SHARED BENEFITS FROM EXPOSING RESEARCH DATA

Malcolm Wolski, Joanna Richardson, Robyn Rebollo
Griffith University, Australia
m.wolski@griffith.edu.au, j.richardson@griffith.edu.au, r.rebollo@griffith.edu.au

Abstract

The Australian National Data Service (ANDS) has taken a national leadership role in implementing a national collaborative research information infrastructure strategy. The strategy is designed to build the eResearch capability of Australia's researchers through national initiatives and strengthened engagement in international initiatives. ANDS has been established with an understanding that (1) data has become an increasingly important reusable product of research, and (2) research collaboration is fundamental to the resolution of the major challenges facing humanity in the twenty-first century.

With the objective of increasing the discoverability of Australian research data collections, ANDS has built the Research Data Australia (RDA) service. With the benefit of an ANDS grant to help populate the RDA, Griffith University has developed a framework for collecting and exposing research data. Critical to its success has been the key role played by library professionals in both its design and implementation.

The system collects appropriate metadata from research collections within the University through customised feeds from the various University content management and corporate systems. This system then acts as a central University repository to feed information in a standard format to Research Data Australia. It can also expose this data to University library discovery tools and other research information federations where appropriate. The end result is that this service not only integrates and aggregates data within the institution but also provides a key link into national and international open systems.

This paper describes this project, including the role played by the University's library professionals, the underlying infrastructure and processes and integration with national repositories, identifier services, and with the overarching national scholarly communication fabric. It concludes with an analysis of the benefits of such an initiative to individual researchers, to the University and to the national research agenda.

Keywords: eResearch, research infrastructure, metadata, research data, library discovery, university information services, information profession, role of the library

Introduction

In a submission to the European Commission, Kroes [High Level Expert Group on Scientific Data 2010] writes: "[A fundamental characteristic of our age is the] rising tide of information, sweeping across our professions, our families, our globe. We create it, transmit it, store it, receive it, consume it – and then, often, reprocess it to start the cycle all over again. It gives us power unprecedented in human history to understand and control our world. But, equally, it challenges our institutions, upsets our work habits and imposes unpredictable stresses upon our lives and societies." The knowledge generated from this deluge is seen as a key to global competitiveness while national prosperity is viewed as underpinned by knowledge innovation (O'Brien, 2010). Within this context, governments worldwide are grappling with the challenges of creating robust research e-infrastructures which can not only manage this information but also ensure its discoverability and accessibility.

Universities operate within both national and international environments, which in turn have their respective agendas. These drivers are having a profound influence on the development of institutional research frameworks which can help (1) researchers maximise the resources
available for research and (2) universities maximise the value of their investments in research so as to increase their research impact worldwide.

**National and International Research Drivers**

Funding bodies and national governments are seeking an improved return on investment for funded research. In a number of countries accountability is measured among universities by means of a research assessment exercise (RAE). The United Kingdom now has its Research Excellence Framework (REF); New Zealand has introduced the Performance-Based Research Fund (PBRF). The Australian government has just rolled out a national research evaluation initiative—Excellence in Research for Australia (ERA)—which is designed to provide benchmarking data for Australian universities compared with international measures. Another important metric is university rankings. A university’s research impact—the extent to which its research informs further research and practice—is a significant component of the university league table measures [O’Brien 2010]. Rankings have become important to stakeholders competing to attract the best students, lecturers, and researchers.

Along with improving the quality of their research output, universities are under pressure to increase the accessibility of their research. Both governments and funding agencies are promoting freely available, publicly-funded research findings. Mandates by funding bodies such as the National Institutes of Health (NIH), the Medical Research Council, the Wellcome Trust, the National Science Foundation and—in Australia—the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) demonstrate the recent change in funding rules based on new research paradigms. In addition many major research funders worldwide either currently have or are implementing policies that require grant holders to submit data management plans for formal approval and to manage their data in accordance with those plans. The NSF (National Science Foundation 2010), for example, has mandated that data management plans will be subject to peer review.

**University Libraries’ Response to the Challenge**

The mission of a university's library is intertwined with that of the university, i.e. making the world’s knowledge accessible to current and future scholars [O’Brien 2008]. Libraries have traditionally seen their role as “one of collection building for the future as much as about current provision” [Law 2009]. This role is now challenged by a scholarly and communication landscape which has changed profoundly and irrevocably. Indeed the very environment in which research is being conducted—and disseminated—is undergoing profound change. University libraries are faced with the challenge of managing these external changes in order to remain responsive and vital within their organisation.

Faced with the problem of a continuously changing context--referred to as “permanent white water” [Vaill 1991], a potential response can be either continued “industrial age attitudes, policies and procedures”, on the one hand, or a mastery of “the art of change”, on the other [Warnken 2004]. The response from most university libraries—and associated bodies—is to position themselves in terms of a larger strategic process, i.e. to be proactive and innovative rather than reactive. A range of possible strategies is highlighted in the current literature.

