City Research Online

City St George's, University of London

Citation: Roland, L. & Bawden, D. (2012). The Future of History: Investigating the Preservation of Information in the Digital Age. Library & Information History, 28(3), pp. 220-236. doi: 10.1179/1758348912z.00000000017

This is the unspecified version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/3186/

Link to published version: https://doi.org/10.1179/1758348912z.00000000017

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

The future of history: implications of preservation of information in the digital age

Lena Roland and David Bawden
Department of Information Science, City University London

Abstract

This study investigates the challenges of preserving information in the digital age, and explores how this may affect the future of historical knowledge.

The study is based on literature analysis and a series of semi-structured interviews with 41 historians, archivists, librarians, and web researchers.

While the results reject the idea of a single 'digital black hole' in historical records, they emphasise the importance of the issue for the future of history, and the complexity of the solutions to be adopted. The need for planning, for education, and for cooperation between historians and the information professions is emphasised.

Introduction

> 'the preservation and re-use of digital data and information forms both the cornerstone of future economic growth and development, and the foundation for the future of memory… Any future scholarship depends upon the survival of the digital resources in accessible and intelligible contexts'. (Seamus Ross, 2000)[1]

> 'historians and citizens of the future will find a black hole in the knowledge base of the 21st century'. (Lynne Brindley, 2009)[2]

The preservation of digital material is of evident importance for historians of the future, if Brindley's 'black hole' is to be minimised. The problems, however, are complex. Most obvious are the technical problems, of the need to cope with obsolete hardware and software. Additionally, the existing legal framework does not reflect the digital age, current law makes it difficult for memory institutions to capture and preserve digital material. There are also social and cultural challenges to be met. Should everything in the digital realm be preserved? If not, what should be preserved, and just as importantly what should not? And who gets to decide? Who should be responsible for safeguarding digital data? How and when can the ever changing World Wide Web be captured, and, with its mutable nature, how will future historians establish provenance?

The purpose of this study is to investigate the implications this uncertainty poses for the future of history. What can be done today to ensure future historians will have access to a rich historical record so they can tell the story of our time to future knowledge thirsty generations? While the history of the information age *per se* has been much discussed, the historical preservation issues raised by communication technologies have been little considered; see Tredinnick (2011) for one of the few such analyses.[3]

A content analysis of the literature was conducted, to identify the main concepts and issues of digital preservation, as it relates to the history profession. These were then explored in a series of semi-structured interviews with historians to understand their awareness of and engagement with digital preservation, and with library/information professionals, including archivists and web specialists, to explore their opinions on digital preservation and how it relates to the history profession. In total, 41 individuals were interviewed, in person, by telephone or email exchange, as practicality dictated. More details are given in Roland (2010).[4]

The findings of the literature analysis and survey are now reported in six sections, dealing respectively with: digital preservation; awareness and engagement; the ephemeral web; selection and curation; the historian's craft; and a 'digital black hole'?

2

DIGITAL PRESERVATION

Beagrie & Jones defined digital preservation as

> 'the series of managed activities necessary to ensure continued access to digital materials for as long as necessary'[5] while Conway points out the purpose for digital preservation is 'to ensure protection of information of enduring value for access by present and future generations'.[6]

The issue of digital preservation supersedes national boundaries and is therefore, a global issue, or indeed as Stewart Brand put it a 'civilisation issue'.[7] A comprehensive bibliography of digital preservation and curation is available (Bailey 2011),[8] and only a few references are given here.

The exponential growth of digital data and the expansion of communication platforms seen in recent years have compounded the need for digital preservation and curation. Digital data is vulnerable to loss or decay and mostly there is no analogue equivalent. As Levi notes:

> 'the digital preservation challenge is most acute when it comes to items that were produced in digital format—that is, born digitally. Such items represent over 93% of the world's information… a fact that makes the preservation of digital information critical'.[9]

The UK does not yet have an official national strategy for digital preservation. The absence of an official framework means the preservation of digital information is mostly done on a case by case basis: hence the process is rather unsystematic and fragmented.

A digital preservation 'community' has emerged to respond to the growing challenge of preserving data, knowledge, history in the digital age and organisations such as the Digital Preservation Coalition (DPC), Joint Information Services Committee (JISC) and the Digital Curation Centre (DCC) have made great progress in raising awareness of the issue and promoting best practice.

The overwhelming focus of the literature regards digital preservation as a technical problem. Without question the rapid pace of technological change is a major threat to the longevity of digital data. Hardware may become obsolete and storage formats such as floppy discs and video cassettes for example outdated and potentially unreadable. As Danny Hillis points out,

> 'thousands of years ago we recorded important matters on clay and stone that lasted thousands of years. Hundreds of years ago we used parchment that lasted hundreds of years… we are now in a period that may be a maddening blank to future historians--a Dark Age--because nearly all of our art, science, news, and other records are being created and stored on media that we know can't outlast even our own lifetimes'.[10]

The digital preservation community rely on just a handful of case studies to illustrate the perils of obsolete technology and the fragility of digital data. There are several technical solutions in place to 'save' digital material, the most commonly applied solutions are migration or emulation techniques. A small number of case studies are frequently cited to show the need for this, including the Domesday Book project (Nath 2010)[11], the NASA Viking Landers (Digital Preservation Coalition 2006)[12], and the US Census (Waters and Garrett, 1996)[13].

