

Looking for public archaeology in the web archives

Lorna Richardson

1.0 Overview

In an internet age, technology is expected to travel on an upwards trajectory, yet there is a significant issue for media historians in the future; the lack of material evidence for these iterations means that we risk losing an understanding of our social, economic, cultural and technological histories and our perception of these developments over time.¹ Without an understanding of the importance of web archiving, rather than simply 'digital preservation', we stand to lose not only a situated understanding of the emergence of the internet but the contextual understanding of the impact of these technologies on our own disciplines and intellectual traditions.

For archaeologists, the issue of archiving digital archaeological data has been the source of some apprehension for a number of years.² Jeffrey warned that the discipline of archaeology faced 'a second Digital Dark Age', as the adoption by archaeologists of social media and collaborative websites is 'running ahead of plans or policy to preserve the material generated'.³ The scholarly discussion of the impact of these issues on the archaeological sector by Morgan and Law and by Jeffery draw attention to the fact that the longevity of much of the material held on social media platforms is ultimately controlled by large corporations, so such data is inherently unstable.⁴ There are great methodological challenges for those who wish to use the web as a source for 'web historiography'.⁵ And for archaeologists, just as letters and site diaries are used to reconstruct the history of archaeology, perhaps the arrival of the 'Second Digital Dark Age'⁶ could see the discipline lose part of its history, if we cannot preserve our digital archaeologies for future research.

Technologies and websites labelled as 'Web 1.0' are, and have been, typically dominated by the presentation of static web pages containing hyperlinked content,⁷ which are far more suitable for web archiving than the rich content and connections of participatory media and what is known as 'Web 2.0'. As the web has evolved and participatory media has become a common experience for web users and web architects, the reality of archiving the 'live web experience'⁸ becomes more and more difficult. We need to understand the limitations of web archives, the incompleteness of data, and that the reconstruction of archived web material will always be subjective, and will never look the same as it did when it was live on the web in the first place.⁹ These provisos provide food for thought as digital analogies for the work of the archaeologist. It is worth exploring the work of Kirsten Foot and Steven Schneider who have conceptualized the 'web sphere' as an analytical approach to web archives:

We conceptualize a Web sphere as not simply a collection of Web sites, but as a set of dynamically defined digital resources spanning multiple Web sites deemed relevant or related to a central event, concept, or theme ... the boundaries of a Web sphere are generally delimited by a shared topical orientation across Web resources and a temporal framework.¹⁰

The categorization of the types of archaeological information that can be found on the internet, and therefore can be expected to be represented in the UK Web Archive, are numerous, and there is much crossover between these types of website records. This causes difficulty when separating out the sources within the web archives, so it is perhaps worth illustrating here the range of information provided online – and this list is not comprehensive:

1. Wikipedia entries on archaeological subjects.
2. Academic archaeological activity, such as university archaeology departments with courses, course materials, research projects and research news from the academic sphere.
3. Websites on archaeological legislation and planning for planners and developers.
4. Websites and blogs from excavations with specific single-site information.
5. Multi-site overviews of specific archaeological subjects or time periods.
6. Websites showing the application of archaeological techniques, 3D imaging etc.
7. Archaeological archives with data provided for public use.
8. Information from museum collections on object and find spots.
9. Social media sites, discussions and shared images.
10. Individual blogs discussing archaeological issues.
11. Thematic discussion lists, some of which are supported and maintained by academic institutions.
12. Pseudo-archaeological information – mythical archaeologies, ancient aliens etc.
13. Information from commercial archaeological organisations.
14. Community archaeology projects and societies.

