

Chapter 8

Cybraries in Paradise: New Technologies and Ethnographic Repositories

Linda Barwick (University of Sydney)

Nicholas Thieberger (University of Melbourne)

Digital technologies are altering research practices surrounding creation and use of ethnographic field recordings, and the methodologies and paradigms of the disciplines centered around their interpretation. In this chapter we discuss some examples of our current research practices as fieldworkers in active engagement with cultural heritage communities documenting music and language in the Asia-Pacific region, and as developers and curators of the digital repository PARADISEC (the Pacific and Regional Archive for Digital Sources in Endangered Cultures: <<http://paradisec.org.au>>). We suggest a number of benefits that the use of digital technologies can bring to the recording of material from small and endangered cultures, and to its re-use by communities and researchers. We believe it is a matter of social justice as well as scientific interest that ethnographic recordings held in higher education institutions should be preserved and made accessible to future generations. We argue that, with appropriate planning and care by researchers, digitization of research recordings in audiovisual media can facilitate access by remote communities to records of their cultural heritage held in higher education institutions to a far greater extent than was possible in the analog age.

The advantages of using audio recording technology for ethnographic and scientific purposes were understood from early in the twentieth century. In 1900, the French anthropologist, Léon Azoulay, was perhaps the first to point out the possibilities for systematic analysis of sound made possible by the ability to reproduce at will and in fixed detail what had hitherto been an evanescent mode of human production:

Nor is it possible for linguistics to faithfully represent the speech of living languages even with the help of its arbitrary and totally insufficient transcriptions, made up for want of better and which no agreement has ever made uniform ... all that of which linguistics is incapable, the phonograph from now on makes possible. (Azoulay, 1900, 175)

Azoulay foresaw the establishment of phonographic museums, and his own collection of 388 cylinders recorded at the Paris Universal Exposition of 1900 was to form the basis of the first French phonographic archive (Pitoëff, 1993). As predicted by Azoulay, audiovisual recordings now constitute primary data for linguistics, musicology, anthropology, oral history, and other humanities disciplines. Much publication in these areas depends on researchers' transcriptions and analysis of their own field recordings of language, music and events.

However, we now stand at a point where the audiovisual documentary heritage created in analog recordings is reaching a crisis of format obsolescence (Council on Library and Information Resources (CLIR), 2001). It is now recognized that digitization is the only effective way to maintain access to analog audio recordings (National Library of Australia (NLA), 2000; Humanities Advanced Technology and Information Institute (HATII) & National Initiative for a Networked Cultural Heritage (NINCH), 2002; DigiCULT Consortium, 2004). This is an international problem. The technical committee of the International Association of Sound Archives has estimated that worldwide 100 million hours of unique analog recordings in archives need digitizing—far more than can possibly be processed in the lifetime of the dwindling number of functional analog playback machines—and this estimate does not include recordings in higher education institutions (Bradley, 2003; Boston, 2003; Schüller, 2004; Wright & Williams, 2001).

While planning for digitization of recordings held in specialist audiovisual archives is relatively advanced, analog research recordings in higher education institutions face particular issues. The solitary nature of much humanities research practice and the ready availability of inexpensive consumer audio recording technologies (e.g., audio cassette) has led to the accumulation of many private research collections of important audiovisual data *held* within higher education institutions, but not necessarily *registered* with them. All too often, when researchers retire or die,

their orphaned collections languish forgotten in filing cabinets or, even worse, are disposed of by relatives who do not understand their cultural significance. Furthermore, as will be discussed further below, our own research institutions often have not realised the significance of these recordings, regarding them as by-products of our research, rather than integral to it.

In our research practice, we frequently confront the discrepancies and inequalities between our relatively privileged Western knowledge institutions and the remote communities with whom we collaborate. In our work, the “digital divide” (Benton Foundation Digital Divide Network Staff, 2004; Gorski, 2001; Macharia, 2004) is only too apparent, most obviously in the geographic and economic impediments to the availability of infrastructure for electronic information and communication technologies within the communities with which we work. But there is another aspect to the digital divide, one that is perhaps less obvious to the outside observer, but equally if not more significant if we wish to ensure the communication of our research results to the tradition-bearers. This relates to the lack of technological development to enable discovery, description and citation of the audio-visual media formats in which much culturally significant content is most appropriately recorded.

Western knowledge institutions and research methodologies have favored and developed formats and media that best suit text-based modes of scholarly communication. Many other societies preserve and transmit cultural knowledge instead by non-textual means, such as musical, dramatic, verbal and visual arts. It is, therefore, no coincidence that audiovisual recordings—the primary formats in which the languages, performance traditions and other knowledge of pre-literate cultures have been captured by researchers for documentation and analysis—are of continuing interest to the communities of those recorded, indeed of more interest than the theoretical articles produced about or based on the recordings (Seeger 2004, p.1). This is an issue that has been poorly understood by many research institutions.

