The Challenges of Web Access to Archival Oral History in Britain

Rob Perks, Curator of Oral History, British Library Sound Archive, London & Visiting Professor in Oral History, University of Huddersfield

We have arrived at a turning point in the development of oral history where access to and control of the interview data we create and collect are creeping out of our ownership and control. We have long debated the nature of that ‘ownership’: recognising that interviews are a co-construction in which ‘shared authority’ is important; and emphasising the need for informed consent and an ethical and legal framework delineating copyright ownership, usage and access conditions. But until recently we - as archivists and academics and ‘keepers’ of the material – have more or less continued to exercise control over access to those interviews. Having decided what to ‘preserve for posterity’ we have largely determined who we will allow into public reading rooms and libraries; we have chosen which extracts from the interview data we wish to present in a lecture or publish in a book or broadcast as a radio programme. Secondary users or readers who have managed to negotiate access to the original data sources have then often been faced with a host of restrictions on the use they can make of the data: through poor or non-existent catalogues; inadequate content documentation; restrictions on copying transcript extracts; the high cost and degree of difficulty in obtaining copies of recordings for analysis back at base; tortuous copyright clearances prior to published use; a risk-averse public sector culture, and so on. In short we have not made it easy for people to get at our oral history holdings, and we have frequently been doing this in the name of our interviewees, to protect them, even though it is rarely clear what we are protecting them from and there are virtually no reported cases of abuse of access to oral history materials for undesirable ends. And we could, of course, carry on in this mode. But if we do we would be alienating new generations of potential users of our collections and, as funding tends to follow users, we would be threatening our own existence as holders of oral history data. Comprehensive web access is unstoppable and we ignore it at our peril.

Yet access to archived oral history data via the web completely transforms this guardianship relationship, which naturally makes us very uncomfortable as creators and curators, partly because we fear a marginalisation of our own role as conduit, explainer and interpreter. Online access to oral history interviews in their entirety sidesteps the archivist-gatekeeper and we now agonise about what we can do to reintroduce our presence through contextual information, site user agreements, copyright warnings, passworded access and other registration procedures. Not that I am necessarily arguing against any of these things, merely urging us to reflect on our own motives. For if we are serious about welcoming the web as an unprecedented democratic opportunity for disseminating hidden voices, challenging stereotypical historical views, and empowering people to share their own experiences with new worldwide audiences, regardless of income, mobility and expertise, why are we

29 This article is based on a paper first presented as part of a panel about web access to oral history at the International Oral History Association conference in Sydney in 2006: see www.une.edu.au/ioha2006/conference%20program.html [all cited websites accessed January 2009]. I have since benefited from useful conversations with Karen Brewster, Sherna Gluck, Kevin Bradley and Mike Frisch. Thanks also to Peter Findlay, ASR Project Manager, and Janet Topp Fargion for their helpful comments. I use the word ‘web’ in this article as shorthand for the World Wide Web and the Internet.

attempting to reintroduce barriers to access? Why are we hesitant? We evoke interviewee uncertainty about web access, but how much do we actually know about interviewee understanding of the web? We think of the web as providing a virtual version of the access we provide on-site and tend too often to think in a rather linear way: how much have we reflected on the new creative uses that can be made of that data, through powerful search engines, linkages, and the blending of related images and textual sources? And yet, perhaps caution is required: how much do we actually know about online users and what they want? More crucially, what do we know about whether mass access via the web will change the nature of the interview relationship itself?

What struck me most forcibly when we, at the British Library Sound Archive, embarked on our own first mass digitisation and web-access project six years ago, was how little reflective writing and thinking there was at that time on oral history and the internet. In 2009 the situation has not changed a great deal: the skills and knowledge base amongst oral historians and heritage professionals around digitisation and web access remains woefully poor. Standards have been slow to emerge and what little agreed best-practice there is relating to consent issues, retrospective renegotiation of rights, site-user agreements and so on, has been remarkably little-debated. A few pioneers – such as the Jukebox project at the University of Alaska Fairbanks and Sherna Gluck's Virtual Oral/Aural History Archive – stand out but otherwise the debate has hardly begun. In getting to grips with the key issues it seemed to me that we needed to better assess the web presence for oral history in Britain, so in the summer of 2005 we worked with a postgraduate student in archives and records management at University College London to carry out the first-ever survey.

