

# Capture, commemoration and the citizen historian: digital shoebox archives relating to PoWs in the Second World War

Alison C. Kay, iSchool, Northumbria University at Newcastle

**General subject areas:** History and Information Science

## **Broad foci:**

- Thematic focus: sites relating to Second World War prisoners of war
- Methodological focus: creating and curating a corpus
- Analytical focus: tracking the digital footprint of citizen historians

## Introduction

Although everyone became a mass publisher of digital data, fifty years from now most people will not have access to any registries of their memories, such as photos of their loved-ones. Memories play a crucial role in human behaviour and losing access to them may have unexpected impact on modern societies. Web archives will probably be the only source of personal memories to many people.<sup>1</sup>

What is the historical record in the light of the digital emancipation of the citizen historian? This case study sits within a wider discussion on who practices history and who can contribute and edit the historical record – and how. It is also a first step in demonstrating how web archives can be used by academic historians to tap into un-catalogued, privately-held archival material. Finally it is part of a wider research project contributing to preservation concerns in light of the accessibility of digital tools.

The intention at the outset was to use the Web Archiving Researcher Bursary to explore the possibilities of a case study on the digital footprint of citizen historians (grass-roots historians and community groups) in the web archive. Setting some parameters on this grand task, the scope was limited to digital shoebox archives (micro-collections and narratives of lived experience) relating to British prisoners of war during the Second World War. This situated the case study alongside my broader, collaborative, 'proof of concept' project funded by Semantic Media on the digital preservation of shoebox archives and the possibilities of semantic technologies (linked data) in connecting with the wider web.

What might be included in these shoebox archives in their physical form? They include items such as photographs, letters and diaries, maps, medals, illustrations, newspaper cuttings and items of personal significance such as medals. The broader project seeks to capture, curate and connect the contents of 'made digital' micro collections in accessible (open-source), sustainable, online collections. The web archive case study approaches from the other direction and represents an early foray into the possibilities of identifying and collating citizen historians' existing online collections, as captured in the web archive.

Why is this important? In recent decades citizen historians have digitised the contents of their shoebox archives in the form of blogs, self-designed websites and contributions to digital community projects. These digital footprints are sites of commemoration but they also represent endeavours to contribute to the historical record. (We know little about their motives beyond this.) Unfortunately, these individual sites are rarely connected to broader historical research nor used by scholars in universities or memory institutions and thus are 'disconnected' from the broader historical record to which they relate. Also, unless subsequently captured in web archives they are vulnerable in terms of preservation. This has become a pressing issue for archivists but one about which historians have been slow to voice their concerns. Yet UNESCO now recognizes digital information as a form of heritage and notes that substantial quantities of this digital heritage are being lost. As Gomes and Costa warn us: 'The fast and unexpected disappearance of information from the web will prevent future historians from accessing and researching valuable information sources. As it has been performed for printed media for centuries, the information published on the web must be archived and preserved to enable future historical research'.<sup>2</sup>

The physical objects on which these digital micro-collections are based may in turn also be vulnerable due to over-confidence in the sustainability of current technology and people's over-estimation of the life-course of the 'made digital' record. Even when captured in web archives, online micro-collections are still at risk if they remain unconnected and un-analysed through lack of curation.

Hence this case study set out to explore the developing search interface of the British Library's web archive and work on a methodology for isolating web pages containing the narrated stories or 'made digital' artefacts of prisoners of war of the Second World War. Proceeding through a series of three explorations while the search interface was evolving, the initial focus was necessarily on answering the first two of three key 'parent' questions:

- 1) Could the footprint of citizen historians relating to PoWs in the Second World War be identified in the web archive?
- 2) Could relevant archived web pages, once identified, be grouped into a study collection for analysis?
- 3) What does this web archive collection reveal about citizen historians' early uses of digital technology, efforts to connect their stories to those of others, and attitudes to who writes history?