The view that the academic librarian is ideally equipped to support data curation [Furlough 2009] [Lowry, Adler, Hahn, & Stuart 2009] [Walton 2010] is gaining traction. The role of the academic librarian can be expanded to include involvement in “early planning and data-modelling phase of eResearch” (Luce, 2008) as well as curation and access [Borgman 2010]. In a response to the newly initiated National Science Foundation Data Sharing Policy, Hswes [Hswe & Holt 2010] outlines a new leadership role for academic libraries including services such as data consultation.

“Relationships between researchers and traditional library and university support for research have shifted radically. ... Libraries must articulate and create their own future. Academic libraries can support research by developing and aggregating discipline-based tools, providing customized services, and emphasizing user-centered services” [Kroll & Forsman 2010]. This may entail embedding information specialists, with relevant subject-based research experience,
in departments and research teams [Research Information Network 2010]. Researchers need to be recognised as both users and creators of an expanding range of digital information [Williams & Pryor 2009]. As Borgman [2010] so aptly encapsulates the new paradigm: “The role of libraries in research institutions is evolving from a focus on reader services to a focus on author services”.

Lowry [2009] advocates the “radical reconfiguration of research library organizations and services” along with an increasingly diverse and talented staff to respond to the rapidly changing environment. Libraries may be part of new hybrid organisations which will emerge as a result of tackling new support paradigms [Luce 2008]. Library staff have a key role in assisting other information specialists to ensure the adoption and usage of technical innovations for research support [Krafft, et al. 2010]. In this new world there are opportunities to “reshape the library staff in dramatic fashion” [Sennyey, Ross, & Mills 2009]. Puente [2010] and Johnson [2010] discuss some of the skills and competencies required in the new roles which are evolving. MacColl [2010] discusses how the strategic role which the library should play can be aligned to skills which are currently lacking in many universities.

At the institutional level, libraries will want to work more closely with Research Offices to review their provision of support for researchers [Research Information Network 2010]. In the new paradigm of collaboration and partnerships, libraries should emphasise proactive outreach and engagement by taking an active role as conveners among the different stakeholders [Luce 2008]. Finally university libraries need to position themselves within their parent organisation to demonstrate value [Oakleaf 2010].

Lougee [2009] summarises the major challenge: “Clearly, the world of . . . libraries will require this facility, this ability to constantly adapt as scholarship changes and as new technologies are added to our repertoire.” It is within this context that this paper describes how Griffith University has built a research e-infrastructure layer which not only connects individual researchers and the University to national Australian research service, but also feeds data into university discovery tools. This initiative has been led by the Division of Information Services, which includes the library.

**Griffith University**

Located in the rapidly growing corridor between Brisbane and the Gold Coast in Southeast Queensland, Griffith University is Australia’s ninth largest higher education provider. The University offers more than 300 degrees across five campuses to more than 40,000 students from 124 countries studying at undergraduate through to doctoral level in one of four broad academic groups: arts, education and law; business; science, engineering, environment and technology; and health. Griffith’s strategic research investment strategy positions it to be a world leader in the fields of Asian politics, trade and development; climate change adaptation; criminology; drug discovery and infectious disease; health; sustainable tourism; water science; music and the creative arts.

Griffith is regarded as one of Australia’s most innovative tertiary institutions and one of the most influential universities in the Asia-Pacific region. This innovation is carried through into the provision of information services, with e-learning, e-research, library, information and communication technology services, systems and infrastructure offered through a single integrated division, Information Services. This provides a distinct advantage to the University in an increasingly complex scholarly information and communication environment [O’Brien 2010] and was instrumental in the development of the service described in this paper.

**Australian National Agenda**

As part of the Australian government’s NCRIS (National Collaborative Research Infrastructure Strategy) initiative, the Australian National Data Service (ANDS) was formed to support the ‘Platforms for Collaboration’ capability. The service is underpinned by two fundamental concepts:

1. with the evolution of new means of data capture and storage, data has become an increasingly important component of the research endeavour, and
2. research collaboration is fundamental to the resolution of the major challenges facing humanity in the twenty-first century [Sandland 2009].

With a view to increasing the visibility / discoverability of Australian research data collections, ANDS has built the Research Data Australia (RDA) service (http://services.ands.org.au/home/orca/rda/). It consists of Web pages describing data collections produced by or relevant to Australian researchers. RDA publishes only the descriptive metadata; it is at the discretion of the custodian whether access, i.e. links, will be provided to the corresponding data. Behind RDA lies the Australian Research Data Commons (ARDC) which is the infrastructure and systems needed to support data and metadata capture, publication feeds, and applications such as data integration, visualisation and analysis.

**Collection of Research Activity Metadata**

Griffith University received NCRIS grant funding in 2010 for research data identification and discovery. Griffith’s Seeding the Data Commons project has captured data about the University’s research datasets, has assessed each dataset and determined appropriate access, and has then published 1,100+ records to Research Data Australia.

Four Academic Services Librarians who expressed an interest in research data management and curation work were selected to conduct data interviews with researchers who maintained research data created from research grants. Research projects funded by the Australian Research Council (ARC), National Health and Medical Research Council (NHMRC) and Australia Council for the Arts dating back to 2002 were identified and extracted from the University's research administrative system in preparation for the data interviews.

