

Barwick, L. (2005). Networking digital data on endangered languages of the Asia Pacific region.
International Journal of Indigenous Research, 1(1), 11-16.

Note: This is a preprint. Consult the original for citation information.

Networking digital data on endangered languages of the Asia Pacific region

Linda Barwick
University of Sydney

In 1900, the French anthropologist Léon Azoulay predicted that “the fixation of sounds by the phonograph [was] about to work in all human fields of activity and knowledge a revolution as great as that produced by the fixation of luminous images in photography” (Azoulay, 1900a). In an essay surveying the possible uses of sound recording technologies to support scientific research, and calling for the establishment of phonographic archives, Azoulay exclaimed: “all that of which linguistics is incapable, the phonograph from now on makes possible” (Azoulay, 1900b).

In the 105 years since that date, scholars of languages, cultures and musics from around the world have enthusiastically embraced the potential of portable recording technologies — initially audio, and since the 1970s, video — to capture the events that they study, whether these are stories, songs, performances or conversations. Because of the changing nature of people, societies and technologies (Barwick, 1991), many of these ethnographic recordings have outlasted the people, traditions and even languages that they recorded. Consequently many of these recordings have become even more significant than either the recordists or the people recorded may have realised at the time the recordings were being made (Schüller, 2004). These recordings now have immense significance not only for researchers but also for the descendants of the people recorded and the cultural heritage communities whose traditions and languages they encode.

One of the reasons that the documents made by ethnographic researchers are of particular interest to cultural heritage communities is because of the emphasis of the ethnographic method on recording and describing real-life events, often with explanations by the culture-bearers themselves. Ethnographic methods are not used only by anthropologists, but also by field workers in a number of humanities disciplines such as field linguistics and ethnomusicology, as well as many social sciences. The social scientist Ann Bowling summarises the broad conception of ‘ethnography’ I am using here:

the study of people in their natural settings; a descriptive account of social life and culture in a defined social system, based on qualitative methods (e.g. detailed observations, unstructured interviews, analysis of documents). (Bowling, 1997)

Thus, the sound- and video-recordings made by ethnographic researchers, especially in the humanities, are typically used as the primary documentation forming the touchstone for subsequent analysis and description. Indeed, the same primary documents may well be analysed by researchers from different disciplines, using discipline-specific tools for annotation and analysis.

For example, a recording of a song performance including discussion by the performers about the subject matter of the songs may be of interest to musicologists, linguists, anthropologists, or oral historians, as well as being highly prized by the local community for its family history, entertainment or educational value, as well as an authentic record of

the voices of dearly loved deceased community members. Each researcher may produce his or her own annotation of the content of the recording, all of which are likely to be of interest to the cultural heritage community.

Furthermore, in the interests of rich descriptive documentation, ethnographic researchers typically produce data in many different media:

- Audio recordings may be used for music sessions, interviews, language elicitations, conversations, story-telling sessions;
- Video recordings may be necessary for studies of movement and dance, social interaction, sign language, gesture studies and performance technique;
- Images such as photos, diagrams, graphs, maps, paintings and sketches may be produced to support primary documentation of events or to illustrate analytical techniques;
- Databases are needed for systematic organization and analysis of, for instance, lexical, genealogical, taxonomic, or geographical information;
- Texts produced in the research process include field notes, transcripts of tapes, analyses, dictionaries, and sometimes use specialist codings or fonts.

In many cases, the primary audiovisual recordings that are arguably of most interest to future generations of researchers and culture bearers need to be made sense of by reference to the supporting documentation that indexes and interprets the recordings. Conversely, the secondary documentation depends on the primary documentation for its validation as accurate research output. As Roberts and Wilson observe:

The qualitative research data that is generated on the basis of these primary documents are likely to be: rich and voluminous, shedding light on the lived experience of the 'being-in-the- world' and the interactions inherent in complex social phenomena. Analysis of such data, however, is complex and time consuming. (Roberts and Wilson, 2002)

However valuable these recordings may be to researchers and to communities, however, many of them are now threatened. We are in the midst of a crisis of format obsolescence for many of the most common audio- and video-recording formats used in the 20th century (Council on Library and Information Resources (CLIR), 2001). Not only are the tape media on which these recordings were typically made deteriorating over time, especially if they have been kept in dusty or humid conditions, but also the very recording formats and platforms are disappearing. Every single audiovisual format in which I recorded my own field collections in the 1980s and 1990s is now endangered:

- reel-to-reel audio tape and playback machines are no longer manufactured;
- good quality audio cassette stock is no longer manufactured, and neither are high-quality machines on which to replay the tapes (although consumer level tapes and playback machines continue to be manufactured);
- VHS and Hi-8 video formats are fast being phased out in favour of DVD; and
- even born-digital formats like Digital Audio Tape are fast disappearing (Sony recently ceased manufacture of the high-quality DAT decks needed to produce high-quality archival transfers of the tapes).