## Methodology and early findings (research questions 1 and 2)

The initial aspirations of this case study were restricted by the global nature of the web. The UK Web Archive focuses on UK hosted domains. Yet the blogging market, even for blog writers based in the UK, has been dominated by WordPress, Blogger and similar platforms. These are largely provided by US-based companies. Unless the creator of a blog had purchased their own UK-hosted domain to use in conjunction with one of these services, this ruled out the possibility of finding blog pages created by citizen historians to share their or their relatives' artefacts or narrated experiences as PoWs. (These could of course be searched for in a separate project using the Internet Archive in the US.) Hence the case study proceeded with the intention of searching for UK domain web pages and contributions to community, media and memory institution projects.

Within this case study, exploration of the web archive took place over a series of five stages and was very much an iterative process as early methodological findings were noted:

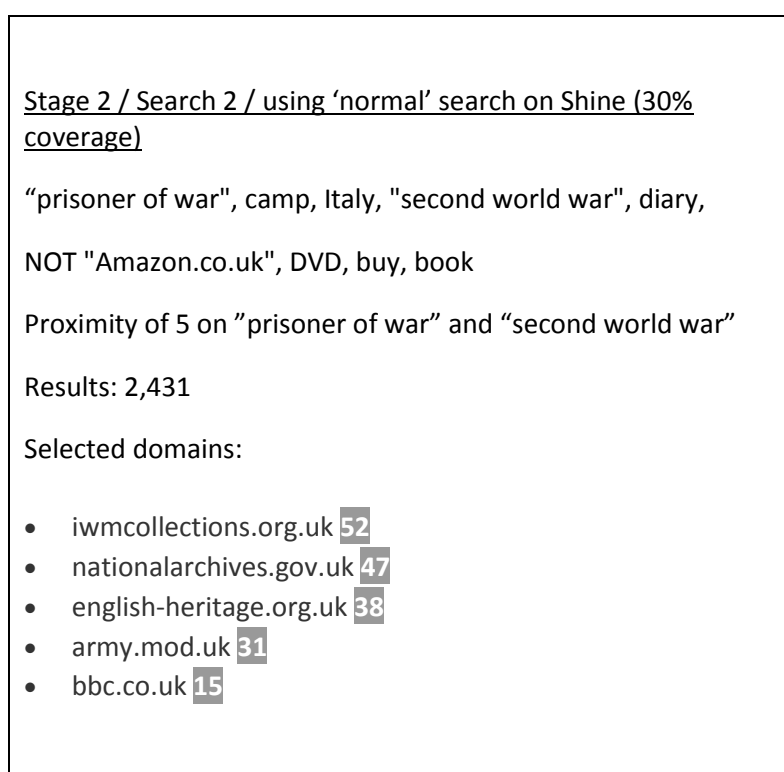
- *Stage 1* (April–May, 2014): Initial exploration, keyword identification, and application of basic filters using AADDA-discovery (subset of UK web domain dataset);<sup>3</sup>
- *Stage 2* (July, 2014): Searching using strings of tested keywords, identification of diverting keywords, application of NOT function;
- *Stage 3* (September, 2014): Adoption of Shine search interface for 'normal' (not advanced) searches. NB: limited to 30% of the UK Web Archive contents;
- *Stage 4* (October, 2014): Full use of Shine – prototype search interface –querying full UK Web archive contents 1996–2010;
- *Stage 5* (January, 2015): Use of login and save function to apply lessons learned in all of the above stages and to create a corpus of PoW-related web pages from the UK Web Archive.

Initially, as with the case studies of the other bursary holders, the first hurdle for researchers was that of being faced with a very large index that needed refinement. In Stage 1 (April–May, 2014) the approach was purely exploratory, using a subset of the UK web domain dataset. There was very limited means of saving each query (and hence building a corpus), although it was possible to use a bookmark in the web browser (Chrome) or to make a copy of the URL and save it in a spreadsheet. Between each of the subsequent stages, with the help of regular bursary holder meetings, the interface for searching the web archive was developed and became more responsive to the needs of the various projects. For example, the early stages were dominated by trial and error searches, some basic filtering for suffix

(e.g. co.uk or gov.uk) and feedback to the British Library's web archive team on the interface. By Stage 3 (September, 2014) however, researchers were able to begin using Shine,<sup>4</sup> the new prototype search interface, which queries the full SOLR index from 1996 to 2010. At this stage a login and search-save function had been added, making it possible to take the step forward from experimental searching (and saving in informal ways such as copy and pasting the URLs into a worksheet) to saving within the Shine search interface itself. Following feedback from the bursary holders, further refinement of the interface followed in which the core filtering commands were added to the advanced search page.

