgery practice in California.)

The fourth source was data derived from another web archive, the UK Web Archive. I myself was a curator of a special collection in the UK Web Archive on religion, law and the state in contemporary Britain, first as a guest researcher from 2011 and then continuing after I joined the staff of the UK Web Archive in 2012. Using data provided by former colleagues at the British Library, I was able to derive further domains, of three types: those which had been archived and still existed on the live Web; those that had been archived but had subsequently gone from the live

[2]The Internet Archive has a single instance of brendabrasher.com but it does not contain the content referred to: https://web.archive.org/web/20010514053625/http://www.brendabrasher.com:80/online_religion.html. Brasher's university pages contained some such listings, but these are only present in the Wayback Machine until 2005, at the URL http://raider.muc.edu/~brashebe/online religion.htm.

Web; and finally those which I had selected but could not at the time be archived because permission from the site owner could be not secured.[3]

From these four sources, then, it was possible to derive an initial list of domains (or, in fact, several lists, concerned with different issues, also categorised by their country of origin). One limitation of this approach is that it tends to foreground those organisations which had troubled to register a domain, or at least a host on a domain, such that content hosted at a location like *prophecy.com* or *prophecy. mywebsite.com* appears in the graph where *mywebsite.com/prophecy* does not. And the evidence of some early listings, both printed and online, suggests that a great many sites in the earliest days did not have their own domains. And, tempting though it may be, we should not assume that the most well-organised or well-funded organisations were the earliest to register domains, since a great many of the most apparently eccentric sites run by individuals had their own domains. In any case, it is clear that this method underestimates the size and density of the conservative Christian Web.

5 The Second Stage: Interrogating the Graph of Links

The second stage was to interrogate web archives themselves, in order both to determine the existence or otherwise of archived versions of these domains and to uncover further relevant domains. The British Library makes available a dataset that serves this purpose, being the UK Host Link Graph. It is derived from the JISC UK Web Domain Dataset, an extraction of the holdings for the *.uk* ccTLD for the period 1996–2013, which forms part of the UK Web Archive's collections. The Host Link Graph records the instance of link relations between .uk hosts and other hosts (both .uk and not). The Link Graph for the period 1996–2010 contains over 2 billion such relationships, expressed in the form:

1997 | *host1.co.uk* | *host2.com* | *7.*

which declares that in 1997, the Internet Archive found some seven resources (usually pages, but not always) on *host1.co.uk* which contained at least one link to a resource on *host2.com*.

Taking the individual categorised lists of domains, a set of shell scripts were created (one for each list) to extract from the UK Host Link Graph all the edges in the graph that represented either an inbound link to, or an outbound link from, one of the domains on that list. The resulting datasets were then manipulated in turn to extract those domain names that had not previously been identified. These newly discovered domains were then individually assessed and categorised, by means of both the live Web (where applicable) and with archived instances in the Internet Archive. In some cases, a domain was one previously used by an organisation

[3]My thanks are due to Nicola Bingham for her assistance in obtaining this data. Some details of the project are available at https://peterwebster.me/2013/01/22/religion-politics-and-law-in-contemporary-britain-a-web-archive/

which later adopted the domain in the initial list; in these cases, the association was recorded and the superseded domain added to the master list. Visual inspection of other previously unknown archived sites revealed numerous webrings, blogrolls and pages of links, from which further domains of relevance were derived and fed back into the master lists.

Using, then, a set of revised scripts, the process was then repeated until no new relevant domains were being revealed by new iterations. Once this point had been reached for each list, the final datasets derived were concatenated to produce a single graph for all the domains under examination, and then duplicate lines were removed. This dataset contained a significant amount of noise—inbound and outbound links with domains that were irrelevant to the study. During each iteration of the process, however, domains that on inspection were classified as irrelevant—for instance, some early general directories of the Web that linked in, or advertising and similar domains to which there were outbound links—had been noted. From these, an additional script was now created, to remove lines that contained those domains from the graph.

By this means it was possible to derive a link graph of the conservative Christian Web in the UK in the earliest days of web archiving. The resulting graph contained two principal types of edge. The first type was outbound links from organisations within the UK ccTLD to related organisations both within that ccTLD and beyond it, and also to churches and other religious organisations, government, civil society organisations and the media. The second type was inbound links to the organisations under examination from other domains within the ccTLD. With this graph, it is possible to begin to analyse the interrelations between campaigning groups on several different issues related to religion, law, politics and ethics in the UK, and their relations with the non-UK Web, and the USA in particular.