Our research institutions tend to see the preservation, indexation and even digitization of ethnographic *texts* as part of their core business, while the research *recordings* on which they depend are seen as mere by-products of the research. Preservation of and access to the ethnographic articles of, say, the 1920s have been well-served by University libraries, which have pushed the development of appropriate technologies for making textual material available in digital form. Written articles of this period are easily discovered through interoperable online library catalogues; if not already available in electronic form, texts can be digitized and distributed quickly and accurately by means of widely available copying and optical character recognition software (all the more so since they are now out of copyright); and the standardized apparatus of scholarly

referencing together with the ability to cite excerpts of textual documents directly within one's own text means that researchers can easily engage with the texts and their ideas at a fine level of detail. By contrast, preservation and access to the wax cylinders of the same period has been the domain of specialist sound archives rather than University libraries. Many if not most ethnographic wax cylinder recordings were unpublished and thus not systematically collected or described in mainstream University library catalogues; only specialist sound archives and a few individual collectors have the means to replay fragile wax cylinder recordings; and standardized means of describing the specialized content and citing it at a close level of granularity are only now beginning to emerge. These are impediments for researchers, but far more so for communities who now wish to regain access to their heritage in audiovisual media.

The application of digital technologies within our research practice may seem incongruous and even inequitable to those who see new technologies as exemplars of the privilege and consumerism of the developed world, complicit in globalization by providing instant access to information across regional and national boundaries. Yet, as Steven Feld has observed, the tendency to homogenize associated with globalization can be opposed by the ability of each locale—if it can afford it—to reflect on itself and its identity, using the same technologies (Feld, 1994). In fact, digital technologies are already being used by community members themselves to preserve ethnodiversity through locally-made recordings, and researchers' access to the means of production and safekeeping of recordings is highly valued by many of our community collaborators.

The next three sections of this paper deal with particular instances of current practice in using digital resources in ethnographic research. The Belyuen Bangany Wangga local repository, established in the Northern Territory community of Belyuen in 2002, is an example of a local repository for digital access to audio recordings of songs from the community, a resource originally compiled for research purposes by Barwick and her collaborators. The linguistic case study discusses Thieberger's use of digital technologies in his research on the language of South Efate, from Vanuatu. Finally, we discuss the rationale and practices used in establishing a new digital repository, from our experiences in establishing PARADISEC in 2003-2004.

Case Study 1: Belyuen Bangany Wangga local repository

The following case study reports on collaborative efforts by researchers and communities in the Daly region of northwest Australia to establish suitable practical platforms for digital repatriation of archival audio

recordings. Barwick and her collaborators, musicologist Allan Marett and linguist Lysbeth Ford, have been centrally involved in establishing such facilities in the Daly region communities of Belyuen (formerly known as Delissaville) and Wadeye (formerly known as Port Keats) in the Northern Territory. The present case study focuses on the digital audio access workstation established at Belyuen, located on the Cox Peninsula southwest of Darwin, current population around 300. The workstation is named *Belyuen Bangany Wangga* (“Belyuen Song and Dance”) in Batjamalh, the largest of the five community languages.

For centuries, Aboriginal people of the Daly region of northwestern Northern Territory have been performing their music for outsiders. Music and dance performances continue to form an integral component of ceremonial exchange and trade relationships with neighboring Indigenous groups, and musical performances for non-Indigenous people have been documented from the time of early settlement of the Darwin area. Sound recordings of such performances, dating from as early as 1942, are held in private collections, state and national sound archives and some recordings, notably those made in the course of anthropological expeditions led by Charles Mountford and A.P. Elkin, have been published, distributed, and broadcast nationally and internationally (Elkin, 1957, 1953; Mountford, 1949). From the 1980s to the present, Marett, Barwick and Ford have been involved in making new recordings of Belyuen singers.

Residents of Belyuen have included some of the most prolific and influential composers and singers of the public didjeridu-accompanied *wangga* song genre, also known as Nyindi-yindi. These singers include Jimmy Bunduck, George Ahmat, Tommy Burrenjuck, Bobby Lane, Jimmy Mulluk, Billy Mandji, Rusty Moreen, Colin Worambu Ferguson, Kenny Burrenjuck, Simon Moreen, David Woodie, and Roger Yarrowin. With funding from the Northern Territory Library and Information Service, a digital audio workstation *Belyuen Bangany Wangga* was established in 2002 to give local public access to archival recordings of these Belyuen singers (Barwick, 2003; Marett, 2003). The collection of some 480 Wangga songs now held in the *Belyuen Bangany Wangga* digital audio workstation represents a rich local tradition comprising many ethnographic recordings that, having been dispersed in different archives all over the world, have now returned to the home community.

