Jun 21st, 3:30 PM - 4:30 PM

Rsquared: researching the researchers. A study into how the researchers at the University of New South Wales use and share research data

Howard Amos
University of New South Wales, h.amos@unsw.edu.au

Maude Frances
University of New South Wales, m.frances@unsw.edu.au

Tom Ruthven
University of New South Wales, t.ruthven@unsw.edu.au

RSQUARED: RESEARCHING THE RESEARCHERS: A study into how researchers at the University of New South Wales use and share research data

Howard Amos
Deputy University Librarian. University of New South Wales, Sydney, Australia
h.amos@unsw.edu.au

Maude Frances
Manager, Library Repository Services. University of New South Wales, Sydney, Australia
m.frances@unsw.edu.au

Tom Ruthven
Manager, Digital Library Innovation and Development. University of New South Wales, Sydney, Australia
t.ruthven@unsw.edu.au

Abstract
Research data management in libraries in the past focused on the published output at the end of the research cycle. We are now of the age of data driven scholarship. E-scholarship (not just e-science) is predicated on data organisation, management, use and reuse. We now need to become more involved with the management (describing and making accessible) of research data itself. To support our researchers we must understand what services are needed to support e-research.

This paper presents a research study of data usage, creation and sharing within different research communities at UNSW. The study identifies emerging data usage and management needs within the e-research life cycle of diverse research communities. Comparison is made with the outcomes of other studies that have examined e-researcher work practices in relation to their data. The paper examines the findings to understand what role researchers see libraries having, and discusses the development of a framework that libraries can use to support the curation and management of data and the development of tools and library support services that can be used across disciplines.

The study addressed the following:
What are the existing data use, storage and sharing practices of academics at UNSW?
What are the differences and similarities between disciplines and fields of research?
What are the differences between project types (e.g. multidisciplinary, cross institutional)?

Using a mixed-methods research design, data from focus groups was used to construct an online survey followed by interviews with selected survey participants to extend and provide in-depth understandings of data from the survey.

The study was carried out by the Library Repository Services unit (LRS) which was established in 2007. Initially charged with designing and building the UNSW institutional repository, LRS now leads the Library's research data management services and contributes to the University's e-research coordinating group. LRS services support the library as a partner and collaborator in the e-research space where the description, management, use and reuse of data in all its forms (primary, secondary, interpreted, analysed and published) is provided.

Introduction
The University of New South Wales Library established the Library Repository Service (LRS) in 2007. Initially the focus was on building and managing an institutional repository as part of the ARROW partnership (www.arrow.edu.au) and in response to initiatives from the Australian Federal Government, which have resulted in new data gathering requirements to report on research activity. As the team became established it became apparent that the services the LRS team provided needed to extend to data librarianship to directly support researchers.

In the past research data management in libraries has focused on the published output end of the data spectrum, yet repositories are about facilitating the use and reuse of material in many contexts [Amos & Ruthven 2008]. We now need to become more involved in the management, describing and accessibility of primary research data.

The literature at the time [Gibbs 2007, Gold 2007, Heery & Powell 2006] saw institutional repositories as logical places for storage of research datasets. As data stores, work was being carried out to assess the role of institutional repositories in management of research data assets and the roles of libraries and librarians in the provision of data services. The UNSW repository team also became involved in this work in support of research projects, establishing research infrastructure projects in 2008 and 2009.[1]

LRS services are not IT infrastructure services. We don't provide raw storage facilities, back up, archiving and preservation services (except as for curation – we may assist with describing for preservation). We see our role as assisting with schema selection (if called upon), developing cross walks to standard schema and quality assurance on marking up data and describing for discovery and reuse.

So what are the services required from the UNSW "digital library" or more specifically from the Library Repository Service that we have established? The US Interagency Working Group for Digital Data (IWGDD) formed under the auspices of the NSTC Committee for Science found that new specialisations in data tools, infrastructures, sciences and management are emerging to support the use and reuse of research data [IWGDD 2009]. How does LRS support the establishment/process of "library as partner/collaborative research space" – i.e. where is the description/management/use and reuse of data in all forms provided, from primary, secondary, interpreted/analysed to published?

How do we mitigate the risk that if we build it they may not come? Do we restrict ourselves to services for specific research projects – but will this allow us to build "library as partner/collaborator"? Are there commonalities that may be exploited to provide cross disciplinary and inter-project services? Economic realities dictate that LRS services must be generic enough to encompass research repository services across a spectrum of research projects across all disciplines.