Sound and audiovisual archivists are well informed about the extent of the problem and most major national sound archives are already well advanced in future-proofing their valuable collections by transferring them to suitable digital formats (Bradley, 2003; Boston, 2003; International Association of Sound and Audiovisual Archives (IASA), 2004; Webb, 2003). However, as the archivist Dietrich Schüller observes, sound and audiovisual archives proper include only about 20% of world-wide holdings of significant cultural recordings: the other 80% are held by universities, museums or collections of private researchers, which are not only “notoriously underfunded” but in some cases neither “aware of the threats facing their unique and irreplaceable holdings” nor “equipped to adequately handle and preserve these valuable sources” (Schüller, 2004). Even if funds become available for digitisation, all too often the digital versions of research collections are created in unsustainable formats (e.g. mp3) or entrusted to unreliable storage media (e.g. CD-R or DVD-R), and little ongoing funding is available for long-term digital curation of collections through auditing, refreshing and migrating the digital files in a mass storage system.

The question of sustainability of digital research documentation is a question that is currently receiving some attention in Australia through a number of government funded initiatives including the Australian Partnership for Sustainable Repositories (Bradley, 2005). There is considerable concern in Australia and internationally not only about the sustainability of digital data (Webb, 2003), but also about the organisational and other impediments to auditing and management of the many small research collections held within the University sector (Cathro, 2004).

To return to the specific question of the sustainability of small audiovisual collections, Schüller goes on to caution:

The challenge facing audiovisual collections in general, and the holdings of small research institutions in particular, which reflect the cultural and linguistic diversity of mankind as recorded over the past 100 years, can only be met by an international action plan which coordinates governmental, scientific, and private interests and activities in this field. Unless systematic efforts are made on an international scale to establish adequate preservation schemes for those collections, most of these documents will be lost in the next 10–30 years, either by deterioration of the materials or by obsolescence of replay equipment. (Schüller, 2004)

The problems facing small collections are indeed grave, and cannot be solved by one researcher or one institution alone. As Schüller points out, it is a problem of coordination as much as one of technical know-how. However, I believe there is some cause for optimism. Because if there is one thing that has emerged as a powerful agent for change in research practice in recent years, it is the emergence of web-based electronic tele-collaboration tools.

As an example, let me turn to one specific implementation of these challenges, the efforts made by a group of Australian field researchers who banded together in 2003 to establish PARADISEC, the Pacific and Regional Archive for Digital Sources in Endangered Cultures. Linguists, musicologists and anthropologists who had undertaken their

fieldwork in the preceding 30 years were now beginning to look to retirement and were dismayed to discover that their research collections were under threat. We were aware that digitisation of our collections was the first step to making them sustainable, but discovered that the collection policies of our national institutions excluded recordings made outside Australia, and that our Universities had neither the resources nor the specialist expertise to process our collections to appropriate archival standards. With financial support from our Universities we won funding from the Australian Research Council's Linkage Infrastructure Equipment and Facilities scheme to purchase equipment for archival-quality digitisation (using the Quadriga audio archiving system) and mass digital storage systems. As well as producing preservation copies of the audio recordings we also routinely produced CD-audio and mp3 derivatives of the files for access by researchers and communities, and we are now trialling delivery of digital images of field notes to accompany some of the older collections, as well as planning to begin video ingestion. Our website (paradisec.org.au) is increasingly used as an information resource and we also run training courses for researchers, postgraduate students and local archives.

