This collection of papers arises from a conference of the same name, held at the University of Sydney in December 2006. Funded jointly by the Australian Partnership for Sustainable Repositories and the Ethnographic e-Research project, and hosted by PARADISEC (the Pacific and Regional Archive for Digital Sources in Endangered Cultures), the conference brought together researchers, technologists and information specialists with a common interest in sharing the results of their recent work at the interface between fieldwork and digital repositories. The timeliness of this event is indicated by the considerable national and international interest generated by the conference, which attracted presenters and attendees from the USA, Canada, UK, Norway and Russia, as well as staff and students of Australian Universities, national cultural institutions and Indigenous cultural centres. A selection of papers offered to the conference were submitted for anonymous peer review to form this volume.

The APSR-funded FIDAS (Fieldwork Data Sustainability) project took place in 2006, and aimed to assist the workflow of researchers who collect digital data during fieldwork by analysing field methodologies and providing tools and resources to support the collection and long-term sustainability of data and metadata created in the field. While academic fieldworkers have been quick to take advantage of digital technologies to enable them to collect and organise their data, standards and workflows are only now beginning to emerge to assist researchers to submit their data for archiving and access. In many disciplines, including my own, ethnomusicology, fieldworkers have often collected data in a rather ad-hoc way during fieldwork. After all, as pointed out by several papers in this volume, the goal of our activities is to support our academic research, not to produce an exemplary data set. Even equipped with the best training, equipment and intentions, fieldworkers very often (if not always) find their data collection activities constrained in unpredictable ways by
contingencies that arise in the field, which may lead to patchy and highly variable metadata quality at the time of submission of the data to a digital repository. It can be very difficult or even impossible to reconstruct some of this information at a later date, yet these resources are often unique and unrepeatable records of highly significant events collected at considerable expense of researcher time, effort and resources. From the repository perspective, lack of metadata (including preservation metadata) can have serious implications not only for ingestion into a repository, but also for subsequent archival management and dissemination of archival information. With the increasing use of digital methods of data capture and organisation, the backup and curation services offered by archives and institutional repositories are needed more than ever, for few researchers have the time or resources to implement an ongoing digital sustainability programme on their personal collection.

The Ethnographic e-Research project (EthnoER), funded by the Australian Research Council in 2005-6 as part of their E-Research Special Research Initiatives Programme, is a consortium of ethnographic researchers (linguists, musicologists and anthropologists) wishing to explore the possibilities for collaborative research based on large digital media data-sets. The project aims to promote standards for description, media formats, rights management and annotation to promote interoperability between collections and re-usability of ethnographic research data by researchers as well as cultural heritage communities. Testbed data was drawn from six projects based in Australian universities: PARADISEC; the Aboriginal Child Language Acquisition project (ACLA); the Auslan Corpus Project; the Melbourne University Reciprocals Project; the Waima'a (East Timor) documentation project; and the National Recording Project for Indigenous Performance in Australia). Collaborating international researchers came from the Alaskan Native Language Center (University of Alaska at Fairbanks), the Hans Rausing Endangered Language Project (School of Oriental and African Studies, University of London), the Archive of the Indigenous Languages of Latin America (University of Texas at Austin), and the Galiwin’ku Indigenous Knowledge Centre. Development support was provided by technologists from the University of Queensland, the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Australian Partnership for Advanced Computing (APAC). In addition to presenting some of the results produced by the EthnoER project itself (for example, the papers by Schroeter, Thieberger and Stephens), this collection includes reports by
project partners on aspects of their own research practice that are significant for repository managers and archivists interested in modelling and designing workflows to facilitate deposit of research data, as well as understanding potential demand and limitations on access to data.

The volume has been divided into four parts, roughly reflecting the workflow from fieldwork to repository and beyond, and includes contributions by botanists, biologists, and librarians as well as linguists.

In Part I, we hear from sign language specialists Trevor Johnston and Adam Schembri about the specialised nature of their work on Australian Sign Language (Auslan). With a corpus dependent on video data, their work has specific requirements regarding conduct of fieldwork, annotation procedures, sustainability, and ethics that have significant correlates in other ethnographic collections. The next two papers, by Honeyman and Robinson, two early-career researchers who have adopted innovative digital methods in linguistic fieldwork in remote locations, provide important practically oriented accounts of the constraints suffered by fieldworkers in such situations. Strikingly similar constraints on creation, management and access to botanical information apply to the online database on plants of Papua New Guinea, created in an international collaboration between Australian and Papua New Guinean researchers, as recounted in the following paper by Conn and Dimas.

Part II addresses the problematic concept of best practice, from the varying perspectives of researchers, archivists, community collaborators and disciplinary champions. Bowden and Hajek point yet again to the competing constraints that researchers face between creating and sustaining digital data on the one hand, and on the other, providing appropriate access to research results within resource-poor cultural heritage communities. David Nathan suggests that the concept of 'sustainability' may be useful in guiding archives and researchers in finding a balance between such competing constraints and demands. Berez and Holton provide a detailed account of the successes and failures of one project to provide community training to members of an Alaskan language community to allow them to develop expertise in digital language documentation and management, concluding that in this context too there is a significant cline between best practice and convenience. Similar issues affected the evolution of a research community education resource described by Boynton, Moran, Aristar and Aristar-Dry, resulting in the development of a more nuanced model of best practice in language
documentation, to recognise the constraints faced by both creators and users of linguistic data.

Part III presents results of a number of projects currently under development to provide better access to digital research data. Under the umbrella of the EthnoER project, the EOPAS (Schroeter & Thieberger) and Annodex (Stephens) tools have each been working to build on existing tools and annotation standards widely used by linguists to create online access and collaboration platforms. Musgrave discusses issues for the field linguist in using XML to facilitate the archiving of annotation data, while Henwood and his collaborators discuss the importance of standards and interoperability in current development of an online interface to aggregated botanical collections, suggesting that such new views of existing data provide scope for exciting global collaborations.

In Part IV, we move to two case studies of projects whose very conception depends on the accessibility of well-structured online repositories. Næss is a Norwegian linguist about to embark on a study of a language family from the Solomon Islands, using data housed in Australia, in the PARADISEC repository; she hopes that access to the primary data, from the collection of the late Stephen A. Wurm, will clarify his theories about the genetic status of this family, which have recently come under question. Finally, Ross Coleman uses the present volume, which will be available online through the University of Sydney's eScholarship repository as well as in conventional print form, as an example of emerging hybrid forms of electronic publication. He suggests that good arguments can be made that criteria for research quality assessment can be applied to well-structured published research databases as well as to conventional publications.

Endnotes
1 The full papers were reviewed anonymously by at least two reviewers selected from an international panel of experts (double blind process). I would like to thank the reviewers for giving so generously of their time to ensure the quality of content of this volume.
2 See web links to these projects and institutions in the reference list.
References