It seemed logical to ask the researchers what their needs are [Gibbs 2009, Granada et al. 2009]. What emerging needs or usage trends of higher degree by research candidates and early career researchers will we need to support? What role do they see us having in their e-research data life? What metadata are our researchers using, and what criteria do they consider important to select metadata schemas. What are the existing data use and sharing practices of researchers at UNSW? What are the differences and similarities between disciplines and fields of research?

To answer these questions we conducted this study, the aim of which was to gain an understanding of changing needs in data use, creation and sharing within different research communities at UNSW. If we are to help researchers identify metadata schemas we need to have a robust understanding of the scope of what they need to describe. Findings from the

---

[1] **Social Science Research Material,** a database of research material collected in the National Centre in HIV Social Research at UNSW migrated to a repository to maximise interoperability with a national social science data store, and includes research proposals, project descriptions and research instruments from qualitative, quantitative and mixed-method studies. See Frances & Croucher [2009].

Human Resources for Health Knowledge Hub Electronic Gateway, a joint initiative of LRS and the School of Public Health and Community Medicine at UNSW, is a repository for document and other file storage, with a wiki and a blog that enables discussion on developing resources.

Membranes Research Environment , jointly developed by LRS and the UNESCO Centre for Membrane Science and Technology at UNSW, combines a repository with wiki database technology to capture publications and data from all stages of the membrane materials research lifecycle. See Carusi & Reimer, [2010]; Cox et al., [2009]

study will assist us to develop the library framework, tools and services relating to data curation and management support.

Data use and sharing: a review of the literature

Recent studies and reports on data use and sharing tend to focus on the data lifecycle and, naturally, are data-centric rather than based on research practices. Some of the studies and reports touch on the practices and workflows of researchers. A few have examined particular aspects in-depth. Yet, significantly, the question posed by Lyon in 2007 remains valid: "How much do we know about individual researchers' practice regarding data creation and capture, data management, curation, preservation, data publication and reuse? The answer is not enough" [Lyon 2007].

The data-centric studies and reports use one of two data lifecycle models. The Digital Curation Centre (DCC http://www.dcc.ac.uk/) Curation Lifecycle Model that places data in the centre surrounded by aspects of curation and preservation, and the IWGDD model that centres on Culture, Use & Reuse and Strategy surrounded by aspects of technical requirements, skills, organisations and entities, and policy. This approach has informed examinations of large research projects [Wallis et al. 2008, Pepe et al. 2009]

Using these data lifecycle models, the roles and responsibilities of people and organisations have been examined. Much of this work comments on organisational roles [Lyon 2007] or roles for a particular aspect [Gold 2007]. Specific parts of the lifecycle have also been examined and well documented such as best practice principles for metadata creation to enhance discoverability [Dinkelmann et al. 2009].

Significant sectoral and national reports have provided directions on policy and infrastructure requirements for research data. The *Harnessing the power of digital data for science and society* report [IWGDD 2009] made recommendations on infrastructure and national services for the US Government. The recommendations of the UK *JISC Strategy 2010-2012 Initial Draft* [JISC 2009] focused on investment in research, business systems, and national services. The Australian federal government's *Final Investment Plan for Research Capability for the National Collaborative Research Infrastructure Strategy* [NCRIS 2006] recommended the establishment of the Australian National Data Service (ANDS) with goals in data management, curation and collecting research resources. A JISC-funded guide [Green 2009] provided well documented advice to institutions on digital data repository planning.

Broad surveys have been conducted to give high level information about frequency of practices by researchers in data use. [Markauskaite et al. 2009, Maron 2008]. Other studies and reports touch on the practices and workflows of researchers. The IWGDD report had a section on Personal Digital Collections [IWGDD 2009]. The CLADDIER Project in the UK found that data producers have certain requirements for citation [Pepler 2008]. The DISC-UK DataShare project [Gibbs 2007] discovered that motivation for management of data is a better driver for researchers than data sharing.

Specific aspects of data use and sharing practices were examined in a study of chemistry at a single institution [Polydoratou 2007] including the types of data researchers produce, formats saved, the metadata considered important, the perceived advantages of sharing their data and how access to data can be achieved, though it did not place the issues within a research workflow. The DCC SCARP Project (http://www.dcc.ac.uk/projects/scarp) has engaged with particular disciplines through a series of immersive case studies though the focus of the project is to better understand how digital curation capacity and support structures may be developed effectively [Lyon et al. 2010]. Attitudes and use by researchers of specific data stores have been examined and the value of covering the complete lifecycle of social science data, from survey design, via data collection to the publication of analyses was confirmed [Mochmann 2009].