Survey of web-based oral history in the UK

The survey found 265 UK-based websites presenting, promoting or giving access to oral history of which 36% were local community groups, 12% were museums, 12% national bodies and special interest groups, 11% from higher education, 7% archives and record offices, 5% corporate/business, 3% libraries (including the British Library) and 3% schools. This broadly reflects known key areas of oral history activity in the UK, and the survey noted that very few of these sites were dedicated only to oral history: oral history frequently featured alongside several other services or activities. More surprising was the finding that only 85 of the 265 sites had any actual oral history content (audio or transcript), the remainder merely describing oral history projects or collections.

32 http://uaf-db.uaf.edu/Jukebox/PJWeb/pjhome.htm and http://salticid.nmc.csulb.edu/cgi-bin/WebObjects/OralAural.woa/
33 Heather Macleod, ‘Web-based oral history in the UK: a study’, MA dissertation, University College London, September 2005. The websites she studied were accessed in June-August 2005. Oral history was defined as involving the spoken word: projects only concentrating on written reminiscences were excluded. I am grateful to Heather, not only for all her hard work on the survey but also for permission to quote from her thesis.
34 As 70% of the sites surveyed had no indication of when they had been created or last updated it is impossible to date the currency of information on every site. It is fair to assume that many lack sustainable resources for regular updating.
35 By contrast Karen Brewster’s survey of US oral history sites found that over half were university-based: Karen Brewster, ‘Internet access to oral history recordings: finding the issues’, 2000, ‘Research Results’ at www.uaf.edu/library/oralhistory/brewster1/index.html.
Only fifty sites had any audio content of any kind and, amazingly, only three sites had any full interview transcripts. In other words, the vast majority of oral history website creators were not seeking to provide access to their holdings – only 33 sites had any kind of online searchable catalogue or interview listing – but were using the web as a showcase or information board for their work. This reflects research in the US by Mary Larson and Karen Brewster which came to similar conclusions. 36

When it came to legal and ethical considerations, the survey's findings were worrying. Only 38 of the 265 sites mentioned copyright at all, despite decades of awareness-raising by the Oral History Society and others, and only nine of these gave a detailed explanation of copyright procedures. Only four sites with oral history content had addressed ethical issues through contextualising information, site-user agreements and/or passworded access, and might be regarded as models of best-practice. 37 The study concluded by calling for improved awareness of copyright and ethical issues amongst web creators whilst noting that the worst transgressors of best practice appeared to be community oral history groups, perhaps through ignorance, perhaps because copyright and consent wasn’t considered to be an issue where the interviewees themselves created the site!

The survey came to some other interesting conclusions: that there was an apparent shortage of digital technical expertise amongst heritage professionals, and a concomitant deficiency in training provision in areas such as audio digitisation and editing, and web authoring. We might add to this more practical problems such as lack of finance, shortage of server space and a frequent lack of understanding amongst IT staff about the particular requirements for online audio delivery (‘you need how much memory?!’), which has often forced oral history online projects down the priority list. All this might help to explain that, whilst 265 oral history websites in the UK sounds a large number, very few indeed actually provide an online resource involving access to archived oral histories. And some sites, knowing the popularity of oral history amongst web visitors and its potential for profile-raising, appear content to headline oral history as a point of interest, rather than attempting to provide online access. This leads me to a reflection on our own oral history web projects at the British Library.

The Way We Speak project 38

For some years the British Library Sound Archive has provided online access to catalogue data about its oral history holdings, now totalling over 300 collections and some 55,000 recordings, 39 but only a handful of audio extracts have been accessible online. 40 In 2003 a £3.25m government grant was awarded to the BL to digitise and make accessible images and sounds, from manuscripts to maps, music to wildlife recordings, all selected on the basis that they relate to ideas of ‘place’.

37 These included www.museumoflondon.org.uk/MOLsite/londonsvoices/ which was established as an online resource.
39 www.cadensa.bl.uk
40 http://www.bl.uk/reshelp/bldept/soundarch/about/soundarchive.html
In March 2005 a new web-based resource was launched aimed at bringing the BL’s collections into the homes and schools of a broader set of public users than it usually reaches. Included amongst 90,000 images and sounds on the site were 681 oral history extracts (fifty-five hours in total), entitled The Way We Speak, reflecting how English accents and dialects (sounds, words and structures) have changed since the Second World War, and how notions and understandings of place and community have shifted during that period. The two comparative audio collections, recorded fifty years apart and held by the British Library Sound Archive were selected for the project: the Survey of English Dialects and the Millennium Memory Bank.