Indeed, the lack of concrete evidence led Ross Harvey to provocatively ask, 'so where's the black hole in our collective memory?' pointing out that many of the cited case studies highlight data recovery, not data loss. He declares there needs to be more effort,

> 'into quantifying the extent of digital information loss or compromise, or, at the very least, to document more examples to supplement the few specific studies currently available'.[14]

Nonetheless, the rhetoric has continued, as evidenced by the declaration of the Chief Executive of the British Library:

> 'researchers of the future (will) find a black hole when researching late 20th Century history as much of our digital history has decayed and become digitally corrupted'.[15]

This dramatic statement captured the attention of the national press receiving extensive coverage and in doing so raised further awareness of the need for more impetus in this area. Other significant literature contributions making essentially the same point, while emphasising that technical issues go alongside economic and social concerns include Lavoie and Dempsey (2004)[16], Smith Rumsey et al (2010)[17], Sierman (2009)[18], and PLANETS (2010).[19]

It seems self-evident that the historian is an important stakeholder in effective long-term digital preservation. Yet there is a notable lack of literature that looks specifically at the historical perspective of digital preservation and the impact on the future of historical records. As Rosenzweig, one of the first historians to ask questions about the impact of digitisation on his profession, observed in 2003:

> 'archivists and librarians have intensely debated and discussed digitisation and digital presentation for more than a decade. They have written hundreds of articles and reports, undertaken research projects, and organized conferences and workshops…(however) academic and teaching historians have taken almost no part in these conferences and have contributed almost nothing to this burgeoning literature. Historical journals have published nothing on the topic'.[20]

Eight years later there are still just a handful of historians who have added their voice to the literature. Most are based in the USA, and appear to be earlier in addressing this issue than their European counterpart; a rare UK perspective is provided by Hampshire and Johnson (2009).[21]

In his paper, ‘Scarcity or Abundance? Preserving the Past in a Digital Era’, Rosenzweig raises many vital questions about the impact of digitisation on the future of his craft. He points out,

> ‘the “system” for preserving the past that has evolved over centuries is in crisis, and historians need to take hand in building a new system for the coming century. Historians also tend to assume a professional division of responsibility, leaving these matters to archivists. But the split of archivists from historians is a relatively recent one digital data requires yet more rethinking–about whether we should be trying to save everything, who is "responsible" for preserving the past, and how we find and define historical evidence’.[22]

This statement raises several important points, discussed later.

There is also a concern that too much emphasis on the digital environment could create a bias in the historical record, as highlighted by Leary (2010), who identified two dangers in digitising historical collections; the ‘illusion of completeness’ and the ‘illusion of accuracy’. He urged his colleagues to be aware that ‘voices are still missing in the online collection’ and that the online realm is still ‘a limited history’.[23] The respondents to the survey largely confirmed these issues.

AWARENESS AND ENGAGEMENT

Unsurprisingly, all respondents in all disciplines are aware of digital preservation; however, there was a very mixed level of engagement. The findings suggest the majority of historians are engaged with digitisation but only one aspect of it – the digitisation of analogue material.

The idea that many historians are concerned with born analogue material, as opposed to the born digital material, was reinforced in the survey, as exemplified by a UK archivist:

> 'Because of the environment within which historians work, necessarily focussing on the short and medium term for their research and teaching, it is difficult for many of them to engage with the changes to the historical record that will inevitably occur over the next 10 to 20 years. This timeframe is sufficiently far away for many of them not to have engaged with such changes so far. The engagement is likely to occur only when sufficient quantities of born electronic records are made available to the public'.

Several respondents think historians should be more concerned about the availability of historical sources of the future and engage more. Some historians clearly believe their colleagues are not sharing their long-term view of history, describing them as "myopic" and displaying a "general inertia: they "have kind of got it conceptually, but are uninterested".

Yet the idea that it is not necessarily the historian's job to focus on the future was also strongly expressed. As pointed out by a UK librarian:

> 'Historians today - their interest is exactly in right place. Historians of today are interested in history let's say that is anything more than 30 years old and the things that were not born digital. But the historian in 2100 will be very interested in the born digital materials. If we do our jobs right they will find it very easy. If we do our jobs wrong they will discover the materials are not there or if they are it will be extremely complicated to figure it out'.

While digitisation of analogue collections is recognised as progressive in that it increases access to historical resources and knowledge, as well as enabling a more democratic, alternative history to be told, others regard the digitisation of born digital material such as blogs and datasets as more pressing due to its fragile and vulnerable nature. There is a sense that without action in this area, future historians will be at a loss.

As one UK historian put it:

> '… Digitisation is very useful but the material that only exists in digital form; that is clearly disappearing. In a sense although the digital preservation of hard copy archives is useful, I think the bigger picture is being lost'.

It should be a concern then that few historians are engaged with the preservation of 'born digital' data, as they will be the records of the future. The necessary

reconsideration of how the historical art is practiced is not taking place universally or uniformly. Even if presevered, they are not likely to be contexualised, catalogued or organised adequately; curation is crucial.

Several respondents were concerned about the 'illusion of completeness' noted above:

> 'I'm more concerned about the born digital, there's very little capacity within the system to deal with it and there is the illusion if you can't find it on the web then it's not there'. (UK historian)

One specific issue of the changing digital environment which emerged from the survey is that of cloud computing; a trend which could have a significant impact on digital preservation and as such on the future of history. Naughton describes this as:

> 'a technology in which we use simple devices (mobile phones, low-power laptops or tablets) to access computing services that are provided by powerful servers somewhere on the net'.

and of its impact comments:

> 'this switch to computing as a utility rather than a service that you provide with your own equipment has profound implications for privacy, security and economic development – and public perceptions are lagging way behind the pace of development. What happens to your family's photo collection if it's held in the cloud and your password goes to the grave with you? And what about your documents and emails – all likewise stored in the cloud on someone else's server? Or your "reputation" on eBay? Everywhere one looks, the transition to cloud computing has profound implications, because it makes us more and more dependent on the net. And yet we're sleepwalking into this brave new world'.[24]

Several respondents had concerns about implications of the cloud for digital preservation and history: its implications are as yet not fully known, and it magnifies many well-known issues. Others, however, believed that it was the best long-term hope for preservation of digital materials. In general, there is agreement only that cloud computing is a good short term solution; whether longer-term is more a solution or a problem remains unclear.