Leaders in various fields have contributed their perspectives on data use and reuse, including expanding the representations of "unit of communication" beyond the journal article to add datasets, simulations, software, and dynamic knowledge communication in their own right in order to facilitate collaborative network-based endeavours and increase the speed of discovery

[Van de Sompel 2004]. In addition, individual researchers have contributed their appraisal of the issues and concerns to be addressed with data use, storage and sharing [Coggins 2009].

Purpose of this study

This study was undertaken to address the significant gap in knowledge of researcher practice by eliciting a detailed and relational picture of data use, reuse and sharing practices. It used the research lifecycle, rather than the data lifecycle, to create a framework to understand the role of different types of data and its value.

To generate the understanding of patterns of practice across the research community, the study was guided by the following research questions:

What are the existing data use and sharing practices of researchers?
What are the differences and similarities between disciplines and fields of research?
What are the differences between research groups and project types?

Different practices and attitudes associated with different research models were examined such as collaborations with non-academic partners and multi-disciplinary research. Practices of early career researchers were investigated with reference to established researchers, particularly prominent researchers whose methods of data curation and attitudes to sharing their data may influence practices in their departments or fields of research.

The study drew on approaches taken by other research into the practices and workflows of researchers. As well as providing a sound approach it enabled comparison of local practices and attitudes against findings from research conducted elsewhere. For example, case studies were identified as a useful tool to engage with researchers especially when possible routes to improved methods are exemplified and where the researchers own use of best practice is highlighted [Lyon et al. 2010]. The approach taken was also informed by studies in related fields based around the research lifecycle, such the insights gained into research practice in using and sharing information resources [Randall et al 2009]. The study also drew on information already known about UNSW researchers. This included a supplementary report of the Intersect study [Markauskaite et al. 2009] provided to UNSW on its researchers responses in the areas of collaborative research, use of data from sources other than their research team and reasons for restricting access.

Findings from the study will inform development of data curation, data management support services and infrastructure at UNSW. It will establish the priorities for researchers from different disciplinary areas who have different research models and use different data types.

Research design and procedures

The study was conducted using a mixed methods research design, comprising focus groups, an online survey and in-depth interviews for data collection. Participants in focus groups were from a range of disciplinary areas including public health, social sciences, chemistry, and engineering, and with a balance of established researchers, early career researchers and postgraduate students. Invitations to participate in an online survey were emailed to UNSW academic staff and postgraduate students. Researchers who completed the survey were invited to participate in a follow-up semi-structured face-to-face interview.

Focus groups participants were asked to respond to questions about data management practices in their disciplines and fields of research, including data use and reuse. Development of the survey, comprising mostly forced choice items, was guided by themes generated from focus group data. Semi-structured interviews were chosen to collect qualitative data to allow interviewees to introduce unexpected themes. Responses to a set of questions from an interview schedule were combined with less formal discussion of particular issues that were raised by participants during their interviews. The survey and interviews were conducted concurrently.

Quantitative data from the survey were analysed using descriptive frequency and chi square statistics. The descriptive analysis was extended with multiple correspondence analysis

(MCA)[1] which provides a graphical representation of practices and attitudes to data use, reuse and sharing, including 'clusters' of related practices and attitudes. Thematic analysis was applied to qualitative data from the interviews, from which a number of case studies were extracted to provide a more comprehensive analysis. Case studies were identified and positioned as 'landmark' individuals in the MCA graphical output.

Summary of findings

Findings reported in this paper include a summary of key themes that emerged in the focus groups, and two case studies extracted from the interview data, of data management and sharing practices. Analyses of survey and interview data will be available in a forthcoming project report.

Focus group participants were from a range of disciplinary areas with varied levels of research experience. The first group consisted of social science and public health researchers, the second included researchers from various physical sciences, and a third group included engineering and physical science researchers. A fourth group comprised postgraduate research students from social and physical sciences.

Existing methods of collaboration framed discussions of procedures for data reuse. While disciplines such as astronomy and climatology have long traditions and established methods for 'open access' reuse of research data, participants from social science and public health were more inclined to share data within research relationships and partnerships with known colleagues.