Discovering archival recordings

During the 1990s Barwick and her research team investigated archival holdings of recordings from the Daly region, initially in order to compare them with current musical practices in the region. It soon emerged that there was great interest within the community in having access to these archival recordings. The compilation of archival recordings

made available on the *Belyuen Bangany Wangga* access workstation is the result of years of collaborative research between researchers and the community.

Some archival detective work and musicological expertise was necessary to identify early recordings as coming from the communities in question. Candidate recordings for repatriation were identified in the first instance by searching the catalogues of relevant national collections such as the Australian Institute for Aboriginal and Torres Strait Islander Studies (AIATSIS), the ABC Radio Archives, and the National Film and Sound Archive. However, initial catalogue searches did not necessarily reveal all relevant recordings. The reasons for this situation are many and complex, and our eventual success depended on close acquaintance with community members and local song styles, as well as knowledge of the historical circumstances in which early recordings were made and disseminated, and a willingness to engage with the difficult cataloguing systems of a number of different archival institutions.

The biggest impediment to discovering the recordings was that early sound recordists sometimes failed to collect the singers' names and communities of origin. For example, the recording place of the earliest recording from Belyuen, made in 1942 by ABC war correspondent, Peter Hemery,¹ was deliberately omitted from the record because of the strategic importance for the Australian war effort of radar stations located near Belyuen community. Language difficulties may have prevented recordists from noting accurate information about singers, most of whom have learnt English as a second language. Anthropological interest in group ownership and participation in ceremonial performances may also have contributed to the failure of anthropologists such as A.P. Elkin to record the names of individual singers (Elkin & Jones, 1958). To further complicate an already complex situation, people from these communities, especially singers, frequently travelled to other places for work or social reasons, so that the identification of the community of origin of a singer was not always known to recordists. For example, one Belyuen singer was variously recorded by four different researchers at many different locations between 1959 and 1988: at Bagot, Darwin, Mandorah, Batchelor, Katherine, Beswick, Daly River in the Northern Territory and Kununurra in Western Australia, as well as his home community of Belyuen. Furthermore, people may be known by different names in different communities. For example, one singer who resides occasionally at both Belyuen and Wadeye is known by a different surname in each place.

Another impediment to identification of candidate recordings was that some old recordings—notably those made by ABC recordists in collaboration with the anthropologists C.P. Mountford in 1948 and A.P. Elkin in 1952-3—were edited for radio broadcasts, film soundtracks, pub-

lished compilations, and sometimes further disseminated on unpublished process discs and tapes to interested researchers. This has resulted in a confusing number of different instances of the same recording appearing in different archives. Further complications in identifying instances of the same recording resulted from the varying attribution of authorship of the same recording to the ABC recordist, supervising sound engineer, or anthropologist who directed the entire expedition (e.g., some recordings attributed to Charles P. Mountford were actually made by ABC sound recordist Ray Giles, supervised by journalist Colin Simpson).

Indexing the recordings

Barwick visited the various archives holding relevant material and obtained Digital Audio Tape (DAT) copies from the access reel-to-reel copies. She originally made cassette copies for replay in the community, but as the project developed, digital audio replayed via computer was increasingly used. Compared to audio cassettes, the digital versions of recordings were significantly easier to work with in replay, transcription and documentation of the song because of improved audio quality, random access to individual songs, and improved ability to isolate problematic passages and where necessary slow them down for transcription. Once candidate recordings were assembled from archival resources, Barwick and her research collaborators replayed them to community members in order to identify the singers and other performers, and document contextual information about the performance, the topic and composer of each song, and where possible the song text and its translation. This information was then combined with the archival description to produce a catalogue of each song in the collection, the occasions on which it had been performed, and the recordings on which it occurred.

Designing local access to archival recordings

Until recently there was limited infrastructure for storing sound recordings and making them accessible within the community. Belyuen school held copies of some recordings on cassette and VHS video but there was no systematic attempt to assemble a collection. From time to time, people from the community visited Canberra to retrieve cassette copies of recordings from the AIATSIS sound archives. As the research project proceeded, Barwick and her collaborators distributed cassette copies of archival recordings to collaborators within the community on each occasion that the research team returned. However, these cassette copies tended to be short-lived. Frequently the cassette copies obtained from archives, being third generation copies of analog originals, were relatively low fidelity

and did not appeal to the ear. Both research team and community members were enthusiastic at the prospect of providing ongoing local access to recordings in digital form.

