

# Oxford e-Research Conference 2008

Topic strand:

**Survey research and in-depth interviews focused on the attitudes and practices of researchers**

Paper title:

**“We have just entered the Wright Brothers phase of the digital era”: data management and practice in the eResearch era**

Submitted by:

**Belinda Weaver**

Manager, eResearch Infrastructure  
The University of Queensland Library  
b.weaver@library.uq.edu.au

**Margaret Henty**

National Services Program Coordinator,  
Australian Partnership for Sustainable Repositories  
Margaret.Henty@apsr.edu.au

## Abstract

Universities around the world are looking to improve the infrastructure that supports their eResearch. However, infrastructure means more than just high quality ICT. While supercomputing facilities and high speed internet connections are crucial components for successful eResearch, also important is the availability of trained technical staff to support researchers in both complex applications and networking. Sound data management practices are also vital to ensure that all research project data is both readily accessible to researchers on demand and safeguarded for the long term, so that the data can be revisited by the original research team, or eventually made available to other researchers.

Academic culture, institutional capabilities and individual skills are critical components if we are to move towards a more effective research system where the potential for eResearch can be optimised.

In the second half of 2007, The University of Queensland conducted a survey of the data management practices of all researchers. The survey was open to academics and post-graduate students. It sought feedback on current data management practices, including the use of formal plans, and the use of backup systems for safeguarding research datasets. The survey also aimed to identify gaps in researchers' knowledge of training and support systems currently available to them.

Responses to the survey were designed to feed into future planning for eResearch infrastructure provision at the University via an eResearch Working Party which collaboratively drafted the original survey instrument. The University of Queensland followed the survey with a series of focus groups. Groups were made up of individuals who had indicated in the survey their willingness to participate in an ongoing eResearch reference group. Survey responses were probed in much greater

detail, and attendees were asked for ideas on formulating a University-wide policy on data management.

The survey was subsequently replicated by two other Australian universities with similar planning needs. The results have given each university valuable information on which to base future decisions about local eResearch infrastructure. Universities are also using the results to identify the most important training or technical support needs. The use of common questions and survey gathering software across all three institutions enabled the aggregation of results, which were then fed into national eResearch planning processes.

The survey included a wide range of questions covering data management. Questions covered

- the types of data generated in research
- the methods of data generation, e.g. from instruments, fieldwork, etc.
- the size of data sets
- how data is used and manipulated
- the types of software used in analysis
- researchers' willingness to share data
- responsibility for data management and storage
- data ownership
- the use of high performance computing

Comments were encouraged – and to good effect, as many useful remarks were received.

So what have we discovered? There is a remarkable consistency in the results across all three universities. This suggests not only that researchers share many common practices and attitudes to their research, but that the problems or hurdles they encounter are also common. Overall, the surveys showed that researchers are more focussed on research outcomes than on good research management practices, especially when it comes to the long term stewardship of data. The heavy reliance on back up media such as USB drives or portable hard drives which, while convenient, are not reliable enough for archival storage, is a concern. However, researchers were eager to find better ways to manage their research data, with many expressing a desire for more training and direction in data management practices.

More promisingly, the survey revealed that many researchers would be willing to share raw data with other researchers provided they were provided with an easy mechanism to enable this. The applications researchers used to manipulate data were very varied, and some complaints were made about vital applications being unavailable or rationed because of cost. The use of so many different proprietary software applications, all with their own proprietary formats, has implications for long term data storage and curation. Some confusion about responsibilities for data management was reported across all disciplines.

Rather than blaming researchers for the negatives in the current state of affairs, we need to examine the level and type of support that they get from their institutions. This does not mean just the technical support, but also the training needed and 'rewards' forthcoming for doing the job of research data creation and management properly. Currently there is no reward when data management is well done, and no sanctions against those who fail to do it well, or at all.

We will present the results of our survey and the conclusions that can be drawn from it for future institutional eResearch planning.