Frameworks for data storage and archiving also differed across fields of research. In social science and public health disciplines, a condition of institutional ethics approval to conduct research is that clearly defined protocols for storage, retention and disposal of research data are followed. Researchers from physical sciences and engineering had a more subjective and pragmatic approach to determining how research data is stored and disposed of, with one participant referring to a balancing act between 'space' and 'value,' in which the expense of storing large datasets means that some data is not retained after a paper is accepted for publication.

Responsibility for maintaining privacy and confidentiality of participant information was a predominant theme raised by social science and public health researchers. It was identified as a major consideration in determining whether to provide access to data generated from a diverse range of research methods involving human participants including ethnographic field work, surveys and interviews, and clinical and pharmaceutical trials. Researchers emphasised the pivotal role of human research ethics committees in regulating access to and reuse of their research data.

Focus group participants from physical sciences and engineering referred to the role of scientific lab notebooks in providing access to results and research data. It is common practice for data and procedures recorded in lab notebooks to be used by other researchers to replicate experiments. Confidentiality and access restrictions were discussed with reference to intellectual property and ownership of data.

Attitudes and understandings relating to data sharing among physical scientists and engineers varied according to the career stage and research experience of participants. While some early career researchers discussed sharing data to advance research and to support colleagues, more experienced researchers indicated that complex and multi-layered issues relating to intellectual property and ownership of data need to be considered. Some early career researchers were enthusiastic about their results being verified by others and saw data sharing as an opportunity for other researchers to improve and add value to their methods and results. With reference to competitive advantage, a more experienced researcher stated that datasets

[1] Multiple Correspondence Analysis (MCA) is a nonparametric technique which examines and plots patterns in datasets to visually display relationships within and between large numbers of variables of categorical data. See Le Roux & Rouanet [2010]

collected over time are used to build future research projects, and cannot be made available for use by other researchers. Postgraduate research students acknowledged the need for supervisors to approve decisions about making research data available to others.

Some key issues were articulated across all focus groups. These included the importance of secure storage for research data, including secure back-up files. Participants in all groups referred to existing informal data sharing arrangements, based on trust, collegiality and expectations of reciprocity with colleagues and research partners. There was agreement about the role of formal data archiving policies and practices in fostering broader research collaboration. Detailed documentation of data across the research lifecycle was seen as necessary for replication of experiments in the physical sciences and engineering or for conducting meta-analyses in social science and public health research.

In summary, discussions about researchers' data management and data sharing practices were associated with specific research contexts. Research practices were associated with pre-existing patterns of communication and collaboration among researchers at different stages of their research career and in particular fields of research.

Case study A
The researcher (Professorial, Level E) is a Chief Investigator on social research projects and has lead a national research centre.

Data management practices in the research program
The research program informs policy development and most projects include non-academic research partners from government and community groups. Access to data is limited to researchers on the project and is provided to stakeholders as analysed results. Each project has one researcher responsible for the management, storage and security of data.

Data may be archived in a national data archive. Well managed and systematically archived data and related questionnaires is important for longitudinal analyses, analysis of aggregated data, and to build on past research.

Conditions for sharing data
Policies and procedures for sharing data were developed as the research program progressed, mainly in response to various requests for access to data.

Outside researchers must show they are bona fide researchers and outline what that they want to do with the data. Access is approved by the Chief Investigators. Re-analysis of the data may require further approval from the institutional ethics committee. A limited dataset or only analysed data may be provided when that meets the need of the outside researcher. It takes considerable time to prepare data and monitor access and outside researchers are required to stay in touch with the Chief Investigators.

Specific issues relating to managing and sharing data
Careful management of sensitive data is required to meet institutional ethics conditions and responsibilities to research partners and participants. In managing release of sensitive findings, advice may be sought from government and other relevant organisations, including those who funded the research.

Case study B
The researcher is a Research Fellow (Level B) working on a multi-disciplinary and multi-institutional engineering research project and has been responsible for storing and sharing project data and publications.

Data management practices in the research centre and field of research
This field of research does not generally adopt good data management practices. Students generally do what the person before them has done to record, maintain and store data. Not having a meaningful way to store and annotate data is an impediment to re-analysis. Standard operating procedures to support data reuse and sharing are sometimes implemented on projects but have not been sustained.

Lab notebooks are used as a primary source of data, to record data processing and are the source of evidence required for patents. Digitally generated data require digital record-keeping: it is hard to replicate digital data in a hard-copy notebook. Previous student notebooks are stored in the laboratory and are openly available for reference.