In deciding on the appropriate platform for local access, the community and researchers wanted to make the collection searchable by locally relevant categories, but also to provide information to allow the recordings to be located in the source archival collections. The components needed to be sturdy, modular and reasonably inexpensive, using locally sustainable technologies and personnel. Community interest tended to centre on particular songs and performers, so it was decided for community access to provide access at the level of the song rather than the whole recording. An iTunes database on an Apple eMac computer provided cost-effective, stable and scriptable access that enabled individuals within the community to choose their own selection of songs.

Permissions from rights holders (i.e., institutions and/or recordists) as well as performers and their families were sought and documented before launching the database. The community decided to restrict initial access to Belyuen itself, but at a future date it is likely that a selection of Belyuen recordings will be made publicly available through the Northern Territory Library and Information Service and/or the University of Sydney. A CD of Belyuen music, *Rak Badjalarr* (Marett, Barwick, & Ford, 2001), published by AIATSIS in 2001, has made available a selection of the results of this research for the repertory of the singer Bobby Lane Lambudju (1941-1993), including archival recordings made by Alice Moyle and LaMont West. Further CDs of other Belyuen singers are planned.

Case Study 2: Language documentation and the ethnographic cybrary

Making data available and reusable are two central foci of Thieberger's description of the indigenous language of South Efate (Central Vanuatu). The documentation of South Efate has taken into account newly emerging tools and processes that can be used to represent spoken natural language. Linguists routinely record endangered languages for which no prior documentation exists. This is vitally important work which often records language structures and knowledge of the culture and physical environment that would otherwise be lost (Maffi, 2001). However, while it is typical for the interpretation and analysis of this data to be published, the raw data is rarely made available. The data—tapes, field notes, photographs, and perhaps video—are often not properly described, catalogued, or made accessible, especially in the absence of a dedicated repository. Developments in technologies now make it possible for audio and video data to be

made widely available and readily searchable, subject to intellectual property issues, the enforcement of which is also gaining more attention.

A field linguist typically engages in recording aspects of a language for analysis in a written grammar. As argued by Duranti, grammars are necessarily partial documents that contain analysis of the parts of the language that we currently consider necessary to include in a style that is currently fashionable (Duranti, 1997). Looking back over grammars written in the past makes one aware of how such fashions change and how difficult it can be to find information about topics not covered in the grammar. Efforts to re-learn languages based on historical materials—as is becoming increasingly important to many Indigenous Australians, for example—have also highlighted the importance of a well-described broad range of language usage data that is securely archived.

There is a growing awareness of the need for a fieldworker to be recording as much information in as many contexts in the field as possible, as their recordings may well be the only documentation made of the language. A concomitant is the importance of data management for the preservation of our audio and video recordings and photographs so that they will be available for others beyond our own use of them. Himmelmann (1998) observes that documentation and description are two parts of the activity engaged in by field linguists, but that documentation has traditionally been considered a secondary task to the production of a language description. In a similar vein, Woodbury (2003) notes that language documentation has always been a part of the linguistic effort, but that new technological approaches offer a way of refocusing our work. Bird and Simons (2003) discuss the technological implications of developing a dataset that will endure over time and remain accessible.

Emphasising the documentation means that certain products of our work, such as text collections and dictionaries, become primary rather than incidental. Similarly, our concern with the reusability of our work takes on a primary focus so that the data has a use for others after we have done our analysis. Reusability is a concept from computer programming and from ecology (“Reduce, reuse, recycle”) whereby we should do a task once and then be able to address the outputs of that process rather than repeating the work involved. We need tools to enable us to work with field recordings in a way that allows their further use as archival objects. The workflow following recording includes creating citable archival files with persistent identifiers. Any annotations that we make must be clearly related to the archival files. Such annotations, from a linguistic point of view, are transcriptions, interlinearized texts, and concordances, together with broader linguistic analysis.

With appropriate tools we can build good archival practice into our normal workflow for minimal extra effort. For example, we have tools

that permit text and audio to be linked so that we can create a corpus for an otherwise unrecorded language. These annotated recordings are of far greater use than they would be as audio alone. In the South Efate data there are currently some 20 hours of digitized field tapes that can be accessed via a textual concordance. This represents a significant part of the work done towards Thieberger's doctoral thesis (Thieberger, 2004b), and has involved preparing a data set for analysis in a manner that will be reusable and archivable.