THE EPHEMERAL WEB

The shift to a digital environment, and the dynamic and mutable nature of digital material, and in particular the ephemeral nature of web 2.0 data poses complex challenges for preservation. In 2010, the British Library estimated the average life expectancy of a website as less than 75 days, with at least 10% of UK websites lost or replaced with entirely new material every six months.[25]

As Uricchio points out,

> 'spread from person to person, with the always-present possibility of manipulation and mutability, they differ in the main from the relative stability and uniformity of the traditional fixed media… They can be apprehended, but the question is, at what point? What constitutes a sufficient 'capture' in a dynamic and fast evolving distributed network where any of the nodes is capable of change?'[26]

Other commentators like Hoorens and Rothenberg, (2008)[27] observe that archiving web content is like 'chasing a moving train', while Wolpert (2010)[28] asks how do you capture 'the river that flows?'

Examples of solutions have included: Brewster Kahle's Internet Archive and Wayback Machine, which captures the web, albeit in basic form, providing a resource for future, and indeed contemporary, historians; the archiving of Salman Rushdie's digital files by Emory University; the 9/11 Digital Archives (Cohen 2004), and the British Library's 'Email Britain' campaign of 2007, which asked members of the public to send an email on one particular day.

Perhaps most notable is the archiving of Twitter by the Library of Congress. Survey respondents were strongly supportive of this, describing it as "marvelous", "fantastic" and "hugely valuable", and allowing historians to "ask more questions, get more answers and make history more interesting". It was also seen as an opportunity to start exploring important questions about preservation: selection, access to large volumes of data, issues of privacy, etc. Yet again, there was some concern about the' illusion of completeness': it is important to be mindful that not everyone tweets.

The survey responses indicated that Web 2.0 material such as blogs, chat forums, status updates, YouTube videos or Tweets, as representations of 21st century life, are certainly worth preserving, particularly as they may provide a unique and deeper insight into historical events, especially for social and political history. They may offer the prospect of a more "democratic archive of the future", capturing the lives of ordinary people, and giving insights unobtainable from formal records. Of course, issues of privacy and intellectual property must be addressed, in the preservation of these kinds of sites.

The majority of respondents felt strongly about preserving the context (metadata) as well as the content, for at least a sample of material, as they can show social phenomena and the context of communication.The current legal framework is however, to a notable degree, obstructive.

Typically, current legislation does not enable easy preservation of digital materials and there is pressure for the current UK copyright legislation to catch up with the digital age.

Future historians face the difficulty of navigating the murky waters of ownership and authorship that so often pervade digital content. Kirschenbaum (2010) highlights an obstacle to interrogation of future history,

> 'before historians can apply any searching programs to collected emails and blog posts of future biography subjects, they must first obtain those emails. Since that data belongs to the companies controlling the email or social networking program, historians might find it difficult to gain access to the data. The biggest challenge to researchers of the future is not finally going to be technological in my opinion, but legal and social'.[29]

The legal framework is an important and urgent issue for digital preservation in a historical context. However the issue is an extremely complex one; an analysis is beyond the scope of this study. For further detail see Roland (2010)[30] and for a detailed international comparison see Besek, Coates, Fitzgerald, Mossink, LeFurgy, Muir, Raseberger and Weston (2008).[31]

SELECTION AND CURATION

The issues of how, and by whom, material is selected for preservation in a digital environment are clearly central. The survey respondents were clear that, in an age of information overload, not everything should be preserved, and selection policies must be transparent; yet it is difficult to determine what should **not** be preserved. There is a particular need to try to preserve the materials of the "unbelievers", who do not yet see the need for managing born-digital materials.

Their views expressed a clear desire for the selection of digital data to mirror the analogue world: there is no sudden change, just because the format of the material has evolved. Just as in the past, not everything can, or should, be preserved. And, just as in the past, there must be careful purposive selection, while allowing for some more speculative element:

> ‘The records of central government and administration has to be a priority. There’s also an argument that publicly funded projects should be preserved that if it’s deemed worthy of public funding then that needs to be kept’.
> (UK historian)

> ‘Fundamental records of society should be kept. As for the random, it’s more difficult, it’s difficult to know who will be the next James Joyce’.
> (UK digital humanist)

Diverse format types, other than text documents, numerical data and images, including such things as 3D engineering models, were also considered important for the future archive, as were such things as supermarket loyalty card data.

The importance of preserving a balanced record for future historians was emphasised by a number of interviewees. 'Unimportant' sites should be archived as well as the obviously significant, and an attempt made to give a balance between the "voice and the voiceless".

The majority of respondents were reluctant to identify data not worthy of preserving. This is perhaps unsurprising as it is impossible to know what will be of value to future historians:

> ‘The historical questions of today are going to be very different from the historical questions of the future which you can never second guess so to a certain extent you can never get it right’.
> (UK historian)

The selection, and subsequent curation, of digital material clearly requires the collaboration of historians and archivists; but the optimal role of each is unclear.

The responses to the survey indicate that the respondents believe strongly that archivists are best placed to have ultimate responsibility for selection decisions, rather than historians; a view shared even by a majority of historians. There is some active collaboration between the history profession and library and archive practice, in the UK particularly through the auspices of the British Library and the

National Archives. However, matters could be further improved, by greater dialogue between all concerned: historians, archivists, librarians, technologists, etc.