I then selected keywords (see appendix Table 1) and these were exploited in various combinations of strings, finally producing 12 core strings (see appendix Table 2). However, it proved challenging to identify a string that would draw up results skewed towards personal narrative, commemoration or the digital preservation of artefacts. The search results were not specific enough yet to create a worthy corpus. Also, as more of the archive came online for searching via the Shine interface, the number of search results for each string vastly increased. The tools for further filtering became more vital at this stage but also more accessible within an advanced search screen or the results page. This enabled filtering by general content type, author, crawl year, public suffix, or domain. The latter was particularly useful in this case study as it was immediately apparent that (with other filters applied), web pages from memory institutions and the media domains constituted a significant proportion of the results (see Figure 1.). Nonetheless, DVD and book advertisements continued to dominate the search results in all the initial strings and the NOT function for 'Amazon', 'DVD', 'book' and 'buy' was identified as an essential to filter.

**Figure 1:**



After eliminating the 'Amazon effect' and experimenting with various proximity limits between key words, the search results still contained a sizeable body of irrelevant material. This said, the search result numbers were significantly more manageable. For example:

**Figure 2:**

Stage 4 / Search 9 / using 'advanced' search on Shine (full coverage)  
"prisoner of war", "Second World War", diary  
NOT Amazon, DVD, buy, book  
Proximity of 25 words between "Prisoner of war" and "Second World War"  
Results: 206 (or 94 on proximity of 10; 40 on proximity of 5)

Note that even a proximity setting of 5 for the terms 'prisoner of war' and 'Second World War' on the above search picked up irrelevant results such as a mathematics evening at the University of Surrey to inspire teachers.<sup>5</sup> However, the same search string without filters applied produced an unmanageable 53,638 results. A fellow bursary holder summed this up precisely as there being 'islands of valuable resources within a sea of irrelevant material'.<sup>6</sup>

In Stage 5 (January, 2015), the 12 search strings were re-run with filters, excluded domains<sup>7</sup> and various proximity limits (see appendix Table 3). This collectively produced 51,788 results, however 27,061 of these were from string 11 alone. Excluding this string, the unfiltered search total was 24,727. After getting rid of the Amazon effect, *n* became 2,894 and this figure held with proximity searches of 25, 10 and 5 on the key words 'prisoner of war' or 'POW' and 'second world war'. On average (excluding string 6 as an outlier), the filtered results represented 11% of the original index, varying between 5.1 and 20.3 for each individual string. This demonstrates the importance of identifying potentially irrelevant sites you might catch at the outset. The proximity results were interesting because for this case study setting limits between significant key words appeared to have little impact at all. Yet judging from the discussions at the Web Archive Researcher Bursary Holder meetings, this seemed to be unusual.

Crucially, the ability to exclude individual search results, save the remainder and add them to a corpus had been added to the Shine interface by this stage. To do this, however, would mean manually visiting each of the 24,727 keyword search string results, one-by-one, and excluding irrelevant and duplicate versions of pages.<sup>8</sup> It was possible to exclude a host (in theory) and this would have speeded the process up but this did not prove very successful. A preview or snippet view function for each individual search result would also have aided

the exclusion process but although discussed in the October meeting did not become a feature of the interface during the timeframe of this case study. This hurdle aside, if the exclusions could be made manually, it was now possible to create a corpus for each search string.