6 Summary

This is not the venue for a full rehearsal of that analysis, which I intend to publish at length elsewhere. However, the method outlined here is one by which the boundary of a "soft" web sphere can be drawn by interrogation of a larger link graph. It also demonstrates the kind of multi-source investigation necessary to uncover the archaeology of the early Web. Big data and small, printed sources, the traces of previous web archiving efforts (even when unsuccessful) and echoes in the scholarly record itself: all these come into play when trying to understand an archive without a catalogue.

References

Ackland R, Evans A (2017) Using the web to examine the evolution of the abortion debate in Australia, 2005-2015. In: Brügger N, Schroeder R (eds) The web as history. UCL Press, London, pp 159–189

Atherstone A (2019) Evangelicals and Islam. In: Atherstone A, Jones DC (eds) The Routledge research companion to the history of evangelicalism. Routledge, Abingdon, pp 127–145

Bates S (2004) A church at war: anglicans and homosexuality. I. B. Tauris, London

Ben-David A (2016) What does the web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain. New Media Soc 18:1103–1119. https://doi.org/10.1177/1461444816643790

Ben-David A (2019) National web histories at the fringe of the web: Palestine, Kosovo and the quest for online self-determination. In: Brügger N, Laursen D (eds) The historical web and digital humanities: the case of national web domains. Routledge, London, pp 89–109

Brasher BE (2001) Give me that online religion. Jossey-Bass, San Francisco

Brügger N (2018) The archived web: doing history in the digital age. MIT Press, Cambridge, MA

Brügger N, Milligan I (eds) (2019) The SAGE handbook of web history. SAGE, London

Campbell RA (2004) Searching for the apocalypse in cyberspace. In: Dawson LL, Cowan DE (eds) Religion online: finding faith on the internet. Routledge, London, pp 239–253

Clark V (2007) Allies for Armageddon: the rise of Christian Zionism. Yale, New Haven

Grimley M (2014) Anglican evangelicals and anti-permissiveness: the nationwide festival of light, 1971-1983. In: Atherstone A, Maiden J (eds) Evangelicalism and the Church of England in the twentieth century: reform, resistance and renewal. Boydell and Brewer, Woodbridge, pp 183–205

Halavais A (2019) How search shaped and was shaped by the web. In: Brügger N, Milligan I (eds) The SAGE handbook of web history. SAGE, London, pp 242–255

Jackson P (2010) Extremes of faith and nation: British fascism and Christianity. Religion Compass 4:507–527

Kellner MA (1996) God on the internet. IDG Books, Foster City, CA

Landes R (2007) Millenarianism and the dynamics of apocalyptic time. In: Newport KGC, Gribben C (eds) Expecting the end: millennialism in social and historical context. Baylor University Press, Waco, TX, pp 1–23

McMinn L (2001) Y2K, the apocalypse, and evangelical Christianity: the role of eschatological belief in church responses. Sociol Relig 62:205–220

Musso M, Merletti F (2016) This is the future: a reconstruction of the UK business web space (1996–2001). New Media Soc 18:1120–1142

Nanni F (2017) Reconstructing a website's lost past: methodological issues concerning the history of Unibo.it. Digit Hum Q 11. http://www.digitalhumanities.org/dhq/vol/11/2/000292/000292.html

Numbers RL (1992) The creationists: the evolution of scientific creationism. University of California Press, Berkeley, CA

Smith G, Woodhead L (2018) Religion and Brexit: populism and the Church of England. Relig State Soc 46:206–223

Sweetnam M (2019) Evangelicals and the end of the world. In: Atherstone A, Jones DC (eds) The Routledge research companion to the history of evangelicalism. Routledge, Abingdon, pp 178–197

Teszelszky K (2019) Web archaeology in the Netherlands: the selection and harvest of the Dutch web incunables of provider Euronet (1994–2000). Internet Hist 3:180–194. https://doi.org/10.1080/24701475.2019.1603951

Webster P (2017) Religious discourse in the archived web: Rowan Williams, Archbishop of Canterbury, and the sharia law controversy of. In: Brügger N, Schroeder R (eds) 190–203. UCL Press, London, p 2008

Webster P (2018) Technology, ethics and religious language: early Anglophone Christian reactions to “cyberspace”. Internet Hist 2:299–314. https://doi.org/10.1080/24701475.2018.1468976

Webster P (2019) Lessons from cross-border religion in the Northern Irish web sphere: understanding the limitations of the ccTLD as a proxy for the national web. In: Brügger N, Milligan I (eds) The historical web and digital humanities: the case of national web domains. Routledge, London, pp 110–123