The survey results also suggest that historians need to think about their own digital research outputs, and individuals who are not connected to a network, or national agency, are more vulnerable to data loss. Perhaps there is opportunity for academic and national libraries to rise to the curation challenge by identifying individuals or owners of 'vulnerable' sites and provide a service to safeguard their material, or at least educate these individuals about data curation. Libraries should seize the opportunity to take a proactive rather than reactive stance in this sphere.

As noted above the UK does not have an official digital preservation strategy, so that allocation of responsibility is unclear. Survey respondents gave a clear view that responsibility for curation of digital data should fall on the institutions that currently steward analogue material; the state, for example, should retain responsibility for government data. Continuation and collaboration were strongly recommended.

Again, as noted above, the current UK legal system does not support digital preservation making it harder for memory institutions to capture this content; as the respondents commented, 21st century materials are being handled within a 20th century legal framework. The point was also made that corporate bodies should be responsible for long term preservation of their content, particularly when there are legal implications.

There were suggestions the British Library or the National Archives should, and probably will, take a leadership or advisory role in the stewardship of the digital realm.

A strong theme emerged that put responsibility in the hands of the creator of digital data, since they are often best placed to determine what should be preserved, and how. This will require considerable effort to be put into education and the raising of awareness; librarians and archivists are best placed to do this. Funding bodies are also in a position to encourage a mindset of responsibility and indeed they have an obligation to ensure the longevity of research if it has been financed by public money.

As stated by this UK librarian:

> 'There should be a responsibility on funders to put obligations on those who are receiving public money to manage and curate outputs in a satisfactory and sustainable way'.

It is perhaps surprising that only two respondents highlighted the role that technology companies play (or don't play) in digital preservation. It is generally agreed as essential that government information is safeguarded; perhaps there should be more onus on hardware and software suppliers to take a more active and responsible approach to preservation. For a government's actions to be interrogated in the future, records need to survive in legible form. A lack of

responsible action will inevitably affect the future archives of governments which will impact on our ability to hold them to account. Surely this will affect the future of evidence and therefore the future of democracy…

To raise awareness of the importance of digital curation perhaps libraries could collaborate with trade associations. Are architects using sufficient metadata to future-proof their plans? Do writers and authors realise that putting their material on a memory stick does not ensure their work will be accessible in fifty years time? Libraries should collaborate with the Royal Institute of British Architects, the Society of Authors, the Medical Research Council for example and educate all trade associations about digital curation – what it is and what it is not. Trade bodies can then promote the long term curation message to their memberships.

THE HISTORIAN'S CRAFT

Although there have been a number of studies on the information behaviour of historians (see, for example, Tibbo 2003,[32] Dalton and Charnigo 2004,[33] Duff, Craig and Cherry 2004,[34] Anderson 2004[35] and Maxwell 2009[36]), there has been little published research specifically concerning historians' views on digital preservation. It is clear, however, that the practice of history is likely to be influenced by these issues, and specifically by the relationship between the historical and archive professions.

Historians have traditionally been able to rely on archives, libraries and museums to research the past. For centuries these 'memory institutions' have safeguarded our national and cultural heritage, our collective store of knowledge, as a legal and moral duty. Yet this system is in a state of flux and so responsibility for preserving much of the digital material created in the 21st century is undefined.

Indeed the literature indicates the relationship between archivists and historians is a long standing and somewhat complex one. Cook (2009), for example observes that the two professions have experienced a distancing in recent years. So much so that 'the archive(s) has become a foreign country for historians'. He notes an aura of 'silence' and 'invisibility' surrounding the archivist and that in order to maintain an objective stance, historians in a sense are, 'blind' towards the archivist's powerful 'co-creator' role, and argues that there is a need to 'break the harmful silence between historians and archivists'.[37]

From a UK perspective the literature suggests there has been little collaboration between the two professions, and this was confirmed in this study by comments from the UK National Archives, suggesting that historians had been rather passive, and reticent in engaging in planning for the future.

As noted earlier, historians seem focussed on digitising existing records, i.e. the records of the past (for the present) rather than the records of the present, for the future. In a sense this does not seem odd; the nature of their craft is to reflect on and analyse the past. Indeed many digitisation projects have revolutionised the history profession and enabled a more democratic history, a history from below. Yet it is evident that born digital material will largely be the records of the future and are most vulnerable to loss.

Indeed Rosenzweig (2003) cautions, 'historians ignore the future of digital data at their own peril' and calls upon historians to have dialogue with memory institutions, helping inform their preservation decisions declaring,

> 'for the foreseeable future, librarians and archivists will be making decisions about priorities in digital preservation. Historians should be at the table when those decisions are made'.[38]

Of course, who knows what will be of interest to future historians? One historian's chaff is another historian's wheat. This raises compelling questions about selection and appraisal of digital data.

Many commentators note it is very difficult to erase the digital footprint; see, for example, -Schonberger (2010)[39] and Naughton (2010)[40]. Yet equally there are the much publicised cases of "source remorse"[41] where an individual regrets a comment or photograph published online and requests web content be removed. Indeed this apparent inability to erase digital fragments of 21st century life is seeing the rise of 'digital reputation managers', an emerging industry whose job it is to optimise the good and 'hide' the bad. Surely this will impact on the future historical account?