The Survey of English Dialects (SED) is still one of the only systematic surveys to have been carried out into the dialects of England. Between 1950 and 1961 a team at Leeds University made audio recordings in 304 locations in England, almost all rural communities where it was felt that traditional dialect had been best preserved. The recordings varied in length from ten to thirty minutes, covering a myriad of topics but often connected with the speaker’s occupation, for example ploughing, harvesting, hedging, pig-killing or breadmaking. The second archival collection chosen was the Millennium Memory Bank (MMB), a joint project between BBC Local Radio and the British Library Sound Archive to create an archival ‘snapshot’ of ‘ordinary’ Britons’ opinion and experience at the turn of the century. During 1998 and 1999 forty BBC local radio stations across the UK gathered 5429 oral history interviews on minidisc, now catalogued and archived at the BL. MMB is one of the largest single oral history collections in Europe, and a unique study of Britain at one particular moment in its history.

For the web project the SED recordings were used as a baseline and cross-matched to equivalent extracts from MMB interviews, and in addition a further 126 MMB audio extracts from key urban centres were added to bring the survey up-to-date and establish a new baseline for future work. After an initial audit to select the most suitable MMB recording, a great deal of attention was given to selecting an appropriate passage for digitisation from both collections. The crucial factor here was combining the occasionally competing aspects of historical content – an intrinsically interesting, coherent passage that made sense in isolation – and medium – an extract that contained linguistic features that enabled a listener to identify a particular speaker with a particular locality. In creating a web-based resource about ways of speaking, every effort was made to select an excerpt containing an interesting word, a revealing pronunciation pattern or a noteworthy grammatical construction.

Alongside the audio extracts, and in the absence of transcripts, new catalogue records were created to provide content descriptive information but also comparative data about lexis, phonology and grammar – the words, sounds and structure that define the speech of a given community. The aim was to provide online users with some context and detailed guidance.

41 The original CollectBritain website was migrated to the main BL website in late 2008. Content from The Way We Speak was moved to http://sounds.bl.uk/
42 Scotland, Wales, Ireland and the Channel Islands were excluded with the single exception of six sites in Monmouthshire (Wales). 313 sites were selected but no audio recordings were made for three sites in Yorkshire and five sites in Lancashire. Collection reference C908.
45 In addition to the archive the interviews formed the basis for 640 half-hour radio documentaries broadcast through the BBC’s local radio network during the final sixteen weeks of the last millennium. Collection reference C953.
concerning the specific linguistic features that identify a speaker with a particular part of the country. The website was searchable by place-name and subject, and a map interface simplified the user’s entry point. Most of the original SED recordings are significantly shorter than the MMB recordings but for each collection five to ten minute audio extracts of intrinsic social history interest were selected, topped and tailed via SoundForge digital editing software, then stored as both .wav files and converted to streamed Windows Media for web access. Audio streaming, rather than downloadable, software was chosen to prevent bulk theft and unauthorised reversioning of the audio extracts.

**Ethical issues (1)**

Mounting such large quantities of oral history recordings and personal data on the web for the first time naturally raised a number of ethical issues. We felt safe in assuming that the SED informants were no longer alive (virtually all were born before 1900, some in the 1870s) but copyright and clearance forms were not used in the 1950s and formal participant consent was not sought in writing, let alone for the kind of world-wide access to the recordings anticipated. Whilst in this case we judged we were acting legally, we debated whether we were acting ethically and whether consent from living relatives of SED contributors should be sought. After discussion with Leeds University (as the originator of the recordings) we took the view that, firstly, it would be difficult, in many cases impossible, to contact relatives; and secondly that all the participants had apparently been made aware of the various public outcomes that the SED team had envisaged, and had given their verbal consent to participate on that basis. Indeed a great deal of the material had already been published in book form. We were also using extracts, not complete recordings. To date we have not received any complaints or requests to remove any SED material from the site and in at least three cases we have been able to ‘reunite’ grandchildren with recordings of their late grandparents through the website and the publicity it has generated.

