Known unknowns: Towards a strategy for preserving digital research data

Owen O’Neill¹, Andy Tseng²
¹University of Melbourne, Australia, ooneill@unimelb.edu.au
²University of Melbourne, Australia, tsenga@unimelb.edu.au

INTRODUCTION
With ever-growing collections of digital research data, researchers and Universities are facing an unprecedented set of digital preservation challenges as part of the wider so-called “data deluge” [1]. Beyond the benefits of preserving data for posterity, reference and reuse, University and funding body policies also stipulate preservation conditions for funded research. However, researchers often lack the resources to address long-term preservation requirements, and a concerted effort to develop appropriate support and infrastructure is required.

Within this context, the University of Melbourne is currently developing a research data preservation and archiving strategy to define and address these challenges. Due for completion in 2013, the development of the strategy has actively engaged a range of key stakeholders including researchers, the Library, IT Services and the Melbourne Research Office.

This presentation provides an overview of the key challenges that have been identified through a series of stakeholder consultations, trials and research. It aims to provoke and contribute to conversations about supporting long-term preservation of research data.

TOWARDS A DIGITAL PRESERVATION STRATEGY FOR RESEARCH DATA
Initially focused on research data of enduring value, the strategy is being progressively developed tested and refined during 2012 through focused trials with researchers and research data custodians, and with input from key stakeholders. The strategy has the following broad objectives:

- Protect and facilitate improved access to University research data of enduring value
- Support cross-disciplinary collaboration
- Meet legal and ethical requirements as defined by University policy, research funding bodies and Government
- Utilise international standards and best practice for digital archiving and preservation of research data.
- Define and clarify recommended workflows, infrastructure and roles and responsibilities for enabling archiving and preservation of research data.

This work will support the University’s Scholarly Information Future strategy [2] by developing a framework for preservation that facilitates cross-disciplinary collaboration and robust scholarly information practices. These practices will also enable wider dissemination of the University’s scholarship.

The strategy is being actively tested and refined during its development through focused trials with researchers and research data collection custodians. These trials are concentrated on a number of perceived barriers to long term data preservation including preparing data for preservation, managing access to sensitive data and strategies long term maintenance.

NEXT STEPS
The strategy is due to be completed by mid-2013 after extensive consultations and research, and endorsement by a high level steering group. It is expected that this strategy will inform the required changes to the current processes and approaches used to maintain digital research data at the University.

REFERENCES
ABOUT THE AUTHOR(S)

Owen O’Neill is a Data Archiving Analyst currently working on a project to develop a strategy for preserving and archiving digital research data at the University of Melbourne. He is based in the University’s Information Technology Services (ITS) Research department. Owen also recently coordinated a program of eResearch projects at the University, working across a range of research domains and centres. Before working at the University of Melbourne, Owen was involved in a number on technical standards and interoperability projects for e-learning content and systems in Australia and Europe.

Dr Andy Tseng is a Data Infrastructure Architect based at the University of Melbourne Information Technology Services (ITS) Research Department. Andy engages with research groups from different disciplines across the University and affiliates with the aim of assisting researchers to develop customised solutions and services to curate and preserve their data collections more effectively. He also helps researchers design formal data models and implement infrastructure for developing tools, platforms to capture, and integrate and share their research data. Andy is also involved in facilitating university plans to describe, catalogue, and preserve research datasets via national initiatives such as “Australian National Data Service” and collaborating with colleagues from around the University for formulating and deploying technical solutions to address data-oriented problems.