Audio-text linkage for archival access

It is now relatively simple to provide indexes of media files in the form of time-aligned transcripts, and these are far more useful as archival material than is a tape with minimal metadata. Personal computers have been around since the mid-1980s, and it has been possible since the late-1980s to link digital audio and text by segmenting the sound into utterance-length chunks (Vallentine, 1992; Thieberger, 1994), but by the mid-1990s there was still no method for linking text and audio suitable for developing a media corpus for use in writing a dissertation. Similarly, there was no analytical tool with which one could access all field tapes via a textual representation. Segmentation of the audio data is not a sensible option as it is too time-consuming and detrimental to the very context of the utterances it is so important to preserve.

Having established a time-aligned transcript—one that had a chunk of text together with a start and end point in the audio file—there was in the late-1990s no simple way of then instantiating those links, that is, of hearing the audio associated with any given textual chunk. Thieberger wrote a working tool called *Audiamus* in HyperCard that allows access to the linked data instantly via a concordance point of entry to the data. A second version of *Audiamus* has been prepared using the cross-platform software *Runtime Revolution* (Thieberger, 2004a). The data in this corpus is citable by timecode, and, in the repository established with PARADISEC, it is locatable via a universal resource identifier (URI). It will be possible to provide streaming access to the audio in this dataset in the future. This means that a URL of the following kind will be able to be resolved: <<http://paradisec.org.au/NT1/NT1-98009-98009A.wav:57.4200-60.2238>>. That is, a data repository can serve selected time-chunks within media files linked to transcripts. By creating the data in a reusable form, it ensures that such steps are possible when the resources required are put to developing a streaming server. It is an important part of Thieberger's dissertation that all possible examples, all exemplary texts, and a representative version of the field tapes can be heard by the reader. In linguistics theses, the data is usually given as an example sentence, often with no indication of its sta-

tus or provenance. In other sciences, data is provided so that claims can be tested and results can be replicated. If we believe that linguistics employs the scientific method, then accessible presentation of the data is necessary for verification. Furthermore, the transcript provides a detailed annotation of the media that will also make the archival form of far greater value than it would be as an untranscribed tape.

Case Study 3: Establishing PARADISEC as a new digital repository

A major benefit of digital repositories is that they can be established without the kind of infrastructure associated with libraries in the past. PARADISEC is a collaborative cross-institutional and cross-disciplinary research resource established in 2003 by the Universities of Sydney, Melbourne, and Australian National University (joined in 2004 by University of New England), with support from the Australian Research Council's Linkage Infrastructure Equipment and Facilities scheme. The project is collaborative and cross-institutional, with our audio archiving unit in Sydney, our project manager in Melbourne, and our website and main data store hosted in Canberra at ANU and the Australian Partnership for Advanced Computing mass digital storage facility. We are also cross-disciplinary, with our chief investigators comprising leading linguists, musicologists, anthropologists, and computer scientists with a common concern to preserve and make accessible in digital form Australian researchers' field recordings of endangered languages and musics from the Asia-Pacific region.

Digitization of audiovisual media allows not only preservation, but also registration of these recordings. Registration in a well-managed repository can allow previously little-known, poorly described and inaccessible recording collections to become discoverable to researchers and communities worldwide if the repository contributes metadata to global networks for discovery and access. Furthermore, digitization facilitates standardized description of recording contents through time-coded transcripts, offering the prospect of more accurate citation than before possible, and, with e-publication and appropriate digital rights management, the prospect of direct citation of primary data within research publications.

Formats and standards

PARADISEC was initially established as a means for transferring existing field recordings to digital format, recognising that audio tapes themselves are becoming endangered as the media deteriorates and the machines for

reading them become obsolete. PARADISEC has been able to move quickly to digitize field recordings that were in a state of disrepair, and to place the data onto an accessible medium together with a metadata set adequate for describing the data. We adopt international archival standards and formats, and archive recordings using the Quadriga audio archiving system, as 24-bit 96 kilohertz Broadcast Wave Format files, which encapsulate the uncompressed linear PCM audio with summary metadata (European Broadcast Union (European Broadcast Union, 1997).

Establishing a good cataloguing (metadata) set for the collection has also required conforming to international standards (cf Hunter chapter XXX, this volume). In particular, metadata standards developed by the Open Language Archives Community (OLAC: <http://www.language-archives.org>) have been central to making our metadata conformant to the requirements of the Open Archives Initiative (OAI: <http://www.openarchives.org>). Adoption of these standards allows us to make endangered language material harvestable and discoverable via standard search mechanisms. Similar work in Europe by other digital endangered languages archives such as that of Dokumentation Bedrohter Sprachen (DOBES: <http://www.mpi.nl/DOBES>) and the Endangered Languages Archive (ELAR: <http://www.hrelp.org/archive>) may also prove useful in the near future, but was not accessible during the initial period of PARADISEC's operation. In 2003, PARADISEC joined with other digital archives with similar disciplinary orientations to form the Digital Endangered Languages and Musics Archives Network (DELAN: <http://www.delaman.org>), which aims to share expertise about digital archive management, to encourage researchers to adopt international best practice, and to explore the feasibility of distributed resource management across the network.

