

Barwick, Linda. (2004). Turning it all upside down . . . Imagining a distributed digital audiovisual archive. *Literary and Linguistic Computing*, 19(3), 253-263.

This is a postprint version, with the page numbers adjusted to match the published version.

## **Turning It All Upside Down... Imagining a distributed digital audiovisual archive<sup>1</sup>**

Linda Barwick

Department of Music, University of Sydney

### Abstract

What could and should be the relationship between research archives of endangered cultural heritage materials and the originating community? This paper argues that recent developments in distributed computing in a networked environment have allowed us to re-imagine this relationship in a way that profoundly changes the role of the archive and reinforces the desirability of establishing ongoing reciprocal relationships with cultural heritage communities. Some possibilities are suggested drawing from experience with PARADISEC (the Pacific and Regional Archive for Digital Sources in Endangered Cultures, established in 2003 as a collaborative venture between the University of Sydney, the University of Melbourne, and the Australian National University) and with local community-based digital archives in the remote Australian communities of Belyuen and Wadeye. Repatriation and rights, planning principles for establishment and sustainability of local digital archives in community cultural centres, and models for a staged approach in setting up ongoing relationships with rights holders are discussed. The paper argues that digital archives, as distributed virtual institutions, need to engage with a number of different communities of interest: not only the individuals, communities, and institutions that own the cultural heritage objects we preserve, but also the wider academic community and international standards-setting bodies. Planning for our archives' digital future means imagining ourselves as actors and creators within that virtual society.

When I was training as a researcher, archives were places of centralization. They centralized the objects, centralized their cataloguing and arrangement, and centralized access. In those days, you could only get access to archives by actually going there, usually down several flights of stairs—but only after having consulted numerous paper or card indexes in other parts of the library system that told you which object was located where, and under which collection number. Access took up a lot of the researcher's time and resources, and required specialist knowledge and training.

The ways in which ethnographic sound archives acquired material were similarly centralized and hierarchical. The sound archives' primary relationship was with the individual collector, who typically travelled to

---

<sup>1</sup> Presentation to DRH Conference, Cheltenham UK, 2 September 2003

remote places to collect the recordings for deposit in the archive. Relationships between the archives and the individuals whose speech or performances were recorded were typically limited by geography, technological differentials, and sometimes language barriers.

## 1 Imagining It Differently ...

Today the advent of digital archiving means that things have changed. While the process of digitization is considerably more costly for creation, indexing, and maintenance of the resource, access has become vastly easier and cheaper if the materials are made discoverable or even accessible online. Specialist knowledge is still useful but not as essential. There has been an explosion of online digital resources that are created and used collaboratively by amateur interest groups in partnership with specialist providers of content, advice, and web hosting. In this presentation I want to put the user/owner, not the institution, at the top of the model, and explore ways in which reciprocal relationships can, indeed must, be acknowledged and implemented in our practice.

This paper draws on recent experience in establishing a digital archive, PARADISEC (the Pacific and Regional Archive of Digital Sources in Endangered Cultures [www.paradisec.org.au](http://www.paradisec.org.au)) (Barwick, 2003b), and on participation in the establishment of digital audio access work-stations in two community cultural centres, or 'knowledge centres', in Australia's Northern Territory (Barwick, 2003a).

## 2 PARADISEC

PARADISEC (the Pacific and Regional Archive of Digital Sources in Endangered Cultures <http://www.paradisec.org.au>), was set up in 2003 to digitize and provide secure online access to Australian researchers' recordings of endangered languages and musics of the Asia-Pacific region. This paper attempts some creative thinking around the question of how to design and implement functional models for distributed digital archiving of cultural heritage audio and video materials.

PARADISEC is a collaborative digital research resource set up by the University of Sydney, the University of Melbourne, and the Australian National University in 2003, with funding from the Australia Research Council's Linkage Infrastructure Equipment and Facilities scheme. PARADISEC's primary goal is to provide digital preservation and access for endangered recorded field material from the region around Australia. At present we are focusing on audio material because there are established archival formats. Preservation of video presents a challenge that we have deferred until clearly agreed archival formats emerge. Existing digital materials, such as transcripts, theses, dictionaries, and so on, are also within the purview of the project.

Our original motivation was to provide a service to researchers by transferring their field recordings to CD, with a central database to keep track of the materials, but exciting recent developments in Australian

research infrastructure including the Grangenet data GRID initiative (<http://www.grangenet.net.au>) have encouraged us to extend our vision: to create an entirely electronic digital archive that could not only store and manage the digitized recordings, but also give networked electronic access to them. We see this as a resource not only for our depositors, but also for the stakeholder communities whose cultural heritage we are dealing with.