As a test and taking the string with the smallest set of filtered results (string 6) for initial assessment of the practicalities of manual filtering, the issue of duplicate versions of the same site became immediately apparent: four of seven results were war-related but not sites of personal commemoration or family history. Rather they were related to papers held by King's College London archives on Major General Sir William Ronald Campbell Penney, KBE CB DSO MC (1896–1964). The remaining three, however, were relevant to this case study and included a newspaper interview and pages from two separate projects to record wartime stories: the independent 'Wartime Memories' and the BBC's 'The People's War'.<sup>9</sup> Given time constraints and allowing for the likelihood of some overlap between the results of the 11 workable strings, two were selected for the manual exclusion process: strings 2 and 7 (plus the seven from string 6 in the test); totalling 445 filtered search results out of a possible 2,894, just over 15% of all the search results. The process proved to be a highly frustrating one, with frequent proxy errors when exclusion requests were made. On the first three attempts, having reduced the string 2 results by half by excluding various hosts, the results number suddenly sprang up from 110 to 2,145 after selecting the next page of results.<sup>10</sup> Seemingly, all NOT filters were removed by the interface. Although a potential workaround was identified – that of selecting for adding to a corpus rather than deleting out of the string results – time did not permit this to be carried out. It would still have necessitated opening each result to assess its relevance.

### Analysis of corpus (research question 3)

3) What does this web archive collection reveal about citizen historians' early uses of digital technology, efforts to connect their stories to those of others, and attitudes to who writes history?

Although more time was needed to create a corpus and analyse its contents, some overview observations of the content of the search strings was possible. What is immediately obvious is that the archive has captured material from a handful of ongoing and completed online projects to gather together wartime memories and retellings of family members' experiences during the great wars. These projects are the citizen-historian driven 'Wartime Memories' project, the media-driven BBC 'The People's War' and the National ex-Prisoner of War Association. These projects are still viewable live on the web. Although 'The People's War' has ceased gathering material, it sought contributions from the public between June 2003 and January 2006, resulting in 47,000 stories and 15,000 images.<sup>11</sup> The 'Wartime Memories' project website is a volunteer undertaking, supported by donations and a grant of £33,800 from the Heritage Lottery Fund in 2005.<sup>12</sup> Corresponding with a project representative, I ascertained that they have been collecting 'stories' since 1999. At the current time they have published online 7,266 stories relating to the First World War and 8,730 stories relating to the Second World War. There are a further 4,000 stories on old

static web pages which they are now transferring to the current dynamic website. The backlog of submissions (unsorted) is huge: 1,163 stories are in the final stages of being checked before they go live; a further 2,689 are being edited before the checking process. Additional submissions yet to undergo either of these two pre-publication stages number in the region of 2,033. In all they estimate approximately 25,000 submissions in their keeping. In addition, the Heritage Lottery Fund monies permitted the creation of an archive of public contributions – not strictly narratives, but diary entries, photos, newspaper clippings and documents.<sup>13</sup> The ‘Wartime Memories’ project is clearly a commendable undertaking, carried on the shoulders of dedicated volunteers for over 15 years.

I was interested to know how much of the ‘Wartime Memories’ or ‘The People’s War’ projects had made it into the British Library Web Archive. An advanced search for ‘wartime memories’ (with a proximity limit of one between these two words) and the domain specified in the ‘within resources’ field revealed 9,785 results. It is difficult to say (given time constraints) how many of these are versions of the same web page and hence how many of the c.15,000 stories have made it into the archive. It is unlikely to be 9,785. There are some useful ‘Wartime Memories’ indexing pages and forum discussion captured in the web archive. The indexes reveal the names of various posted stories, e.g. ‘Seven brothers went off to war’, ‘Italy and back again’ and ‘The Russians are coming’.<sup>14</sup> These illustrate a desire to tell a packaged story; one with drama, humour and affection.

A similar attempt was made to identify how many of the 47,000 Second World War ‘People’s War’ stories had been saved in the web archive. A search for ‘WW2 Peoples War’ with a URL specified of [www.bbc.co.uk/history/ww2peopleswar/](http://www.bbc.co.uk/history/ww2peopleswar/) provided 346 results. As there would appear to be one contribution per web page on the live site and the project claims 47,000 stories, the archive would seem to have captured a very small proportion.