The PARADISEC repository has been constructed by specialists in the content area represented by the collection. The consortium of researchers who successfully applied for funds and then implemented the project are fieldworkers who have recognized the need for an archive to house their recordings. The team extends into information technology areas in modelling the data to be archived and in discussing with technologists the relationships between objects in the collection and how these relationships can be managed in the repository. Where specialist skills are required we have been able to buy them: for example, employing an audio preservation officer who is skilled at digitizing data from analog tape.

Archiving requires that data be in a form that allows it to be useful in future to researchers other than the depositor. To encourage researchers to produce well-formed data, we provide training in recording techniques and in description of the data recorded. We encourage fieldworkers to build these techniques into their ordinary practice so that the recordings they produce are optimal and are able to be lodged with PARADISEC for safekeeping and persistent identification. These citable media files can

then be referred to in all further analysis and can be used by others to verify the analysis (always subject to deposit and access conditions).

We have identified a number of collections of audiotapes that need to be processed, some in urgent need of treatment before they can be played due to their storage conditions, or their age, or both. We prioritize the digitization of the tapes based on a number of factors, including the immediate need that a community or a researcher may have for access to that material. For example, we knew of tapes made by an anthropologist in Papua New Guinea in the 1960s and were approached by a linguist wanting to work on the language from the same area. We accessioned the tapes into the collection and made copies available to the linguist who was then able to begin learning and analyzing the language. Similarly, recordings made by Stephen Wurm in the Reef Islands of the Solomon Islands in the 1960s will be taken back to that community by researchers working there in 2005.

In August 2004, little more than a year after commissioning the audio digitization facility, PARADISEC had identified over 3000 hours of unique field recordings (1466 individual items) that require digitization, representing over 150 languages from sixteen different countries in the Asia-Pacific region. More collections are emerging all the time; 600 hours of recordings have been digitized to date, provided to depositors both on access audio CD format and via password-protected access to our online repository, where we provide the archival master and another access version in MP3 format. All items in the queue, whether digitized or not, are discoverable via the OLAC search engine (<http://wave ldc.upenn.edu/olac/search.php?archive=paradisec.org.au>), and in 2004 we are in the process of developing a web interface to the collection.

Regional access to digital archival data

As digital ethnographic repositories are built, the metadata describing their contents can be made available via the internet, and providing appropriate international standards are adopted and shared, searches across repositories can make metadata discoverable by the relevant research or cultural communities wherever internet access is available. Digital data can be accessed remotely via the internet, and copies can be made and distributed with no loss of quality, in contrast with analog data which suffers a reduction in quality with each generation of copying. Where appropriate local facilities are available, a local repository mirroring the holdings of remote archives can be set up as a point of access and distribution of recordings, with far lower entry and infrastructure costs than a traditional archive or library.

Of course it is crucial to observe and enforce intellectual property and moral rights in this material, but a major difficulty in establishing appropriate arrangements for re-use of field recordings is that those

recorded typically live in small and perhaps remote locations. By establishing relationships with national or local cultural centres it becomes more feasible for ethnographic repositories to locate speakers or performers and their descendants. This will only be possible if there is sufficient metadata associated with a tape to allow us to locate its source—if not, for example, a speaker's name, then at least a geographical location is crucial.

Whilst there is a clear need to ensure safe long-term storage of ethnographic materials, it is fair to observe that access to digital data is quite uneven, especially in the region around Australia (Molnar & Meadows, 2001). In 2004, for example, the University of the South Pacific in Vanuatu had a local area network with comparatively slow access to the Internet, and the Vanuatu Kaljoral Senta had dial-up access only on individual computers. The Vanuatu National Library provided no Internet access and no public-access computers (Williams, 2002). Furthermore, as Williams notes:

... in all the Pacific Islands, the value of the archives and museums are not recognized as the repositories that hold and preserve the national and cultural heritage and identity of a country. These institutions are given minimal recurrent funding and are barely surviving. (Williams, 1998, 1)

Some support for regional archives has been provided by the Pacific Manuscripts Bureau (PAMBU: <http://rspas.anu.edu.au/pambu>), which has microfilmed many thousands of pages of unique material, but, while copies of these films are held in various locations in the region, microfilm readers are rare and are a difficult-to-maintain technology in this context where personal computers are actually more familiar and easier to find.