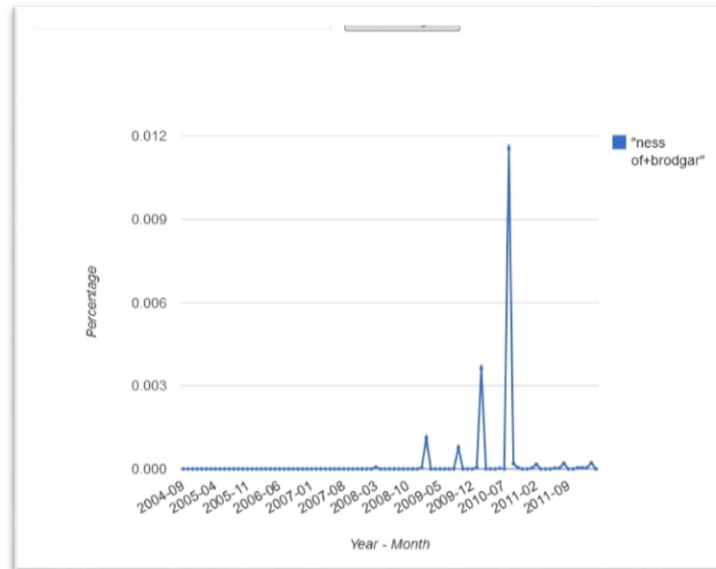
2.0 Methods and issues

With these provisos and the concept of a web sphere in mind, I have attempted to use web archives to create a web sphere for public archaeology as part of my research into the intersection of public archaeology and web archives. My research project 'Public archaeology online: a historical perspective' was devised with the intention of exploring the archive of UK web space from 1996 to 2013, using a prototype user interface developed by the 'Big UK Domain Data for the Arts and Humanities' project. I started out wanting to collate previous iterations of the presentation of archaeological information by archaeological organisations through the internet; and I also wanted to look at examples of popular, alternative/pseudo and academic archaeology websites and blogs, and explore the evolutions of digital

archaeologies during those 17 years. While this project was begun with good intentions, there were a number of problems along the way that affected my ability to undertake the type and depth of research that I had originally set out in my proposal, and this must be borne in mind when examining the results.

The main objectives of my original research proposal were quite complex. The intended methodology and subject area were adapted on several occasions over the course of the project: first, in April 2014, then later in August 2014, and finally, using the most functional interface, a significantly smaller scope and smaller datasets, in January and February 2015. I searched initially for public archaeology-related websites, but realized my choices were too large and reduced my scope to individual archaeological sites chosen because of their relatively high profile in popular understanding of archaeology. I searched for Stonehenge as my initial main archaeological site, expecting that this would provide a reduced number of returns and a manageable dataset, while including the widest possible range of archaeological material as well as pseudo-archaeological websites. Using both the prototype and the public UK Web Archive interfaces, this search created difficulties. The amount of data returned from the open UK Web Archive was relatively straightforward to gather. Unfortunately, however, this information could not be stored and interrogated further; and the prototype web interface, which would have supported deeper analysis, would not perform the same search without crashing.

I reduced the search scale, and was able to gather data on 3 sites in the final run of searches using the prototype interface. I searched for the 'Ness of Brodgar', an archaeological find made in the late 1990s in Orkney; 'La Cotte de St Brelade', an archaeological site on the island of Jersey, internationally famous for the Neanderthal discoveries made there; and 'Wayland's Smithy', a prehistoric barrow site in Oxfordshire. These searches yielded around 500 returns each, which were downloaded in CSV format directly from the prototype interface. The Ness of Brodgar N-gram below visualizes beautifully the release of information as the archaeological site progressed in its discoveries, and further work on marrying the contents of the web archive up with year-by-year developments on the site and with press coverage would develop this to its fullest conclusion. The process of interrogating this data was never fully realized due to time pressure, but I hope to take this forward into future research projects as outlined below.



Each of these attempts to explore the archives during the period of my research project proved to be frustrating, unsatisfactory and slow. The main lessons learnt from this exercise are as follows:

- The scope of the original search term matters. Given the enormity of the dataset, working with the least complex aims provides the easiest data set to handle, although this may not be the 'best' dataset to use for the starting hypothesis
- Be aware of the random nature of the returned data search, and be prepared to spend a lot of time weeding out returns that are irrelevant.
- The search process can be very, very slow, given the size of the archive; and the browser you use may affect this. I had most success with speed of return using the search facility in a single Firefox window. Using other tabs slowed the process down significantly; one search took over two hours to process a 'search failure', which was very frustrating.

It would be very useful to have the following suggestions considered for improvements to the web archive search interface:

- A better filtering mechanism for the types of site retrieved, with the addition of filtering for social platforms such as blogs as well as domain filters
- A more intuitive search function box, that is explicit about the years that can be searched and an easier interface for refining the data returned
- The ability easily to open a link in a new window/tab from the search interface
- A more in-depth tutorial on how to use the search interface effectively, both on the public facing interface and on the one that we were able to access as a prototype, would help with getting the most from these tools.

- The ability to save your searches on the prototype was fantastic, and the lack of this function on the public interface is frustrating.
- A system that indicated if the search had timed out, or there was an issue with the server – the number of occasions I received repeated proxy error notices (see below) led to a lot of frustration with the search facility.