Yet for those concerned with the long term access to and preservation of digital data, there is still the problem with the ease with which much digital material can be deleted or wiped from the historical record. This creates problems for establishing provenance. There is clearly a tension between long-term and short-term persistence of digital data. As Smith (2002) put it, 'digital systems are sprinters, whereas history is a marathon runner'.[42]

The shift from a 'filter then publish' to 'publish then filter' system (Shirky, 2003)[43] poses a rather novel challenge for historians of the future (arguably the not too distant future). How will they find information in an information abundant age? The digital age, according to Rosenzweig:

> **'**seems likely to confront historians–who were more likely in the past to worry about the scarcity of surviving evidence from the past–with a new "problem" of abundance'.[44]

He suggests they may need to learn new skills such as data mining and urges historians to be more proactive in addressing the challenges of information abundance. Survey respondents generally agreed with this, but suggested that this would best be done by collaboration with other professions – in particular the library/information sector - rather than by changing the historical skill set.

It seems clear that the historian's craft is changing, with an increasingly digital environment bringing needs for new technical skills, and for appreciation of issues such as metadata. One respondent cautioned that "imperfect retrieval" could lead to "a form of censorship" and that users of online tools will need better understanding of the nature of online material. For more information on how the online environment could actually limit our world view see work by Eli Pariser (2011)[45].

Respondents were also asked to comment on the impact of Rosenzweig's "age of abundance" on the historian's craft, and it was generally agreed that this is a major issue:

> 'Historians of the digital age may find that everything is relevant, or nothing, and that the sheer volume of opinion drowns out the voices that they are looking for, especially if those voices are marginalized'.
> (UK historian)

There was a view that material which was not digitized, or even that which did not appear on the first page of a Google search or for which the link was broken, would be discounted.

Interviewees were also asked about the effect of technology on the methodology of doing history. Several expressed concern that historical research and study in the digital age may lead to a skewed history, and worried about the dangers of not understanding the mechanisms of the digital environment:

> 'It has certainly affected my methodology. I don't go to archives very much. I do my research from my study here in Cambridge. I use different sorts of material that I would never have dreamt of using 10 years ago. I don't go to a library to read books, I Google them. The downside of this is I'm conscious that I don't spend enough time in the archives. I tend to go into datasets online like the Old Bailey Online and so there's a danger that the sort of history I write becomes determined by the sort of public data sets I can get hold of. The fine grained historical research that used to go on… there might not be as much of it anymore. So the potential for the sort of history I research to be skewed by what is preserved digitally is possible yes'. (UK historian)

Establishing provenance is, of course, crucial to the historian. Yet in the digital age when material is regularly changed, updated or removed this seems to be an even more intricate and complex task. As Auchard points out,

> 'for all its openness, the web has proven to be a leaky vessel for historical preservation, with much of its treasure trove lost in a maze of altered web pages, broken links and deleted sites'.[46]

When asked about this issue, interviewees were unanimous about its importance and complexity. Provenance will retain, or even increase its importance, but may be increasingly difficult to resolve: skills of "digital archaeology" will become necessary. Time-stamp digital signatures, such as those developed by the British Library, may be the best way to ensure an understanding of when and how a digital object was entered into a collection.

Some commentators have suggested that historians need to rethink the definition of an historical record in the digital age. The UK National Archives state,

> 'technology has made the traditional concept of the 'record' more complex as new channels and tools for recording and communicating information proliferate.'[47]

Indeed in the digital age a historical record is more than a written journal or music sheet - surely President Barack Obama's winning election tweet qualifies as an historic record. Reflecting on the public response to the 9/11 tragedy in America, Cohen observed,

> 'the nature of the historical record had changed in many ways… Far more expansively, the record of 9/11 was to be found in new media such as

> websites, email, and other forms of electronic communication and expression, forms that have become an increasingly significant part of America's and the industrialized world's cultural output'.[48]

However, the majority of survey respondents did not agree historians need to redefine a historical record stating 'no', 'not really' or it's 'already in hand', or even that "archivists waste time discussing this issue". It is all evidence for the historian, even if presented in new forms of records.

However, there were some dissenting voices, suggesting that 'records' must now be understood in a much broader way, and that traditional categorisations – primary and secondary, published and unpublished – are breaking down.

This seems to be a significant issue, which deserves further investigation.

A 'DIGITAL BLACK HOLE'?

As noted above, there is a polarised debate between commentators such as Brindley on one side and Harvey on the other, as to whether the idea of the dramatically stated 'digital black hole' has any validity.

The survey responses suggest that this group consider that while this is an effective slogan to raise awareness of digital preservation issues, the reality is not so stark. There is a 'gap' in digital knowledge, and it will impact on the future of history; but there have always been gaps in the historical record, and the digital gap will be at least partly covered by printed materials. Rather than one big black hole, there are more likely to be scattered gaps in knowledge.

> 'There is no digital black hole. This is exaggerated. The important stuff may be lost per se; but there will be other evidence for them. It may not be complete and it may not be perfect but it will allow an interpretation to be made'.
> (UK data archivist)

Several respondents point out that digital data cannot be expected to survive surreptitiously and that active intervention is required: appropriate hardware and software (as well as a reliable electrical supply), in support of an appropriate preservation management strategy.

Indeed this theme, that digital data requires a more determined approach, is expressed clearly in the literature. Smith declares:

> 'long term preservation of digital data requires "up-front planning" and "as frustrating as it may be, we cannot avoid trying in the present to determine the future values in digital content'.[49] While Wolpert points out, 'persistence in the digital world does not happen by luck but through intentional action and explicit investment. The odds that bits will survive in a shoebox in an attic are pretty small'.[50]

It was pointed out that 'casual' communication, such as email and text messaging, creates casual administration, with data in these forms more likely to be lost to future historians without clearly stated, and effectively enacted, policies. Indeed the recent phone hacking scandal highlights the necessity for clear curation strategies, particularly regarding email communication.[51]

The 'digital divide' is seen by some respondents as a further issue. Despite the proliferation of mobile devices in particular, there is still a substantial community which is not online in any form. Archivists and librarians can take a wider role in preserving the 'silent voices' of the digitally excluded.