As a historian I have to remind myself that the online web is gone. We have ‘representations’ in the form of web-archived material and this is incomplete in the sense that the full version of a multi-page commemorative website may not have been captured. Although there may be multiple versions of what is captured, this does not mean that we can see key edited changes over time with any consistency – a personal website may go live well before the narration of the display of photos has been uploaded. Hence in any study based on web-archived material we have to be aware that there are gaps caused by temporal and spatial inconsistency in the archiving process.<sup>15</sup>

## Reflections

In a web archive you simultaneously have too much and too little.

Niels Brügger, *Twitter*, 2014<sup>16</sup>

Addressing research questions (1) and (2), the practicalities of using the search interface as it developed took up the lion's share of the time on this case study and discussions in this regard were always illuminating for those of us who had not taken part in the previous project using the AADDA-Discovery interface.<sup>17</sup> An obvious hurdle in working with an evolving interface was that not every function worked perfectly. As the Shine interface was not always fully accessible when I had research windows, tight scheduling around other research projects did present some challenges. The keyword search function was also only partially successful in drawing relevant material out of the archive. It drew a great deal of irrelevant material too, even in the advanced search. This is partly due to the differing 'information intents' of scholars like myself banging up against the text-based searching limitations and results being based on instance occurrence with no consideration of context, relevance or relationship between web pages. It will be interesting to follow the development of prototype search engines with specific historical search intents, such as that under development at the L3S Research Centre in Hanover.<sup>18</sup> No doubt searching performance will improve and providing broader access to the archive via other means is hopefully high on the agenda.

Using web archives is a relatively new experience for all the researchers and the right approach is still very much up for discussion. One of the most beneficial aspects of the Web Archive Researcher Bursary format has been the meetings – interdisciplinary gatherings to discuss the needs of the case studies and to exchange experiences with other researchers from other disciplines (and my own). This was an important part of the process of coming to understand the nature of web archive content as distinct from more familiar sources: how to access it and connect it to broader narratives. As with any exploratory study, it was heartening to share negative experiences in addition to positive ones.

The Web Archive Researcher Bursary has provided an excellent opportunity to get to know the British Library's web archive better, to read around the topic and to start thinking about how to unleash the archives' potential for scholars at the level of the coalface. Reflecting on my own experience, the need to establish a best practice for historians working with web archives is clear, if results are to be reproducible and hence verifiable. Working with web archive content is more slippery than with material in traditional archives. Web pages are not even copies of the original; they are versions and there can be many versions captured or none (and hence whole sites are not always intact within the archive). Getting relevant web pages out of the archive was not as obvious as it had at first seemed, not only because key words can be culturally and temporally specific but also because the search picks up advertising banners as well as the main contents.

Documenting the use of web archive material is also going to be an ongoing issue. Indeed, the bursary holders had several useful discussions on citation and we were not always in agreement. Although there is a danger of concealed research paths when using digital sources if we can't easily 'unpack the blocks and shards of detail that make up an argument',<sup>19</sup> the footnote is less a pillar of replicability than many would like (even if not openly admitted).<sup>20</sup> Nonetheless, there is a growing disciplinary awareness of the possible disconnect between the research process and representation, further complicated by the evolution of key word searching. Tim Hitchcock warns that this 'removes the framework of



source criticism and classification that we have come to rely upon’ and puts us in danger of ‘research roulette dressed up as formal scholarship’.<sup>21</sup> The use of web-archived material will require transparency if it is going to prop up REF-worthy publications.

Carrying out this case study brought home to me the emergent discussion that there is a clash between the way we now access, store and share material and the way in which we think, discuss and write about the past.<sup>22</sup> Historians need to be contributing to discussions today about the sources of tomorrow. Engaging with web archives is a good starting point for analyzing what is different in working with this new source and what that means for the future of historians. We need to be more aware of how these sources and their archives are constructed and link together, and their limitations.

