Introduction

Nick Thieberger

This volume presents papers focused around the theme of creating and reusing digital research data, particularly from a Humanities perspective. Whatever the disciplinary focus, modes of enquiry and scholarship in the humanities often rely on a body of shared knowledge that becomes the object of fine-grained reference, commentary and analysis. Whether the object of enquiry is a literary or linguistic corpus, a musical repertory, a collection of historical or philosophical texts, an archaeological site or the creative output of an artist or period, scholars need to be able to communicate their insights to each other by referring to the work of other scholars as well as the target body of data. In the past, analog research data has been relatively robust—paper manuscripts and notebooks can still be consulted centuries after their creation. Dynamic media recordings (audio, film, and video) have been less robust, and earlier media (such as wax cylinders or wire recordings) are no longer playable without extremely specialised equipment available in only a few places in the world. Digital data is even less robust but provides many more possibilities for reuse than does analog data. How can we ensure that the records we create will survive and will be reusable? As our libraries and other scholarly repositories increasingly share information and data online, how can we take advantage of the possibilities offered by these infrastructures and collections to create new ways of sharing and reusing the resources?

A key to sustainable research data lies in the creation of repositories committed to housing it and curating it in the long term. In chapter one, **Sebastian Drude** and **Paul Trilsbeek** discuss a large-scale linguistic archiving project based at the Max-Planck-Institute for Psycholinguistics in Nijmegen. Initially planned to support a program funding research on endangered languages, the archive has now extended to a number of regional centres around the world.

A common issue for research uses of primary data centres around the degree to which that data is made accessible in its native form, and how much it is packaged in databases that make the primary data available only via a fairly complicated interface. Further, without an established set of standard formats for data created by practitioners

within a disciplinary community it becomes almost impossible to create higher level processes that deal with that data. **Alexander Borko**wski and A**ndrea Schalley** observe that language archives store material that should be used for research purposes, and that databases created to allow typological work can themselves be unsustainable. They discuss a new typological database (TYTO) based on user-oriented design principles and Semantic Web technologies in order to provide sustainable ways to integrate, analyse, and access cross-linguistic data.

The primary data that is created during fieldwork can have multiple uses. Audio and video recordings can serve as linguistic examples, but have many other potential uses, depending on the content and on the quality of the recording. Video can be posted on web-based hosting sites for access and the same video can be reused as a stimulus for discussions in several languages. **Anthony Jukes** describes a process in which videos of common processes (for example, making palm sugar) are a focus for discussion and then are reused as elicitation stimulus for neighbouring languages, resulting in various audio tracks for the same video.

The use of video for documenting Indigenous sign is the topic of the next chapter by **Jennifer Green**, **Gail Woods** and **Ben Foley**. They argue that involving signers in the process of recording their own performance and deciding how to represent their signing practice on the web is critical to the long-term availability of the recordings, and that the presence of these signs on the internet is a major motivation for signers to participate in their recording.

Having performers agree to recordings being made requires careful negotiation, and the further step of archiving the resulting media, especially archiving in another country, is discussed in the chapter by **Catherine Ingram, Wu Meifang, Wu Pinxian, Wu Xuegui,** and **Wu Zhicheng**. The conditions under which recordings are made available via an archive are not easily communicated to people with no experience of the internet, and the potential problems associated with publication of music are also discussed in this chapter. The authors suggest that archives could present short guides to their practice in languages and media appropriate to the potential depositors so that communities can make their own informed decisions about access.

A critical problem for long-tem storage of research recordings arises from the fact that a recording fixes a transient event in time, perhaps giving an erroneous

impression of the true variability associated with the original context. As **Stephen Morey** points out, it needs to be made clear that a recording is just one performance that happened to be recorded. He explores the implications of archiving records for the societies from which this material originates. A benefit of archiving is that (otherwise disrupted) cultural transmission of traditional practices may be supported by a return of material from archives.

Critical to future use of digital collections is the ability for users to enrich interpretations via annotation services. Such services are becoming more familiar thanks to the semantic web and **Jane Hunter** and **Chih-hao Yu** outline their annotation system for three dimensional objects that builds on the Open Annotation Collaboration (OAC) data model.

Establishing a repository of research data is only a first step, as **Kerry Kilner** and **Roger Osborne** discuss with reference to AustLit, one of the oldest digital scholarship projects in Australia, with their critical perspective on the sustainability of research data. This chapter gives a frank overview of the history of the project, and of the ways in which it has been reconfigured over time, to become the AustLit Digital Research Commons.

In the next chapter, **Toby Burrows** characterises humanities research data as being made up either of entities or of annotations, both of which have to be stored and added to as scholarly work proceeds. At the same time there is a problem of heterogeneity of data formats that defies automated linking of research data. He suggests the Linked Open Data framework would allow links to be made between objects, each with its own. Examples are provided from research on Medieval manuscripts.

Simon Musgrave and **John Hajek** consider how to provide incentives for the creation of sustainable data, and suggest that the approach taken by the 'Excellence in Research in Australia' (ERA) is likely to do exactly the opposite, with its emphasis on publication in traditional formats. By ignoring trends in Open Access publishing for research outputs and in the use of institutional repositories for primary data, the ERA is failing to provide a suitable model for 21st century research practices. The authors suggest that embedding data in publications to assist the verification of analyses should

now be the norm, as should support for Open Access publication as a means of getting research out to a wider audience.

Each paper in this volume was peer reviewed by at least two readers and revisions were entered in a timely fashion which allowed us to produce the book in advance of the conference of the same name in Melbourne in December 2011. In accord with the theme of the conference, the book is available as print-on-demand, and each chapter is downloadable from the University of Sydney's open access digital repository (http://hdl.handle.net/2123/7890).

Acknowledgments

Thanks to each of the authors who acted as reviewers of papers in this volume. Thanks also to the following for help in reviewing papers and otherwise assisting in the production of this volume: Andrea Berez, Steven Bird, Lauren Gawne, John Henderson, Gary Holton, Sebastian Nordhoff, Desmond Schmidt, Sally Treloyn, Myf Turpin, and Aidan Wilson.

Funding for the conference and the publication of this volume has come from the School of Languages and Linguistics and the Faculty of Arts at the University of Melbourne.