This study illuminates a paradox in the digital age. Among all the 'noise' - blogs, emails, status updates, chat forums, Tweets - there is also much silence. In the UK alone there are 9m digitally excluded adults. It is therefore important to consider the unheard voices and the untold stories. There is clearly an opportunity for the LIS profession to capture this silence.

Interviewees were asked if they had encountered examples of the loss of critical data, with severe consequences. Several cited instances of regrettable data loss, others remarked on the significant cost and time attached to data recovery while still others pointed out the loss of opportunity that can be a consequence of data loss. Nevertheless, none of the respondents were able to demonstrate an instance of truly critical digital data loss, so that this study is unable to add another case study to the limited number the digital preservation community relies on.

CONCLUSIONS

The meaningful preservation of digital information will determine the stories future historians will tell (or not), the information they will access (or not) and the knowledge available for future generations to build upon, (or not.)

Digital curation, or the lack of, will impact upon the future of every profession, not just history. It will influence the future of government records, business records, research data and indeed the future of our own personal archives. The issue is not just about preserving historical narratives or safeguarding our national digital heritage. Effective curation of digital data will also help safeguard our ability to hold governments, institutions, corporations or individuals to account. Ultimately it's about preserving access to the truth, (or at least a version of the truth.) It is therefore vital to highlight the issue and emphasise education.

Indeed all respondents were unanimous on the importance of education and believed that more education and training on these issues was paramount, beginning in schools, and not restricted to technical areas: it should form part of the learning of literacy or life skills.

Respondents were also unanimous on the need for the adoption of and implementation of explicit digital preservation policies:

> 'If they don't have one then they're crazy. Or at least they should have a policy that includes digital preservation…. But the key word here is implementation – it needs to be done'.
> (UK digital archivist)

The future of history will be determined by the decisions made today – legal solutions, the successful implementation of policies, education and raising awareness. It will also be determined by the level of collaborative dialogue, vision, and inclusiveness. This study has highlighted some opportunities for the struggling library sector to become more relevant in the 21st Century.

Asked for additional suggestions on how to ensure the continuity of access to digital data for future historians, most interviewee responses were related to amplification of issues raised earlier: technical, economic, policy and legal. Four novel and interesting comments were made:

> 'The key thing is trying to get people aware of it as a discussion and perhaps more aware of their own behaviour. I'm as guilty as the next person of deleting things, trying to tidy up my in-box and so on'.
> (UK historian)

> 'We should be encouraging historians to think differently about what stuff is being created today that is different from the past. There's less worry about the continuation of the paper stuff – there are people ensuring that will survive. It's the entirely new forms of information that is emerging, the social networking stuff. They are yet to wake up to the fact that this will be of value to them. But I think more historians are responding to that now'.

(UK digital curator)

'… nearly everything ever thought, spoken or written has been lost, including the true reasons for most historical events. So perhaps history is bunk, but it's important bunk. Over the next 50 years or so a much larger proportion of what we think, say, do and write will be accessible to historians, and this probably will change the game much more than the nature of the losses'.
(UK data curation consultant)

'My main plea is for someone to look at it more strategically; at a national and international level. There needs to be a much more strategic approach. Who would do it? The British Library and The National Archives should carry on leading the way and researchers should get a bit more clued up about it as well and feed in what they'd like to happen'.
(UK historian)

This alerts us to the wide scope of the digital preservation issue: from international strategies, through education at all levels, to individual awareness and behaviour, and with the nature of history changed for the positive more than the negative. While there may be no single digital black hole, there is sufficient reason for concern, for active planning and implementation, and for collaboration between historians and the information professions. In light of the recent uprising in the Middle East[52] historians of the future will no doubt have a richer view of the past because the Library of Congress had the vision to archive Twitter.

So the paradoxical question to consider is, should historians of the present contemplate the future in order to ensure access to an expansive and rich record of the past? Arguably a myopic outlook will limit the future of history, and therefore the future of knowledge. Ultimately, while it is impossible to know the needs of future historians, educated guesses and an imaginative outlook will help ensure the future historical record is a rich one.

References/Notes

[1] Ross, Seamus (2000) 'Changing Trains at Wigan: Digital Preservation and the Future of Scholarship', p24. HATII, University of Glasgow
Available at: http://www.bl.uk/blpac/pdf/wigan.pdf
Cached version: http://www.webcitation.org/5vC9RzSSq

[2] Brindley, L. (2009a) 'Filling in a Digital Black Hole'. British Library. Available at: http://www.bl.uk/pdf/digitalblackhole.pdf
Cached version: http://www.webcitation.org/5vC6jiOcG

[3] Tredinnick, L. (2011), Rewriting history: the information age and the knowable past, in Weller, T. (ed), Information history in the modern world, Basingstoke: Palgrave MacMillan, pp 175-198

[4] Roland, L. (2010), *The future of history: preservation in the digital age and the implications for historical research and study*, unpublished MSc dissertation, Department of Information Science, City University London

[55] Beagrie, N., & Jones, M. (2001). Preservation management of digital materials: a handbook. London: British Library.

[6] Conway, P (1990). "Archival Preservation in a Nationwide Context", American Archivist, 53, No. 2: 204-22.