In the case of the MMB recordings the legal position was clearer as a comprehensive BBC clearance and copyright form had been used which assigned all interviewee rights for use in ‘all media now known or which may be developed in the future’; sought agreement that the recording could be edited, adapted or translated; and (more controversially) waived all moral rights. Yet, as with all such forms, a lingering ethical question remained about how far interviewees had fully understood the implications of web access to their interviews and were thus said to have given ‘informed consent’. In this case we were persuaded by arguments that they had all knowingly participated in a broadcast-based project, must therefore have understood the probability of their appearance on local and national radio, and that no further clearance was necessary. In terms of a website user agreement we decided that a simple copyright and user statement was sufficient, backed up with an extensive contextualising introductory essay with links to related sites and suggested reading, together with a glossary of terminology.

---

46 This feature was sadly not migrated with the rest of the content in 2008 (but see discussion below about Sounds Familiar: http://www.bl.uk/learning/langlit/sounds/index.html)


49 All recordings copyright © British Library Board. The recordings on this site are for private listening only. Copying, broadcasting or reproduction is prohibited. More recently we have also added a background section on the ‘Ethical use of oral history’ and a ‘Notice and Takedown Policy’.
User impact

The project did a great deal to reconnect the BL Sound Archive with a growing popular web-based interest in local history, accents and dialects. From the online launch of the first set of extracts featuring recordings from the north of England in February 2004, the level of international, national, regional, local press, radio and television coverage was extensive. When web hits are added (41,000 in the launch month) this represented an initial reach of many millions, and a notable aspect of the coverage and user interest was its diversity. New audiences – actors, teachers, local historians, casual web-browsers, school students - were reached, and the BL’s existing service to users enhanced. People ‘visited’ the BL online who had never been there before.

And usage remained high: putting the launch period aside, more typical monthly hits on the CollectBritain site are averaged 20,000, of which between 3000 and 4500 are accent and dialect-related, far outstripping the number of onsite users for similar collections. There is a lot we still don’t know about these new users – how they find the site, how long they spend, what their interests are, how they make use of the content - but measured in numbers alone it was a successful project.

Archival Sound Recordings (ASR) Project

By contrast to The Way We Speak, our second digitisation and web-access project set out not to present audio extracts as an interpreted resource, but to allow unmediated online archive access to whole interviews. With a government grant of £2.3m over six years (2004-9) 10,000 hours of BL Sound Archive recordings have been digitised for web access, including 3,400 hours of oral history interviews with artists, sculptors, photographers, architects, scientists and Jewish Holocaust survivors recorded over the past twenty years. 650 hours of interviews with jazz musicians were also included, plus 800 hours of other spoken word material. The original analogue cassette recordings have been digitised to .wav format (24 bit 48kHz) in up to thirty-minute segments, stored initially on DVD-Rs and more recently on hard-drives (using the Quadriga audio-archiving workstation) and versioned to both MP3 and Windows Media for access purposes. The new audio files were renamed and matched to metadata using METS (Metadata Encoding and Transmission Standard) based on extracting existing catalogue data in MARC format converted into Dublin Core in XML mark-up. METS also allows for preservation and structural metadata to be added in order to describe the complete digital object including audio, images, transcripts and other related materials.

---

50 It has included all the broadsheet newspapers, a full page in the Daily Mirror (1.9m readers), Channel Four News, CBBC Exchange, BBC Radios 2, 4 and 5 Live, ITV This Morning, 19 local radio stations and 20 regional newspapers.


52 From the Bow Dialogues (C812) and the Oral History of Recorded Sound (C90).


54 For enthusiasts of metadata discussion, Guy Marechal of Memnon Audio Archiving Services (Brussels), the ASR Project’s chosen digitisation contractor in its initial phase, based his presentation at the IASA Conference in Barcelona in September 2005 on the ASR project: “New technologies and models of dissemination”, ASR phase 2 brought digitisation in-house, utilising the Quadriga system. All the project metadata will be compliant with Protocol for Metadata Harvesting from the Open Archives Initiative ([http://www.openarchives.org/](http://www.openarchives.org/)).
With this second project full web access is limited to educational internet users via so-called ‘Shibboleth authentication’. Only those remote users with an ‘ac.uk’ email whose institution or employer has a licence with the British Library have access to the complete unedited sound recordings (a potential estimated student audience alone of 4.7 million). Again we chose streamed Windows Media to deliver the sound and registered users can download an unlimited number of recordings for later use. Transcripts, where they exist, are available in searchable .pdf format but there is currently no direct electronic link between the audio and text versions (largely due to lack of resources), though both are available simultaneously. There are also detailed content summaries for every item and, unlike the summaries on the current BL Sound Archive online catalogue, they are searchable via Google and other web search engines, giving the interviews a much higher web presence. These summaries are available to all web users, regardless of authentication.