We thus have three goals for implementation of PARADISEC:

- Exploit the potential of distributed digital systems to build a collaborative cross-institutional resource;
- Ensure the future viability of the resource through linkages between Australian research and archival institutions, stakeholders in the region, and international bodies;
- Develop and implement appropriate models for electronic management of a digital research archive.

PARADISEC recognizes that our resource will only have future viability if we can recognize, build, and maintain collaborative relationships in every aspect of our work.

- Distributed resource creation—audio recordings are created by communities in collaboration with researchers, and ingested into our archive at a number of access points.
- Distributed management of the resource—by a consortium of research institutions.
- Distributed discovery—by users accessing our online metadata catalogue.
- Distributed data—the resource is preserved through backup and mirroring in multiple locations.
- Distributed ownership and rights—different arrangements apply depending on the researcher, institution, and cultural community involved, administered through a digital rights management mechanism built into our access architecture.
- Distributed access—digital files are accessed through online or CD distribution to users.
- Distributed support and training—participants contribute and access support and training resources from a number of different sources, including other institutions and consortia with similar concerns and interests.

These distributed systems can only work into the longer term if they are built on shared standards, formats, and procedures designed for long-term viability (Bird and Simons, 2002).

## 2.1 PARADISEC implementation

In 2003 we have three full-time staff. In Melbourne our Project Manager (Nicholas Thieberger) designs and maintains the metadata and database design, while data entry and administration (Amanda Harris) and audio digitization (Frank Davey) are taken care of in Sydney. The staff commu-

nicate by email and via our shared server 'Azoulay', which has an archive partition where preservation master files are periodically backed up to dedicated tapes in the University of Sydney Mass Storage System, and also regularly mirrored to the Australian Partnership in Advanced Computing national facility in Canberra over the Grangenet research and education data network. The Project Manager has administrator access to our project storage space there. We have just implemented an access system to allow depositors, content owners, community cultural centres and authorized general users online access to relevant files in the collection, via a system of password authentication, which will be discussed later.

### 3 Two Community Cultural Centres in Indigenous Australia

I believe that community cultural centres can play a key mediating role between individuals in remote communities and the digital archives where their cultural heritage is held (Grenoble and Whaley, 1998). When there are effective and rapid communications between individuals and their community cultural centres on the one hand, and between the cultural centres and the digital archive on the other, it becomes practical to reassert the cultural authority of the home communities and individuals. The archive can then take on its most effective role in providing a service of managing, backing up, and providing access to data, rather than having to assume the additional burden of administering the data it 'owns' at arms' length from the communities involved.

Because PARADISEC has only just proven its systems architecture and has not yet established functional partnerships with local communities in its geographical area of interest, I will illustrate my remarks about the potential functions of community cultural centres by reference to my recent experience in Australia (an area outside PARADISEC's current geographic focus). While there are still limitations of geography, technology differential, and language to contend with, recent advances in provision of network infrastructure in Australia mean that it has now become much easier to establish points of presence in local archives, known as 'Knowledge Centres' in Australia's Northern Territory, where I have been involved in helping to set up several local access units.

#### 3.1 Belyuen

The 'Belyuen Bangany Wangga' digital audio access workstation was established in 2002 in Belyuen, a community of about 300 people located on the Cox Peninsula south of Darwin. Funded by the Northern Territory Library and Information Service, the workstation houses a digital 'jukebox' of about 480 Wangga songs recorded in the community over the last sixty-one years (accessible via iTunes on an Apple eMac computer), complemented by other digital resources including photographs.

Here is just one example of the power of community partnerships with archives to properly describe resources, and hence make them findable

and useful. In 1952, University of Sydney anthropologist Professor A. P. Elkin visited Belyuen (then known as Delissaville), and made sound recordings and photographs of song performances there (described in Elkin and Jones, 1958). These photographs are now housed in the Elkin Collection in the University of Sydney Archives, while the sound recordings have become separated from the main collection. The location of the original tape recordings, in fact, is not currently known, but fortunately multiple copies were made and distributed by Elkin on process master discs. In 1995 I made DAT copies for Belyuen Community from the process master disc copies held at the Australian Broadcasting Commission Radio Archives, and since 2002 the individual songs have been locally accessible through the 'Belyuen Bangany Wangga' workstation.