Individual computer users located anywhere in the Vanuatu archipelago with a phone line can, in theory, have dial-up access, but most villages have intermittent or non-existent electricity supply. With a basic wage of around AUD \$200 per month there is little scope for most Vanuatu residents to own computers. Nevertheless, computers are common in all government offices and local centres. For audio material, CD players are increasingly replacing cassette tape players and could be considered the current baseline technology with which to deliver archival audio material at the village level. PARADISEC has been actively working to provide appropriate support to the Vanuatu Kaljoral Senta by providing CD copies of relevant holdings in its collection, and by assisting with advice on backing up of data.

Conclusion

In this chapter we have presented several examples of the importance of digital technologies for the practice of ethnography, especially when working with small and endangered cultures and languages. The interaction

between field recording (as in the Belyuen and South Efate examples) and the long-term storage of its products (as in the PARADISEC example) is crucial to both citability of data, and access to the data by speakers and their descendants. Indeed, we see that there is a crucial interaction between recordings made in the field with sufficient metadata for their discovery, and their subsequent location in a suitable repository. Later, this is followed by their reintroduction to their community because they are located (using web-based search mechanisms) by those recorded or their descendants.

The humanities disciplines are currently going through a major paradigm shift in methodologies and relationships with primary data. The creation of new communicative fields through technological developments has generated an increasing emphasis on teamwork and collaboration between humanities researchers, cultural heritage communities, and technological specialists who can help us to realize our aspirations and envision new ones. Humanities scholars and community bodies need to engage in the fora that decide the strategic directions of research policy and funding, to ensure that technological development understands and embraces our needs. We also need to make sure that present and future generations of humanities researchers are educated to understand and embrace a role in creating and managing resources that will contribute to the future of both our disciplines and the cultural heritage communities with which we interact.

As fieldworkers, we welcome the opportunity to use digital representations of data to facilitate access by those recorded and their descendants in a way that was not possible using analog recordings (Barwick, 2004). For this reason, we wish to ensure that our recordings are archived using the best possible methods, with sufficient descriptive material (metadata) to provide for their discovery, on media that allow the data to be migrated over time, and in sustainable institutional contexts. From a scholarly perspective, we wish to ensure that the digital data we create as part of our intellectual endeavor as ethnographers is reusable, both by ourselves and by others, first because any claims that we make based on that data must be replicable and provable by others, and second, because the effort and expense of creating a digital representation of the data should not be duplicated, but rather serve as a foundation for others to build upon.

Despite the apparent gap between the consumerism of new technologies and the (cash) poverty of those recorded, we argue that, as professional ethnographers, we need to use the best current tools to do the work expected of us. As Anthony Seeger (2004, p.1) comments: “The digital future will only be as rich as the materials supplied to the institutions distributing them. Many details will need to be worked out over time, but they all begin with the researcher.”

Notes

¹ Australian Broadcasting Corporation Radio Archives Disc NAT2-3, Australian Broadcasting Corporation archives tape number 72/10/543 with commentary; 72/10/544 without commentary; AIATSIS sound archives tape A2915.

References

- Azoulay, L. (1900). 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. (2003). The Endangered Cultures Research Group's digitisation project: using digital audio for musicological research. In C. Cole & H. Craig (Eds.), *Computing Arts 2001: Digital resources for research in the humanities* (pp. 147-159). University of Sydney: Research Institute for Humanities and Social Sciences, University of Sydney, in association with the Australian Academy of the Humanities.
- Barwick, L. (2004). Turning it all upside down . . . Imagining a distributed digital audiovisual archive. *Literary and Linguistic Computing*, 19(3), 253-263.
- Bird, S., & Simons, G. (2003). Seven dimensions of portability for language documentation and description. *Language*, 79, 557-582.
- Boston, G. (2003, November 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*. Retrieved 14 August, 2004, from http://portal.unesco.org/ci/ev.php?URL_ID=13437&URL_DO=DO_TOPIC&URL_SECTION=201&reload=1073385878&PHPSESSID=019c5625ccdffa7095faf0aade5c128
- Bradley, K. (2003). *Critical choices, critical decisions: Sound archiving and changing technology*. Retrieved 14 August, 2004, from <http://conferences.arts.usyd.edu.au/viewabstract.php?id=57&cf=2>
- Council on Library and Information Resources (CLIR). (2001). *Folk Heritage Collections in Crisis*. Retrieved 14 August, 2004, from <http://www.clir.org/pubs/reports/pub96/contents.html>
- DigiCULT Consortium. (2004). *DigiCULT: Technology challenges for digital culture*. Retrieved 14 August 2004, 2004, from <http://www.digicult.info>
- Duranti, A. (1997). *Linguistic Anthropology*. New York: Cambridge University Press.
- Elkin, A. P. (1953). *Tribal Music of Australia* [LP record]. Washington, D.C.: Folkways Records FE4439.
- Elkin, A. P. (1957). *Arnhem Land Volume 1: Authentic Aboriginal Songs and Dances* [LP record]. Sydney: EMI (Australia) OALP7504.
- Elkin, A. P., & Jones, T. (1958). *Arnhem Land Music*. Sydney: University of Sydney.
- European Broadcast Union. (1997). *Specification of the Broadcast Wave Format: A format for audio data files in broadcasting*. Retrieved 14 November, 2004, from http://www.ebu.ch/trev_t3285.pdf
- Feld, S. (1994). From schizophonia to schismogenesis: on the discourses and commodification practices of "world music" and "world beat". In C. Keil & S. Feld (Eds.), *Music Grooves* (pp. 257-289). Chicago: Chicago University Press.