**Ethical issues (2)**

Despite the limitations on online access, ethical issues have remained paramount. All closed and restricted interviews were excluded from the project at the outset and have not been digitised (which is a cause for concern from a conservation point of view in the long term). For the rest we reapproached every interviewee and, in the case of deceased interviewees without a clearance form, relatives. There were three categories of interviewee, each requiring a different personalised letter accompanied by a question and answer sheet about the project:

1. In cases where we had a clearance form mentioning internet access and assigning copyright to the BL, we wrote to the last known address informing the interviewee of our intentions and asking them to contact us with any objections.
2. Where clearance forms were missing we wrote asking for permission to include the interview in the project through completion of the missing form. We excluded from the project any that failed to reply and/or grant permission.
3. With the third category of interviewees, for whom we held ‘pre-internet’ clearance forms (those that had not mentioned the internet or any kind of worldwide remote access), we wrote explicitly drawing the new kind of access and use to their attention and inviting them to contact us with any questions or objections.

As we had expected, the response from interviewees was muted, but it was nonetheless a very useful exercise. Some contacted us asking for copies of their recordings; some requested we bring the recording up to date; and relatives wrote with news of the death of interviewees, enabling us to update catalogue records. Many were simply delighted that there was interest in their interviews, that we were keeping them in the picture, and emailed their support (‘it is all quite amazing!’). Only a handful objected to their interview going online. One photographer, also a teacher in higher education, wrote that ‘increased access makes me feel very uncomfortable...my interview drew out a very strong emotional response’.

---

55 Initially the interface is only available to the academic communities of the UK and conforms to British Library web usability guidelines, ensuring that users with physical disabilities can gain access to, and make full use of, the content. The implementation of Special Educational Needs Disability Access legislation is key to the notion of access for all, as is egf compliance. The interface is accessible by password on JANET, the UK’s education and research network.

56 Comprising 2,175,115 higher education students (Office of National Statistics, 2002/03), plus a further 2,602,600 further education students (Office of National Statistics, 2001/02).
Another commented that ‘there is far too much personal stuff in there about my children and named individuals’. The most reflective reply came from an artist who wrote: ‘I was told it was for a select and specialized audience and that was the way I gave the interview. Digitized is entirely different – I lose my privacy to anyone in the world to become a voyeur – I can’t possibly accept that...I’m afraid I see the hand of globalization – which is being pushed without reflection onto a wide group now searching the internet for a quick fix – rehashed as scholarship. I am against this. It is reckless.’

Such views were very much a minority but they helpfully reminded us to be vigilant of our own practice. It is an inescapable fact that the digitisation and web delivery agenda is at the forefront of thinking in most major archives and libraries: in fact most funding for digitisation of sound archive materials is well-spent for preservation reasons alone, but we should be under no illusions that it is the access implications of digitisation that is the key driver. Whether the web itself will continue to be as open, free and democratic as it is today indefinitely remains to be seen. Some concerns have recently been aired in the US about security service ‘fishing expeditions’ involving online searching of oral history archives in a quest for personal data. Such breaches of trust and original intention, if proven, are likely to make interviewees more nervous and circumspect about divulging sensitive or confidential information, even on a closed basis. Quite apart from undermining the entire research process, this might also lead to an understandable reluctance by interviewees to allow web access to their recordings. More hard information about alleged ‘abuses’ of online access is required before any of these fears can be confirmed, but in the meantime a lively debate about the secondary use of oral history and personal interview data has been developing in the social science community.

Since the ASR site went live in 2006 more content has gradually been added, more user licences have been signed and the entire new service is available to users onsite at the BL in London. User numbers are growing and stood at 1.7 million unique visitors at the end of 2008. Most actively debated internally has been whether or not a click-through site-user agreement is required for the site, in line with best-practice emerging in the US. In our own case BL web services staff felt it would discourage users and would anyway have limited impact. Instead a page of contextualising information has been added, including a section of the ‘Ethical use of oral history’, a suite of links to BL web policy on privacy, copyright and licensing, and a ‘Notice and Takedown’ procedure for handling requests to remove recordings (which has only been evoked once). Most importantly the ASR initiative has moved from the status of a discrete time-limited project to the main portal for all online web access to BL Sound Archive recordings, subsuming other sites such as CollectBritain. The implication of this is that the

---