[7] Brand, S. (2003) 'Escaping The Digital Dark Age'. Library Journal vol. 124. Issue 2, p46-49.
Available at: http://www.rense.com/general38/escap.htm
Cached version: http://www.webcitation.org/5vC6ftQRt

[8] Bailey, C.W. (2011), Digital curation and preservation bibliography, Version 2, Houston TX: Digital Scholarship,
available at: http://www.digital-scholarship.org/dcpb/
Cached version: http://www.webcitation.org/5ytfK22cY

[9] Levi, Y (2008) 'Digital Preservation: An Ever-Growing Challenge'. The ex-Librian Newsletter, July 2008.
Available at: http://www.exlibrisgroup.com/?catid=%7B5F787D34-4B0F-400C-B18B-E198197AF4A5%7D
Cached version: http://www.webcitation.org/5vC8ovsT3

[10] Hillis, Danny (2003) quoted in 'Escaping The Digital Dark Age'. Stewart Brand. Library Journal vol. 124. no 2, p46-49. Available at: http://www.rense.com/general38/escap.htm
Cached version: http://www.webcitation.org/5vC8IsPhL

[11] Nath,. C. (2010) 'Digital preservation' Parliamentary Office of Science and Technology. Four page POSTnote, spring 2010. Available at: http://www.parliament.uk/mps-lords-and-offices/offices/bicameral/post/current/physics/
Cached version: http://www.webcitation.org/5vC968iTZ

[12] Digital Preservation Coalition (2006), Mind the Gap Report. DPC. Available at: http://www.dpconline.org/advocacy/mind-the-gap
Cached version: http://www.webcitation.org/5vC7cGTKq

[13]Waters, D. and Garrett, J. (1996) Report of the Task Force on Archiving of Digital Materials. The Commission of Preservation and Access and Research Libraries Group. Available at: http://www.clir.org/pubs/reports/pub63watersgarrett.pdf
Cached version: http://www.webcitation.org/5vC9sDPA2

[14] Harvey, R. (2008) So where is the black hole in our collective memory? Charles Sturt University. A DigitalPreservationEurope (DPE) position paper. Available at: http://www.digitalpreservationeurope.eu/publications/position/Ross_Harvey_black_hole_PPP.pdf
Cached version: http://www.webcitation.org/5ytfPnENW

[15] Brindley, L. (2009b) 'Copyright for Education and Research'. British Library. Available at: http://pressandpolicy.bl.uk/imagelibrary/downloadMedia.ashx?MediaDetailsID=563
Cached version: http://www.webcitation.org/5vC71E3iU

[16] Lavoie, B. and Dempsey, L. (2004) 'Thirteen ways of looking at… digital preservation', D-Lib Magazine, Volume 10 Number 7/8. Available at: http://www.dlib.org/dlib/july04/lavoie/07lavoie.html Accessed on:
Cached version: http://www.webcitation.org/5vC8m1n24

[17] Smith Rumsey, A et al (2010) 'Sustainable economics for a digital planet: Ensuring long-term access to digital information'. Blue Ribbon Task Force on sustainable digital access and preservation. Available at: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
Cached version: http://www.webcitation.org/5vC9d2H4B

[18] Sierman, B. (2009) "The Jigsaw Puzzle of Digital Preservation — an Overview". Liber Quarterly vol 19, iss 1, pp 13-21
Available at: http://liber.library.uu.nl/publish/articles/000262/article.pdf
Cached version: http://www.webcitation.org/5vC9Ufg7B

[19] PLANETS (2010) 'The Digital Divide: Assessing organisations' preparation for digital preservation'.
Available at: http://www.planets-project.eu/docs/reports/planets-market-survey-white-paper.pdf
Cached version: http://www.webcitation.org/5vC9ILpXI

[20] Rosenzweig, R (2003) 'Scarcity or Abundance? Preserving the Past in the Digital Age'. American Historical Review 108, 3, pp 735-762. Available at: http://chnm.gmu.edu/digitalhistory/links/pdf/introduction/0.6b.pdf
Cached version: http://www.webcitation.org/5vC9OqvtD

[21] Hampshire, E. and Johnson, V. (2009) 'The Digital World and the Future of Historical Research'. Twentieth Century British Research, vol. 20, iss. 3, pp 396-414

[22] Rosenzweig, R (2003) 'Scarcity or Abundance? Preserving the Past in the Digital Age'. American Historical Review 108, 3, pp 735-762. Available at: http://chnm.gmu.edu/digitalhistory/links/pdf/introduction/0.6b.pdf
Cached version: http://www.webcitation.org/5vC9OqvtD

[23] Leary, P. (2010) Keynote presentation at 'Digitised History' Conference, London: British Library

[24] Naughton, J. (2010) 'The internet: everything you ever need to know'. Quoted in The Observer. Available at: http://www.guardian.co.uk/technology/2010/jun/20/internet-everything-need-to-know
Cached version: http://www.webcitation.org/5vC99B1UF

[25] British Library (2010) 'UK Web Archiving'.
Available at: http://britishlibrary.typepad.co.uk/inthroughtheoutfield/2010/02/lynne-brindley-appeals-for-uk-web-archiving-.html
Cached version: http://www.webcitation.org/5vC7IFpMA

[26] Uricchio, W. (2007) "Moving beyond the artifact: lessons from participatory culture," in Yola de Lusenet and Vincent Wintermans, eds., Preserving the Digital Heritage (The Hague: Netherlands National Commission for UNESCO and the European Commission on Preservation and Access, Amsterdam.
Available at:
http://web.mit.edu/uricchio/Public/pdfs/pdfs/moving%20beyond%20the%20artifact.pdf
Cached version: http://www.webcitation.org/5vC9pcxFW

[27] Hoorens, S. and Rothenberg, J. (2008) 'The Uncertain Future of Saving the Past'
http://www.rand.org/pubs/research_briefs/2008/RAND_RB9331.pdf
Cached version: http://www.webcitation.org/5vC9gRLoB

[28] Wolpert, J.A. (2010) 'Archives and History'. MIT podcast. Available at:
http://mitworld.mit.edu/video/674/
Cached version: http://www.webcitation.org/5vC83zY6b