Elkin had not recorded the names of any of the people involved in the performances, and the sound recordings are too poor in quality to allow identification of the individual singers and dancers. In 2003, as a result of the community's request to the University of Sydney Archives, I located the corresponding photographs in the Elkin Collection, and made digital copies to take back to the community for lodgement on the workstation for local access. Fifty-one years after the photographs were taken, we were finally able to identify every person shown in them, and from this, determine who the performers were in the sound recordings. Now that the University of Sydney archives has this information, it can make the resource available to the descendants of the performers, and also identify the appropriate people to ask if there are any requests for access or use of these parts of the collection.

### 3.2 Wadeye

The community of Wadeye (formerly Port Keats) comprises some 2000 people and is located about 6 hours' drive south of Darwin. Despite being the fifth largest town in the Northern Territory, until recently it had no library service. The Wadeye Aboriginal Languages Centre, established in 1991, embarked on a programme of recording stories and songs from speakers of the seven different languages represented within the community. Approximately 300 hours of sound recordings have now been transferred to CD by University of Sydney postgraduate student Alberto Furlan, and form just one part of the collection held at the Wadeye Knowledge centre, funded by the Northern Territory Library and Information Service. This resource is being used to create a multimedia Marri languages dictionary in collaboration with researchers Lysbeth Ford and Maree Klesch at Batchelor Institute of Indigenous Tertiary Education (Ford and Klesch, 2003). The Wadeye Languages Centre and Knowledge Centre also has sought out links with AIATSIS, the University of Sydney, and other institutions holding their cultural heritage materials, in order to obtain copies for local consultation, and also to ensure that these institutions can improve their descriptions and care of the materials.

## 4 Community Cultural Centres—Planning Principles

Here are some issues and principles that have emerged in setting up community cultural centres in Australia. Although the situation in the Pacific region that is PARADISEC's brief is much more variable and generally much more under-resourced (Batiri Williams, 2002), I believe that there are likely to be significant overlaps in concerns and planning issues. PARADISEC has already begun to explore some of these issues and to make partnerships with cultural communities in our region.

### 4.1 Components

The technological base needs to be locally sustainable. Because of the remote locations, harsh climatic conditions, lack of infrastructure to establish and maintain proper archival storage conditions, and frequently, the lack of local IT support personnel, the equipment needs to be sturdy, modular (so that components can be easily replaced), and reasonable in cost. To maximize local support, the equipment needs to be interoperable with other local resources such as language centres, schools, and councils. The use wherever possible of open source and standard formats and software increases the long-term sustainability of the resource.

### 4.2 Access interface

The access interface needs to reflect locally relevant means of knowledge management and organization, and to provide a bridge to external resources. It must thus be searchable by locally relevant categories (such as family names) and provide assurance that the right people have authorized access to the resource. It is preferable for the act of searching to provide the means of access to the resource (by hyperlinking) and to provide links to the object's location in external archival collections so that users can browse for other related materials.

### 4.3 Backup

Because of risks presented by the harsh climatic conditions in many remote areas it is essential for digital data to be backed up offsite in a trusted repository. Data backup alone is not sufficient for sustaining the resource. Long-term local access also depends on the provision of training and career paths for local staff, and ongoing planning and support for future use and development of the resource.

### 4.4 Rights

Permissions must be obtained and documented in accordance with local law as well as providing whatever protection is available from national and international legal requirements.

## 5 Repatriation and Rights

Many of us dealing with digital archiving find it very difficult to think about rights. At one conference I went to not long ago, the organizers

declared that rights were off the agenda because we could spend the whole meeting talking about them and never sort out the many technical, systems design, and planning issues that confronted us.

The question of rights presents a particular difficulty for ethnographic audiovisual collections that include recordings of works of verbal and performance art such as stories, songs, music, or drama. Australian and international law, ethical academic practice, and, in most instances, local laws of the originating community all recognize that the intellectual property and moral rights in these verbal and performance artworks remain with the original creators or their descendants (Seeger, 2001; HATII and NINCH, 2002). Since access to digital objects involves copying, in effect publishing them via the web, we cannot provide unfettered public web access to such digital recordings without explicit rights and permissions (IASA, 2002). Unfortunately, in the past many recordists of ethnographic data failed to collect or pass on to the archival institution the names of the individuals involved in creating and performing these works. Yet archives now need the permissions of the originating speaker/singers or their descendants to deliver intellectual property and moral rights for distribution.

The situation on recordings of other linguistic events that are not generally covered by international copyright law—such as conversations, oral histories or elicitations of word lists or verb paradigms (to name just a few)—is even more complex. In such cases, the speakers, their descendants, or other people with interests in the language or its expression may or may not hold strong opinions as to the proper way to authorize access to such material. Ethical academic practice would require that any such views be respected and taken account of, but once again many field recordings have been made without the recordist being aware of or taking note of local protocols. In some instances, indeed, these views may have developed in the home community since the original recording was made, and may continue to change over time. There is thus a need for the repository to establish ongoing reciprocal relationships with the individuals and communities involved.