However I have also been reflecting that dealing with a perceived ‘sea of information’ is not at all a new problem. In 1869 the *Pall Mall Gazette* published the grumble:

human eyes and human hands cannot possibly work through a century of such agglomeration. The human mind will despair, perhaps, of power to deal with the illimitable mass. May we hope that when things come to such a crisis, human labour of the literary sort may in part be superseded by machinery? Machinery has done wonders, and when we think of what literature is becoming, it is certainly to be wished that we could read it by machinery, and by machinery digest it.

*Pall Mall Gazette* (15 September, 1869)<sup>23</sup>

Mass printing was worrisome in its volume in years past, in the way that the archived web is challenging today. Centuries of print have not all been catalogued and saved and the material that remains made it through various preserving filters. It is much the same for material published on the web. It is not all captured in web archives, indeed whole websites are not even captured – although the filter has arguably less human persuasion.<sup>24</sup> E. H. Carr likened the trade of the historian to fishing in a vast and sometimes inaccessible ocean.<sup>25</sup> The archived web certainly provides that opportunity but historians will require a new rod and some practice in using it effectively.

## Appendix

**Table 1: Key search words**

Prisoner of war	POW	P.O.W	prisoner
camp	campo	stalag	oflag
Italy	Germany		
Second World War	2WW		
diary	letter	letters	
captured			
airman	soldier	officer	guard

*Note: Although other countries had PoW camps, those in Italy and Germany are relevant to the broader research project within which this case study sits.*

**Table 2: Keyword strings**

1	"prisoner of war", camp, Italy, "second world war", diary
2	"POW", camp, Italy, "second world war", diary
3	"prisoner of war", camp, Germany, "second world war", diary
4	"POW", camp, Germany, "second world war", diary
5	"prisoner of war", camp, Germany, "second world war", stalag
6	"prisoner of war", camp, Italy, "second world war", campo
7	"prisoner of war", camp, Italy, "second world war", letters
8	"POW", camp, Italy, "second world war", letters
9	"prisoner of war", camp, Germany, "second world war", letters
10	"POW", camp, Germany, "second world war", letters
11	soldier, captured, "second world war"
12	airman, captured, "second world war"

**Table 3: Stage 5 key string search results with filters and proximity limits.**

String	Total <i>n</i>	<i>n</i> with NOT function applied	<i>n</i> with NOT + Proximity 5 applied	Filtered <i>n</i> as % of Total <i>n</i>
1	1772	239	239	13.5
2	2195	249	249	11.3
3	2387	258	258	10.8
4	2625	293	293	11.2
5	1507	306	306	20.3
6	137	7	7	5.1
7	2256	189	189	8.4
8	2582	267	267	10.3
9	3265	302	302	9.3
10	3503	362	362	10.3
11*	27,061	25,712	25,712	95
12	2498	422	422	16.9
<b>Total</b>	<b>51,788</b>	<b>28,606</b>	<b>28,606</b>	<b>55.2</b>
Total excluding string 11	<b>24727</b>	<b>2894</b>	<b>2894</b>	<b>11.7</b>

\* String 11's results were clearly unmanageable and an outlier in terms of working out averages.

<sup>1</sup> Daniel Gomes and Miguel Costa, 'The importance of web archives for humanities', *International Journal of Humanities and Arts Computing*, 8.1 (2014), 106–23, at p. 114.

<sup>2</sup> Gomes and Costa, 'The importance of web archives', p. 107.

<sup>3</sup> <http://www.webarchive.org.uk/aada-discovery>.

<sup>4</sup> <http://www.webarchive.org.uk/shine>.

<sup>5</sup> 'Mathematics evening at UniS', [http://web.archive.org/web/20030116231401/http://www.maths.surrey.ac.uk:80/announce/Outreach/evening\\_total](http://web.archive.org/web/20030116231401/http://www.maths.surrey.ac.uk:80/announce/Outreach/evening_total) [Internet Archive, site captured 16 Jan. 2003].