Very few other archives have a similar sized dataset – over 2 billion resources are recorded in the index, and this is growing each year.¹¹ The difficulty of dealing with the size of the archive has proven to be a common thread among the researchers on the project, and my own project returned hundreds of thousands of web pages with archaeological content. The implications of a dataset of this size for historians of the web and contemporary archaeologists with an interest in digital practice are diverse. Search queries have had to be very specific, and as algorithmic prioritisation does not take place within the returns from the prototype web archive search, my initial project plan has been reduced in size to a manageable scoping exercise for a handful of key archaeological sites and terms. Interpreting patterns and contextualising the meaning of labels for archaeological sites has been undertaken with a great deal of subjectivity.

I remain unsure if my lack of technological expertise or inability to attend the support sessions provided by the project led to me struggling to use the search interface successfully. Perhaps these difficulties are simply part of attempting to interrogate this amount of data and there are refinements in the online search facilities that will be made in future to accommodate these issues.

3.0 Future research possibilities for public archaeology

There remains a huge amount of potential for a concerted effort to uncover the development of the presentation of public archaeology information in the web archive. There is a significant knowledge gap in the understanding of the early adoption of web-based forms of communications in archaeology, and further work in this area would be worthwhile. It is likely that the early adoption of websites in the discipline of archaeology in the United States will be reflected in the archives, and information available may be heavily oriented towards more recent iterations of public communications. There is an interesting opportunity to undertake small longitudinal case studies of organisational websites and pages through history (for example for English Heritage), alongside their contemporary iterations, in order to gain an understanding of how technological constraints might influence the making of meaning to varying degrees over time.

I can see value for the digital archaeology sector in exploring the iterations and content of websites associated with specific archaeological sites, to explore the evolution of archaeological interpretation to the public, and interpretation by archaeologists, as well as the adaptation of churning technologies and web trends for public archaeology. There is enormous scope for further multi-layered work on the impact of web archiving on our understanding of archaeology and for using an archaeological approach to explore, reconstruct and reimagine the technologies of past iterations of the web.

In terms of methodological approaches to these data, what Brügger refers to as ‘web philology’,¹² is just as much an ‘archaeology’ of web archives. An understanding of the applications of archaeological approaches to these data will be useful to the Internet Studies community as much as to scholars of media archaeology, digital archaeology or even digital public archaeology, since there are no longer many cultural phenomena that ‘cannot be analyzed and explained exhaustively if the Internet is not part of the analysis’.¹³ It is an area ripe for further research, at the almost-invisible boundary between media archaeologies and archaeologies of media.

¹ C. Van den Heuvel, ‘Web archiving in research and historical global collaboratories’, in *Web History*, ed. N. Brügger (New York, 2010), pp. 279–303; N. Brügger, ‘Web archiving – between past, present and future’, in *The Handbook of Internet Studies*, ed. M. Consalvo and C. Ess (Oxford, 2011), pp. 24–42; ‘Cowls, ‘Recreational bugs: the limits of representing the past through web archives’, Big UK Domain Data for the Arts and Humanities <http://buddah.projects.history.ac.uk/2014/06/20/341/> [accessed 30 Jan. 2015].

² Richards 2002; Richards *et al* 2010; de Silva & Henderson 2011; Hicks 2013.

³ S. Jeffrey, ‘A new digital dark age? Collaborative web tools, social media and long-term preservation’, *World Archaeology*, 44: 4 (2012), 553–70, at p. 555.

⁴ M. Law and C. Morgan, ‘The archaeology of digital abandonment: online sustainability and archaeological sites’, *Present Pasts*, 6(1):2 (2014), DOI: <http://dx.doi.org/10.5334/pp.58>; and Jeffrey, ‘A new digital dark age?’.

⁵ N. Brügger, ‘Web historiography and internet studies: challenges and perspectives’, in *New Media & Society*, 15: 5, (2012), 752–64, at p. 753.

⁶ Jeffrey, ‘A new digital dark age?’, p. 555.

⁷ L.-J. Richardson, ‘Public archaeology in a digital age’ (unpublished University College London PhD thesis, 2014) <http://dx.doi.org/10.6084/m9.figshare.1122524> [accessed 1 Feb. 2015].

⁸ Cowls, ‘Recreational bugs’.

⁹ Brügger, ‘Web archiving – between past, present and future’, p. 32.

¹⁰ K. A. Foot and S. M. Schneider, *Web Campaigning* (Cambridge, Mass., 2006), p. 20.

¹¹ P. Webster, 'Method in the web archive for the arts and humanities: a conference report', Webstory blog (2015) <http://peterwebster.me/2015/01/20/method-in-the-web-archive-for-the-arts-and-humanities-a-conference-report/> [accessed 14 May 2015].

¹² Brügger, 'Web historiography', p. 39.

¹³ Brügger, 'Web historiography', p. 39.