However, this situation is mitigated somewhat by the fact that, where the means of discovery exist, some of the main users wishing to gain access to archival audiovisual cultural heritage material are themselves holders of significant rights and interests in it. This reality may be obscured by accepted models for digital archiving based on scientific practices of knowledge management, such as the ISO standard Reference Model for Open Archival Information Systems (OAIS, 2002). The underlying structure of the model presupposes a clear differentiation between the ‘producers’ and the ‘consumers’ of the data (Lavoie, 2000). For scientific data, this may well reflect prevailing conditions of production and use. Rights and intellectual property interests are typically much less straightforward for humanities data. The distributed nature of production and use of these audiovisual recordings of endangered cultural material means that there is often significant overlap between ‘producers’ and ‘consumers’. The ‘producers’ include the speakers or singers recorded, the researchers who



commissioned and usually recorded the events, and in some instances the researchers' employers, who may assert ownership of researchers' intellectual property. The main categories of 'consumers' likewise include the community members recorded or their descendants, and academic researchers.

### 5.1 A staged approach

PARADISEC plans to follow a staged approach to setting up ongoing relationships with rights holders.

#### Stage 1: Digitization for preservation

The repository indexes and maintains the data but refers to the depositor all access decisions. Ownership and ethical arrangements remain as pre-digitization. Distribution of digital objects is restricted to depositors, and users are authorized by them (password authentication).

#### Stage 2: Partnerships with community cultural centres

The repository provides metadata and digital objects to the local organization. Interoperable systems are established, including hardware, software, training, and planning. Appropriate means of local community access and interaction are negotiated and developed.

#### Stage 3: Community-based documentation projects

The repository collaborates with researchers and community cultural centres to collect, correct (if necessary), and annotate metadata and digital objects, including rights and access permissions and protocols.

#### Stage 4: Authorized outsider access to the collection

The repository refers access requests to rights holders via the community cultural centre, and on receiving appropriate permissions provides secure online access.

As further food for thought, here I present two idealized models that suggest some practical means of distribution of recordings to owners and to outside users once there exist good means of communication between a central digital repository and a remote area cultural centre. These models are based on recent experience of documenting archival recordings made forty to sixty years ago, and on my understanding of the networking architectures currently at our disposal. The barriers to implementing such models are largely to do with a lack of appropriate agreements and relationships rather than technical impediments.

##### 5.1.1 Distribution model 1: owners

The digital repository, let us say PARADISEC, passes the digital media files and what metadata they have, together with technical advice if required, to the community cultural centre, which distributes CDs or cassette copies of the digital recordings to the local owners. The cultural centre also collects additional information on the recordings (as a minimum, information on who holds the rights to them, but perhaps also corrections or additions to the metadata), and then passes back to the digital repository whatever information the community decides it would like to have noted there. Of

course there is an ongoing loop here. It is desirable to allow the community to add their own annotations to the digital objects, perhaps in the form of voice recordings, and to provide an effective way of linking such annotations to timecoded segments of digital audio or video.

### 5.1.2 Distribution model 2: outsiders

If an outside person, let us say having discovered the existence of the recordings through querying our online database, wishes to have access to those recordings, an access request is made through the repository and then passed on to the remote cultural centre, which contacts the rights holder, and passes permissions back to the repository. A password is then issued to enable the outside user to download the file for a set period of time. Outside users, too, should be asked to return to the repository any indexes or annotations they make that can add value to the resource, and to indicate any publications that refer to it.

## 6 'Designated Communities'

The Open Archival Information System reference model defines an archive as 'an organization that intends to preserve information for access and use by a designated community' (OAIS, 2002). When we have a distributed archive, some complex answers emerge to the question as to who our 'designated community' might be, and more work needs to be done to adequately describe the environment in which PARADISEC operates. Let me finish by placing PARADISEC itself in relation to the various communities and interest groups in which it participates.

In this paper I have focused on the 'designated community' of community cultural heritage interests, a network of individuals, communities, and institutions with whom we need to engage to perform our core functions. Of course, we are also answerable to and serving the academic community made up of our depositors, who also hold significant intellectual property interests in the material, and the broader academic research community, another significant user group that is likely to wish to engage with the collection, perhaps through quite different interfaces and at quite different levels of information from the community users. We also contribute to discussion, the setting of standards and procedures, and the sharing of relevant information resources within our local institutional digital repositories, and more broadly within national academic networks such as the Australian E-humanities network (<http://www.ehum.edu.au>). And of course there are important international communities and consortia with which we must engage, such as international standards-setting bodies, the Open Language Archives Community (<http://www.language-archives.org>) and the emerging Digital Endangered Languages and Musics Archives Network (<http://www.delaman.org>).