<sup>6</sup> *Twitter*: @jfwinters Saskia Huc-Hepher: Islands of valuable resources within a sea of irrelevant material #buddah14 2:56 PM – 26 Feb 2014. (This remark was made during the AHRC-funded workshop held on Wednesday 26 February at the University of London Senate House. The workshop was organised as part of the Big UK Domain Data for the Arts and Humanities (BUDDAH) project.)

<sup>7</sup> For example, it is possible to exclude Amazon at this level rather than via the NOT function.

<sup>8</sup> In a web archive, there may be multiple versions (not copies) of an original web page.

<sup>9</sup> 'Desmond's GREAT escape', 16 Jan. 2004 [http://web.archive.org/web/20040910084750/http://iccheshireonline.icnetwork.co.uk:80/100news/chesterchronicle/content\\_objectid=13823454\\_method=full\\_siteid=50020\\_headline=Desmond-s-GREAT-escape-name\\_page.html](http://web.archive.org/web/20040910084750/http://iccheshireonline.icnetwork.co.uk:80/100news/chesterchronicle/content_objectid=13823454_method=full_siteid=50020_headline=Desmond-s-GREAT-escape-name_page.html) [Internet Archive, site captured 10

---

September 2004]; ‘The World War Two Memories Project – Can you Answer?’ <http://web.archive.org/web/20050215190706/http://www.wartimememories.co.uk:80/answer4.html> [Internet Archive, site captured 15 February 2005]; ‘WW2 People’s War: James Kenna – The Story of a Prisoner of War’ <http://web.archive.org/web/20091123133543/http://www.bbc.co.uk:80/ww2peopleswar/stories/83/a7381983.shtml> [Internet Archive, site captured 23 November 2009].

<sup>10</sup> This happened on both 24 and 29 January 2015.

<sup>11</sup> WW2 People’s War <http://www.bbc.co.uk/history/ww2peopleswar/> [accessed 28 January 2015].

<sup>12</sup> The Wartime Memories Project <http://www.wartimememories.co.uk/main.html> [accessed 28 January 2015].

<sup>13</sup> Correspondence received from Angela on behalf of the Wartime Memories Project, 29 January 2015.

<sup>14</sup> WW2 People’s War <http://web.archive.org/web/20041115005113/http://www.bbc.co.uk:80/dna/ww2/C1182> [Internet Archive, site captured 15 November 2004].

<sup>15</sup> Niels Brügger, speaking at the Web Archive Researcher Bursary Holder’s meeting, 16<sup>th</sup> September, 2014.

<sup>16</sup> *Twitter*: @jfwinters Brugger: in a web archive you may have multiple versions (not copies) of an original which no longer exists #buddah14 4:12 PM – 26 Feb 2014.

<sup>17</sup> <http://www.webarchive.org.uk/aada-discovery> [accessed 19 June 2015].

<sup>18</sup> Jaspreet Singh, ‘Search results for historical material’, Big UK Domain Data for the arts and Humanities (published 22 October 2014) <http://buddah.projects.history.ac.uk/2014/10/> [accessed 24 January 24 2015].

<sup>19</sup> Tim Hitchcock, ‘Confronting the digital or has academic history writing lost the plot?’, *Cultural and Social History*, 10.1 (2013), p. 12.

<sup>20</sup> On the politics of footnotes and history, see Ludmilla Jordanova, ‘Historical vision in a digital age’, *Cultural and Social History*, 11.3 (2014), p. 346.

<sup>21</sup> Tim Hitchcock, ‘Confronting the digital’, pp. 14, 18.

<sup>22</sup> Andrew Prescott, ‘I’d rather be a librarian’. A response to Tim Hitchcock, ‘Confronting the digital’, *Cultural and Social History*, 11.3 (2014), p. 335.

<sup>23</sup> Mark Knight, ‘The implications of social media’, *Cultural and Social History*, 11.3 (2014), pp. 332–3.

<sup>24</sup> Or, at least more democratic if link-tracking has been used to build the archive as opposed to selection-by-archivist.

---

<sup>25</sup> E. H. Carr, *What is History?* (Penguin, 1987; first published 1961), p. 23.