[29] Kirschenbaum, M. (2010) quoted in ''Digital Age Presents New Problems for Historians'. Fox, S., Tech News Daily. Available at: http://www.technewsdaily.com/information-overload-digital-age-presents-new-problems-for-historians-0867/
Cached version: http://www.webcitation.org/5vC8ibrFd

[30] Roland, L. (2010), *The future of history: preservation in the digital age and the implications for historical research and study*, unpublished MSc dissertation, Department of Information Science, City University London

[31] Besek, J.M, Coates, J, Fitzgerald, B, Mossink, W, LeFurgy, W.G, Muir, A, Rasenberger, M, and Weston, C.D (2008) 'Digital Preservation and copyright: An international study'. The International Journal of Digital Curation Issue 2, Volume 3. Available at:
http://www.digitalpreservation.gov/library/resources/pubs/docs/digital_preservation_final_report2008.pdf
Cached version: http://www.webcitation.org/5vC6cMnCd

[32] Tibbo, H.R. (2003) 'Primarily History in America: How U.S. Historians Search for Primary Materials at the Dawn of the Digital Age.' American Archivist 66, no. 1, pp 9-50

[33] Dalton, M.S. and and Charnigo L, (2004) 'Historians and their Information Sources.' College and Research Libraries 65, no. 5, pp 400-425

[34] Duff, W., Craig, B. and Cherry, J. (2004) 'Historians' Use of Archival Sources: Promises and Pitfalls of the Digital Age.' Public Historian 26, no. 2, pp 7-22

[35] Anderson, I. G, (2004) 'Are you being Served? Historians and the Search for Primary Sources', Archivaria, No 58
Available at: http://journals.sfu.ca/archivar/index.php/archivaria/article/view/12479/13592
Cached version: http://www.webcitation.org/5ytfX78F1

[36] Maxwell, A, (2010) "Digital archives and history research: feedback from an end-user", Library Review, Vol. 59 Iss: 1, pp .24 - 39

[37] Cook, T. (2009) 'The Archive(s) Is a Foreign Country: Historians, Archivists, and the Changing Archival Landscape'. Canadian Historical Review, vol. 90, no 3, pp 497-534

[38] Rosenzweig, R (2003) 'Scarcity or Abundance? Preserving the Past in the Digital Age'. American Historical Review 108, 3, pp 735-762. Available at:
http://chnm.gmu.edu/digitalhistory/links/pdf/introduction/0.6b.pdf
Cached version: http://www.webcitation.org/5vC9OqvtD

[39] Mayer-Schonberger, V (2010) 'Delete. The virtue of forgetting in the digital age', Princeton NJ: Princeton University Press.

[40] Naughton, J. (2010) 'The internet: everything you ever need to know'. Quoted in The Observer. Available at: http://www.guardian.co.uk/technology/2010/jun/20/internet-everything-need-to-know
Cached version: http://www.webcitation.org/5vC99B1UF

[41] English, K. (2009) 'The longtail of news: To unpublish or not to unpublish'. The Toronto Star. Available at:
http://www.journalismproject.ca/en/attachments/Long%20Tail%20report_Kathy_English.pdf
Cached version: http://www.webcitation.org/5vC7jD858

[42] Smith, B. (2002) 'Preserving Tomorrow's Memory'. Newsletter of the IFLA Core Activity on Preservation and Conservation. No. 29, pp 3-10

[43] Shirky, C. (2003) 'Clay Shirky's writings about the internet'. Available at:
http://www.shirky.com/writings/broadcast_and_community.html
Cached version: http://www.webcitation.org/5vC9YCR9n

[44] Rosenzweig, R (2003) 'Scarcity or Abundance? Preserving the Past in the Digital Age'. American Historical Review 108, 3, pp 735-762. Available at:
http://chnm.gmu.edu/digitalhistory/links/pdf/introduction/0.6b.pdf
Cached version: http://www.webcitation.org/5vC9OqvtD

[45] Pariser, E (2011) 'Beware online filter bubbles', Ted Talk, March 2011
http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.html
Cached version: http://www.webcitation.org/613H6uLRX

[46] Auchard, E. (2009) 'The black hole. How the web devours history'. Reuters.
Available at: http://www.reuters.com/article/idUSTRE51Q3KE20090227
Cached version: http://www.webcitation.org/5vC62DCv2

[47] The National Archives (2009) 'Archives for the 21st Century'. London: The National Archives, Available at: http://www.nationalarchives.gov.uk/documents/archives-for-the-21st-century.pdf
Cached version: http://www.webcitation.org/5vC8x7uVR

[48] Cohen, D (2005) 'The Future of Preserving the Past" CRM Journal, vol. 2, iss. 2, pp 6-19
Available at: http://www.dancohen.org/files/future_of_preserving_the_past.pdf
Cached version: http://www.webcitation.org/5vC7SCOOo

[49] Smith, A. (2007) 'Valuing preservation'. Library Trends, Vol. 56, No. 1, pp. 4-25

[50] Wolpert, J.A. (2010) 'Archives and History'. MIT podcast. Available at:
http://mitworld.mit.edu/video/674/
Cached version: http://www.webcitation.org/5vC83zY6b

[51] Wikipedia (2011) 'News International phone hacking scandal'
http://en.wikipedia.org/wiki/News_International_phone_hacking_scandal
Cached version: http://www.webcitation.org/61KPfWnHz

[52] Channel 4 (2011) Special Report. 'Arab Revolt: Middle East Uprisings'
http://www.channel4.com/news/arab-revolt-middle-east-uprisings
Cached version: http://www.webcitation.org/61KPcqISh