Because PARADISEC is an entirely digital archive that does not hold physical objects, we do not have a single point of presence. In some important ways, this virtual archive can be said to exist only socially, in our interactions with our various communities of interest. Planning for

our archives' digital future means imagining ourselves as actors and creators within that virtual society.

### Acknowledgements

This paper is based on research funded by the Australian Research Council and the School of Society, Culture and Performance at the University of Sydney. Thank you to Nick Thieberger and Amanda Harris at PARADISEC, Allan Marett (University of Sydney), Mark Crocombe (Wadeye Knowledge Centre), Cathy Winsley (Belyuen Community Council), the University of Sydney Archives, the Northern Territory Library and Information Service, Maree Klesch and Lys Ford at Batchelor Institute of Indigenous Tertiary Education, and the Belyuen and Wadeye communities.

### References

- Barwick, L.** (2003a). The Endangered Cultures Research Group's Digitisation Project: Using Digital Audio for Musicological Research. In Cole, C. and Craig, H. (eds), *Computing Arts: Digital Resources for Research in the Humanities Proceedings*. Sydney: Research Institute for Humanities and Social Sciences in association with the Australian Academy of the Humanities, pp. 147-159.
- Barwick, L.** (2003b). *Planning for PARADISEC: The Pacific and Regional Archive for Digital Sources in Endangered Cultures*. Presentation to the Ozeculture Conference, Brisbane Powerhouse, 31 July 2003. Published online at <http://www.acn.net.au/conference3/barwick/barwick.pdf> (accessed 9 September 2003).
- Batiri Williams, E.** (2002). *Digital Community Services: Pacific Libraries and Archives: Future Prospects and Responsibilities*. New Delhi: UNESCO, 2002.
- Bird, S. and Simons, G.** (2002). *Seven Dimensions of Portability for Language Documentation and Description: Proceedings of the Workshop on Portability in Human Language Technology*. Third International Conference on Language Resources and Evaluation, Las Palmas, Spain, May 2002.
- Elkin, A. P. and Jones, T. A.** (1958). *Arnhem Land Music*. Oceania Monograph 9. Sydney: University of Sydney.
- Ford, L. and Klesch, M.** (2003). *'It Won't Matter Soon: We'll All be Dead'*, Wadeye Aboriginal Languages Project. Second International Inside Out Conference on Higher Education and Community Engagement, Ipswich, Queensland, 3-5 July 2003. Published online at <http://www.uq.edu.au/insideout/pdfs/fordklesch.pdf> (accessed 9 September 2003).
- Grenoble, L. A. and Whaley, L. J.** (1998). *Endangered Languages: Language Loss and Community Response*. Cambridge, UK: Cambridge University Press.
- Humanities Advanced Technology and Information Institute (HATII), and the National Initiative for a Networked Cultural Heritage (NINCH)** (2002). *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*. Chapter IV, 'Rights Management'. Published online at <http://www.nyu.edu/its/humanities//ninchguide/IV/> (accessed 9 September 2003).
- International Association of Sound and Audiovisual Archives (IASA)** (2002). *Copyright and other intellectual property rights*. Available online at <http://www.iasa-web.org/iasa0016.htm> (accessed 20 September 2003).

**Lavoie, B.** (2000). Meeting the Challenges of Digital Preservation: The OAIS Reference Model. *Online Computer Library Centre (OCLC) Newsletter*. 243: 26–30 (January/February 2000). Available online at <http://www2.oclc.org/oclc/pdf/news243.pdf> (accessed 9 September 2003).

**OAIS Consultative Committee for Space Data Systems (CCSDS)** (2002). CCSDS 650.0-B-1. Reference Model for an Open Archival Information System (OAIS). *Blue Book*. Issue 1 (January 2002). Recommendation adopted as ISO 14721:2002. Published online at [http://ssdoo.gsfc.nasa.gov/nost/isoas/ref\\_model.html](http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html) (accessed 9 September 2003).

**Seeger, A.** (2001). Intellectual property and audiovisual archives and collections. In *Folk Heritage Collections in Crisis*. Washington: Council for Library and Information Resources. Online version at [www.clir.org/pubs/reports/pub96/rights.html](http://www.clir.org/pubs/reports/pub96/rights.html) (accessed 8 September 2003).