Conditions and issues relating to sharing

It is important for inexperienced researchers to learn to generate their own data. Data sharing with others outside researchers is not standard practice in engineering. Mostly data is shared to replicate and verify results. Conditions of use often stipulate that the outside researcher can use it as a quality check for their own data or use shared techniques to confirm their results but cannot publish shared data. Permission is approved by the Chief Investigator. Access to students' data is through their supervisor.

Restrictions on data are applied where projects have a commercial interest and are sponsored by industry. Research designed for patent purposes is 'closed' and data is not shared from the outset. If the idea for a patent emerges after the research has commenced, restrictions are applied to define how the data may now be used.

Specific issues relating to data management and sharing

The reliability of data and trust in the skills of the researcher who generated and manipulated the data is paramount. For this reason raw datasets are preferred.

Top down approaches to data management and sharing may be too generic but without institutionally supported policies and procedures it is difficult to generate good practices.

Next steps

Initial work has begun on developing tools that will help researchers manage their data. The UNSW Library is building a chemistry portal that will assist PhD students with data creation on their electronic lab notebooks. Working with the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO) we have begun to design and build a text mining tool, to provide semantic tags for unformatted text stored on the elab notebook. This tool will mine data to find relevant words and phrases, creating semantic tags for the words and phrases based on the research area's ontology. The semantic tags will be used to augment the metadata records applied to the unformatted text. In addition, the semantic tags will be used to build links between electronic notebook entries and published research outputs.

Our findings will also contribute to UNSW's response to building the national research data store. In the 2009 Australian Federal budget, ANDS was allocated $48 million from the Education Investment Fund (EIF) to create and develop an Australian Research Data Commons (ARDC). Over the intervening months ANDS consulted with major stakeholders, including UNSW, and has developed plans to seed the data commons by building datasets describing shareable Australian research collections. This will bring information about Australian research together in a coherent way [ANDS 2009]. The service has been designed so that participating agencies will provide an automated feed of collection description information to the ANDS Collections Registry using an XML format called Registry Interchange Format – Collections and Services (RIF-CS), based on ISO 2146.

UNSW has received funding for data capture projects and seeding the commons, designed to extend Research Data Management Services at UNSW. LRS will oversee the development of processes, tools and services to capture research data with its associated metadata and contribute RIF-CS metadata to the data commons. Eventually all researchers who contribute to UNSW data stores will have RIF-CS metadata harvested and contributed to the data commons and researchers will be able to deposit metadata records directly into ARDC. UNSW researchers will also be able to identify and contribute metadata about relationships between data collections in ARDC and research outputs in UNSWorks, the UNSW institutional repository.

The findings from this study will assist in the identification of important research data collections at UNSW and will help determine how policies and procedures for long term research data management will be developed at UNSW. Discovery of our research data collections by outside

researchers through access to the ARDC will increase the potential for reuse of data sets. This will provide increased opportunities for research collaborations nationally and internationally.

The final project report for this study is planned for release in September 2010 and will include analysis of the survey and interview data.

References

Amos, H. & Ruthven, T. (2008, April). *Becoming the authoritative source: Taking repositories to centre stage.* IATUL 2008: 29th Annual Conference of the International Association of Technological University Libraries, Auckland.
http://handle.unsw.edu.au/1959.4/40399

ANDS (2009). *A proposal for an Australian National Data Service*. Australian eResearch Infrastructure Council.
https://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf

Carusi, A. & Reimer, T. (2010). *Virtual Research Environment Collaborative Landscape Study*.
http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx

Coggins, J. (2009, February). *A researcher's perspective: the value and challenge of data.* A National Research Data Service for the UK? An International Conference on the UK Research Data Service Feasibility Study.

Cox, S., Frances, M., Croucher, J., Sidhunata, H. & Leslie, H. (2009, June). *MemRE: Collaborative development of an integrated research environment.* CERN Workshop on Innovations in Scholarly Communication, Geneva.
http://handle.unsw.edu.au/1959.4/40692

Dinkelmann, K., Edwards, M., Fry, J., Humphrey, C., Nakao, R. & Thomas, W. (2009). *Work flows - data discovery and dissemination: User perspective* (Data Documentation Initiative, Working Paper Series)
http://www.ddialliance.org/sites/default/files/bp/DDIBestPractices_Workflows-DiscoveryAndDissemination.doc.pdf

Frances, M. & Croucher, J. (2009, November). *Infrastructure for problem-based collaborative research: Aligning research, policy and practice.* eResearch Australasia, Sydney.
http://handle.unsw.edu.au/1959.4/44532

Granda, P., Hansen, Sue E,. Ionescu, S., LeClere, F. & Vardigan, M. (2009, May). *DDI across the life cycle: One data model, many products.* IASSIST 2009 International Association of Social Science Information Service and Technology (IASSIST), Tampere, Finland.