Fortunately, PARADISEC was being planned just at a time when the Australian higher education system was benefiting from the establishment of high-bandwidth dedicated networking through the Grangenet and AARnet networks, and networked storage via the Australian Partnership for Advanced Computing's (APAC's) Store facility. In fact, PARADISEC is a new kind of entity that has been enabled by this infrastructure and that could not exist without it. We are cross-institutional in our governance and our functioning. We are governed by a steering committee consisting of representatives from each of the four participating institutions (the University of Sydney, the University of Melbourne, the Australian National University (ANU) and the University of New England). Our online catalogue is designed and managed by Nicholas Thieberger, at the University of Melbourne, and our main audio ingestion unit is at the University of Sydney. Our website is hosted in the Research School of Asia and Pacific Studies at ANU, and the APAC Store facility that hosts the mirror of our collection is also located at ANU. Data is contributed and continues to be owned by the four participating institutions.

Not only has the development of this national networking and storage infrastructure for higher education enabled PARADISEC to establish secure processes for backup and mirroring of our collection, but it has also enabled us to participate in global networks for discovery and management of our collection. We have adopted international standards for archiving and generic description of digital media files developed by bodies like the International Association for Sound and Audiovisual Archives (International Association of Sound and Audiovisual Archives (IASA), 2004): our preservation audio files, for example, are archived as 24-bit 96-kiloherz stereo Broadcast Wave Format files and our backup, mirroring and collection auditing is supported by various automatic management scripts developed at the University of Sydney and APAC. Our metadata schema has been designed to allow metadata harvesting via the Open Language Archives Community (www.language-archives.org), a sub-community of the Open Archives Initiative, which also provides a simple search interface to our collection and a tool for monitoring metadata quality. And we collaborate through various international and global fora such as the Digital Endangered Languages and Musics Archives Network (www.delaman.org),

which is exploring models for global federated discovery of resources on endangered languages and musics from research archives around the world, and the Asia Pacific Advanced Network (apan.net) meetings, which have recently seen us sharing a platform with earth climate systems modelers in discussing the potential of the semantic web for large-scale digital data archives.

For humanities researchers like me, perhaps one of the most exciting applications of the emerging globally-networked ICT environment is the potential to include cultural heritage owners within our networked research communities. As Sally Jo Cunningham observed in her review of the 6th International conference on Asian Digital Libraries, “it is hoped that digital libraries won't simply archive indigenous documents, but will become a part of an infrastructure that supports local cultures, heritage, and values” (Cunningham, 2004). For ethical and rights management reasons, curators of digital collections of cultural material must reach out to indigenous communities whose cultural heritage is represented in our collections, not only to try to overcome the digital divide (Barwick and Thieberger, 2005), but also to find new ways of doing research (Barwick, 2004; Seeger, 2004).

We do not simply need to make our research results available in the relevant community through appropriate formats and technologies and locally-relevant access means (Barwick, 2003). We also need to align our recording and archiving practices to support traditional modes of authority and maintenance of traditional practices, and we need to make sure that wherever possible we establish and maintain an ongoing dialogue with the cultural heritage owners. For each of these steps we need practical exemplars and feedback from our research collaborators, for the specific implementations of these ethical principles will differ over time, place and social environment. While some collaboration and interaction can be supported via remote electronic access, collaboration will continue to be most effective when it is based on personal contact and engagement. PARADISEC offers

Crucially, we also need to recognize that indigenous communities themselves are increasingly making their own recordings and creating their own publications and local repositories, often building on the basis of repatriated materials originating in academic collections. An emerging Australian example is the proposed National Recording Project for Indigenous Performance in Australia, being developed in 2005 by the Yothu Yindi Foundation in collaboration with researchers from the University of Sydney, the University of Melbourne, the University of Queensland and Charles Darwin University, and with the spread of readily available portable recording technologies many other examples are emerging worldwide. These initiatives by indigenous communities to take charge of looking after their own cultural heritage and managing their own representations to the outside world are as much an opportunity for new forms of engagement as they are a challenge to traditional models of the relationship between researchers and their collaborators.

I believe that humanities researchers today stand at an exciting juncture. Through collaboration within and across national, cultural and disciplinary boundaries — between researchers and indigenous communities, between humanities scholars and ICT

specialists, between likeminded colleagues in different institutions —we can see a new research landscape opening up before us. As for any journey, we need to be well-prepared, have a good heart and to be ready to take some risks as well as to enjoy ourselves. I look forward to learning more from my Malaysian colleagues and hosts at this meeting, and to hearing the input of all those present as to which direction you think we should set out on from here.