- Gorski, P. (2001). *Multicultural education and the digital divide*. Retrieved 14 August, 2004, from <http://www.mhhe.com/socscience/education/multi/philosophy/4divide.html>
- Himmelman, N. P. (1998). Documentary and descriptive linguistics. *Linguistics*, 36, 161-195.
- Humanities Advanced Technology and Information Institute (HATII), & 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*. Retrieved 14 August, 2004, from <http://www.nyu.edu/its/humanities/ninchguide/>
- Maffi, L. (Ed.). (2001). *On biocultural diversity: Linking language, knowledge and the environment*. Washington, D.C.: Smithsonian Institution.
- Marett, A. (2003). *Sound recordings as maruy among the Aborigines of the Daly region of north-west Australia*. Retrieved 14 August, 2004, from <http://conferences.arts.usyd.edu.au/viewabstract.php?id=40&cf=2>
- Marett, A., Barwick, L., & Ford, L. (2001). Rak Badjalarr: Wangga songs by Bobby Lane, Northern Australia [audio compact disc]. Canberra: Aboriginal Studies Press.
- Molnar, H., & Meadows, M. (2001). *Songlines to satellites: Indigenous communication in Australia, the South Pacific and Canada*. Annandale, NSW and Wellington: Pluto Press and Huia Publishers.
- Mountford, C. P. (1949). American-Australian Scientific Expedition to Arnhem Land 1948 [9 78rpm records]. Sydney: Australian Broadcasting Commission.
- National Library of Australia (NLA). (2000). *National Library of Australia Digitisation Policy 2000-2004*. Retrieved 14 August, 2004, from <http://www.nla.gov.au/policy/digitisation.html>
- Pitoëff, P. (1993). Du cylindre au disque compact. Les archives sonores du Musée de l'Homme (Paris). *Annuario degli Archivi di Etnomusicologia dell'Accademia Nazionale di Santa Cecilia*, 1, 143-149.
- Schüller, D. (2004). Safeguarding the Documentary Heritage of Cultural and Linguistic Diversity. *Language Archive Newsletter*, 1(3), 9-10.
- Seeger, A. (2004). *SIMS 2004 New Technology Handout*. Unpublished manuscript, Melbourne.
- Thieberger, N. (1994). *Australia's Indigenous languages information stacks*. Retrieved 14 August, 2004, from [oai:aseda.aiatsis.gov.au:0531](http://oai.aseda.aiatsis.gov.au:0531)
- Thieberger, N. (2004a). Audiamus (Version 2) [computer program]. Melbourne.
- Thieberger, N. (2004b). *Topics in the Grammar and Documentation of South Efate, an Oceanic language of Central Vanuatu*. Unpublished PhD thesis, University of Melbourne, Melbourne, Australia.
- Vallentine, J. R. (1992). *Rook: A system for authoring descriptive grammars in HyperCard for the Apple Macintosh computer*. Retrieved 14 August, 2004, from [oai:aseda.aiatsis.gov.au:0499](http://oai.aseda.aiatsis.gov.au:0499)
- Williams, E. B. (1998). *Information needs in the Pacific islands: Needs assessment for library, archive, audiovisual collection and ICT development in the Pacific islands*. Apia, Samoa: UNESCO Office for the Pacific States.
- Williams, E. B. (2002). *Digital community services: Pacific libraries and archives, future prospects and responsibilities*. Suva: UNESCO.
- Woodbury, A. C. (2003). Defining documentary linguistics. In P. K. Austin (Ed.), *Language documentation and description* (Vol. 1, pp. 35-51). London: Hans Rausing Endangered Languages Project.
- Wright, R., & Williams, A. (2001). *PRESTO-W2-BBC-001218 Archive preservation and exploitation requirements*. Retrieved 14 August, 2004, from <http://presto.joanneum.ac.at/Public/D2.pdf>