Gibbs, H. (2007). *DISC-UK DataShare: State-of-the-art review*. DISC-UK.
http://www.disc-uk.org/docs/state-of-the-art-review.pdf

Gibbs, H. (2009). *Southampton data survey: Our experience and lessons learned.* University of Southampton.
http://ie-repository.jisc.ac.uk/304/

Gold, A. (2007). *Cyberinfrastructure, data, and libraries, part 2. Libraries and the data challenge: roles and actions for libraries. D-Lib Magazine, 13*(9/10).
http://www.dlib.org/dlib/september07/gold/09gold-pt2.html

Green, A., Macdonald, S. & Rice, R. (2009). *Policy-making for research data in repositories: A guide. Version 1.2*. JISC.
http://www.disc-uk.org/docs/guide.pdf

Heery, R. & Powell, A. (2006). *Digital repositories roadmap: Looking forward*. UKOLN.
http://www.jisc.ac.uk/uploaded_documents/rep-roadmap-v15.doc

IWGDD. (2009). *Harnessing the power of digital data for science and society*. *Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council.*
http://www.nitrd.gov/about/harnessing_power_web.pdf

JISC. (2009). *JISC Strategy 2010-2012: Initial draft.* JISC.
http://www.jisc.ac.uk/aboutus/strategy/strategyreview1012

Le Roux, B. & Rouanet, H. (2010). *Multiple correspondence analysis*. Thousand Oaks, Sage.

Lyon L. (2007). *Dealing with data: Roles, responsibilities and relationships*, *consultancy report.* UKOLN. http://www.jisc.ac.uk/publications/publications/dealingwithdatareportfinal.aspx

Lyon, L., Rusbridge, C., Neilson, C. & Whyte, A. (2010). *Disciplinary approaches to sharing, curation, reuse and preservation: DCC SCARP final report*. JISC.
http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf

Markauskaite, L., Aditomo, A. & Hellmers, L. (2009). *Co-developing eResearch infrastructure: Technology-enhanced research practices, attitudes and requirements*. Sydney, The University of Sydney & Intersect.

Maron, Nancy L. & Smith, Kirby (2008). *Current models of digital scholarly communication: Results of an investigation conducted by Ithaka for the Association of Research Libraries*. Washington DC, Association of Research Libraries.
http://www.arl.org/bm~doc/current-models-report.pdf

Mochmann, E., Mauer, R. & Zenk-Möltgen, W. (2009, June). *E-infrastructure for comparative social research: From implementation to evaluation of the ZACAT data portal*. 5th International Conference on e-Social Science, Cologne.
http://www.merc.ac.uk/sites/default/files/events/conference//2009/papers/Mochmann.pdf

NCRIS. (2006). *An Australian e-research strategy and implementation framework. Final Report of the e-Research Coordinating Committee.*
http://ncris.innovation.gov.au/Documents/eRCCReport.pdf

Pepe, A., Borgman, C., Mayernik, M. & Van de Sompel, H. (2009). *From artifacts to aggregations: Modeling scientific life cycles on the semantic web.*
http://arxiv.org/ftp/arxiv/papers/0906/0906.2549.pdf

Pepler, S. (2008). *Citation, location and deposition in discipline and institutional repositories (CLADDIER): Final report.* JISC.
http://ie-repository.jisc.ac.uk/223/1/final_report_v7-1.doc

Polydoratou, P. (2007). Use of digital repositories by chemistry researchers: Results of a survey. *Program: Electronic library and information systems, 41,* pp386–399.

Randall, R., Smith, J., Clark, K. & Fried Foster, N. (2009). *The next generation of academics: A report on a study conducted at the University of Rochester*. Rochester, University of Rochester.
http://hdl.handle.net/1802/6053

Rice, R. (2009). *DISC-UK DataShare Project final report*. JISC.
http://ie-repository.jisc.ac.uk/336/1/DataSharefinalreport.pdf

Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C. & Warner, S. (2004). Rethinking scholarly communication, building the system that scholars deserve. *D-Lib Magazine, 10*(9).
http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html

Wallis, J., Borgman, C., Mayernik, M. & Pepe, A. (2008). Moving archival practices upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. *International Journal of Digital Curation, 1*.
http://www.ijdc.net/index.php/ijdc/article/viewFile/67/46