REFERENCES

- Azoulay, L. (1900a). The future of the phonograph. *Literary Digest* **21** (2): 43
- Azoulay, L. (1900b). L'ère nouvelle des sons et des bruits: Musées et archives phonographiques. *Bulletins et Memoires de la Société d'Anthropologie de Paris* **1** (3): 172-178
- Barwick, L. (1991). Partial records: the problematics of loss and change in historical ethnomusicology. In *Tradition and its future in music* (Eds, Tokumaru, Y., Ohmiya, M., Kanazawa, M., Yamaguti, O., Tukitani, T., Takamatsu, A. and Shimosako, M.). Mita Press, Tokyo and Osaka, pp. 139-45
- Barwick, L. (2003). The Endangered Cultures Research Group's digitisation project: using digital audio for musicological research. In *Computing Arts 2001: Digital resources for research in the humanities* (Eds, Cole, C. and Craig, H.). Research Institute for Humanities and Social Sciences, University of Sydney, in association with the Australian Academy of the Humanities, University of Sydney, pp. 147-159
- Barwick, L. (2004). Turning it all upside down . . . Imagining a distributed digital audiovisual archive. *Literary and Linguistic Computing* **19** (3): 253-263
- Barwick, L. and Thieberger, N. (2005). Cybraries in paradise: new technologies and ethnographic repositories. In *New libraries and knowledge spaces: critical perspectives on information and education* (Eds, Kapitzke, C. and Bruce, B. C.). Lawrence Erlbaum Associates, Mahwah, N.J.
- Boston, G. (2003). Survey of Endangered Audiovisual Carriers, 2003. Survey conducted by George Boston, Secretary, Technical Committee of the International Association of Sound and Audiovisual Archives (IASA) with assistance from the International Council of Archives, on behalf of UNESCO's Information Society Division. UNESCO, Paris. Available online at <http://portal.unesco.org/ci/ev.php?URL_ID=13437&URL_DO=DO_TOPIC&URL_SECTION=201&reload=1073385878&PHPSESSID=019c5625ccdffaf7095faf0aade5c128> (accessed 14 August 2004)
- Bowling, A. (1997). *Research methods in health*. Open University Press, Buckingham
- Bradley, K. (2003). Critical choices, critical decisions: Sound archiving and changing technology. In *Researchers, Communities, Institutions and Sound Recordings* (Eds, Barwick, L., Marett, A. and Simpson, J.). Online refereed conference proceedings. University of Sydney, Sydney. Available online at <<http://conferences.arts.usyd.edu.au/viewpaper.php?id=57&cf=2>> (accessed 14 June 2005)
- Bradley, K. (2005). APSR sustainability issues discussion paper. Australian Partnership for Sustainable Repositories, Canberra

- Cathro, W. (2004). Preserving the outputs of research. Online conference paper, International Conference on Archiving Web Resources. National Library of Australia, Canberra. Available online at <<http://www.nla.gov.au/nla/staffpaper/2004/cathro1.html>> (accessed 14 June 2005)
- Council on Library and Information Resources (CLIR) (2001). Folk Heritage Collections in Crisis. Council on Library and Information Resources, Washington. Available online at <<http://www.clir.org/pubs/reports/pub96/contents.html>> (accessed 14 August 2004)
- Cunningham, S. J. (2004). Report on the Sixth International Conference on Asian Digital Libraries (ICADL 2003). *D-Lib Magazine* **10 (1)**. Available online at <http://www.dlib.org/dlib/january04/cunningham/01cunningham.html> (accessed 14 June 2004)
- International Association of Sound and Audiovisual Archives (IASA) (2004). *Guidelines on the production and preservation of digital audio objects (IASA-TC04)*. International Association of Sound and Audiovisual Archives, Aarhus, Denmark
- Roberts, K. A. and Wilson, R. W. (2002). ICT and the research process: Issues around the compatibility of technology with qualitative data analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* **3 (2)**. Available online at <<http://www.qualitative-research.net/fqs-texte/2-02/2-02robertswilson-e.htm>> (accessed 14 June 2004)
- Schüller, D. (2004). Safeguarding the documentary heritage of cultural and linguistic diversity. *Language Archive Newsletter* **1 (3)**: 9-10
- Seeger, A. (2004). New technology requires new collaborations: Changing ourselves to better shape the future. *Musicology Australia* **27**
- Webb, C. (2003). Guidelines for the preservation of digital heritage. UNESCO document CI-2003/WS/3. UNESCO Information Society Division, Paris. Available online at <<http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>> (accessed June 14 2005)