Not only were these Academic Services Librarians responsible for collecting potential research data but this work, especially the conversational interviews with researchers concerning their data practices and needs, was identified as a way of building deeper relationships and a better understanding of the researchers’ needs.

A questionnaire (http://tiny.cc/ve35r) was developed and implemented in a customised content management system used by the librarians to gather information about researcher background and research practices. Important information on data identification, data volume, data storage and data management was also collected and analysed. Part three of the questionnaire included several 'library focused' questions that addressed current awareness practices and web resource preferences. Armed with a print version of the questionnaire, the Academic Services Librarians proactively began making appointments with researchers in the various disciplines across the five campuses of Griffith University to record their responses. After three months, the majority of researchers had been interviewed and research data identified, described and uploaded into the customised content management system. Parenthetically this work has continued past the completion of the project; at the end of 2010 more than 600 researchers had been contacted across the University’s five campuses.

**Creation of a University Framework for eResearch**

In addition to the Seeding the Data Commons project, Griffith University was the lead partner with the Queensland University of Technology in an ANDS-EIF (Education Investment Fund) project to develop a middleware software solution which would aggregate data sources from within the University for uploading to Research Data Australia.

In a university with an active, broad research programme, it is to be expected that research data collections will reside in a range of different repositories, e.g. specialised discipline-specific repositories for stem cell research, historical data, film and music objects, and environmental data. In order to participate in Australia’s collaborative research infrastructure, universities will need to generate and collate a consistent metadata feed in order to populate Research Data Australia (RDA).

The Metadata Exchange Hub was a joint Griffith University and Queensland University of Technology (QUT) project, funded by ANDS, for the purpose of developing a master collection
of research data within the respective institutions, along with an automated update (feed) to Research Data Australia (RDA). The Hub collects appropriate metadata from research collections (at the content metadata level where possible) within the University through customised feeds from the various university content management systems. Also where authoritative source metadata is held in University corporate systems, feeds extract data directly from those databases. This Hub then acts as a central university repository to feed information in a standard format to RDA as well as university library discovery tools and other research federations where appropriate. The overall project objectives are:

1. to develop a sustainable solution to automate the collection of new research data held within the University and to populate RDA; and
2. to provide exemplars / good practice for Australian universities which want to be part of the national collaborative research infrastructure.

The following diagram (Figure 1) depicts the role that the Metadata Exchange Hub plays in aggregating data and identifying the relationships between key data elements. In addition two other external interfaces are required to complete the metadata picture. They are required to utilise persistent identifiers from ANDS and the National Library Australia People Australia service (http://www.nla.gov.au/initiatives/peopleaustralia/). The end result is that this service not only integrates and aggregates data within the institution, but it also provides a key link into national systems.

![Figure 1: Metadata Exchange Hub](image)

Given that the architecture that defines the hub must be open source, the decision was taken to implement a loosely coupled solution based on the Vitro software (http://vitro.mannlib.cornell.edu/) developed at Cornell University, which is an open source Integrated Ontology Editor and Semantic Web Application. The project has used this solution to support a research-focused ontology and to establish relationships between researchers and organisations, research collections, research activities (e.g. projects) and services. The project uses several other open source components, e.g. Persistent ID generator, Kepler workflow, OAI-PMH provider and data integrator. This approach has enabled maximum use of existing software and best use of programming time.

The following diagram (Figure 2) is a simple illustration of the Metadata Exchange Hub components. VIVO, which is based on technology developed at Cornell, has been implemented with minimal changes to the underlying software architecture.
Research activity metadata is uploaded to Research Data Australia (RDA) using the Registry Interchange Format - Collections and Services (RIF-CS). This data interchange format is based on ISO 2146:2010 Information and documentation -- Registry services for libraries and related organizations (http://www.iso.org/iso/catalogue_detail.htm?csnumber=44936). In addition an important part of the project has involved the development of a national research-focused ontology, based on the core Vitro ontology, which has been successfully deployed in the first version of the tool (http://www.ands.org.au/resource/rif-cs.html). This ontology employs components of a number of established ontology standards and describes the relationships between them. Collectively they provide a coherent framework for mapping the bulk of institutional research activity in Australia.

The architecture of the Metadata Exchange Hub has been designed to allow for automatic machine-to-machine communication for the ingestion of university research activity data. Previously identified relevant metadata is harvested from university repositories, data stores and corporate systems in its native form. The connectivity between research data and researchers is important, especially for purposes of reuse and in cross-disciplinary research. Identifying relationships between people, projects and institutions, for example, enhances opportunities for collaboration and new research [Thelwall, Li, Barjak, & Robinson 2008]. An important part of Griffith University’s Metadata Exchange Hub is to expose the relationships – using RIF-CS – among people, institutions, projects and the relevant research data created. These relationships form a linked graph as illustrated in Figure 3.
Both projects (NCRIS’ Seeding the Commons and EIF’s Fast Start Activity) were running in parallel. However the Seeding the Data Commons project commenced a month earlier; therefore the interviewing of 254 researchers was completed several months before the Metadata Hub was ready for pilot. As a result, one area of overlap was the development of a standard ontology as this dictated what data needed to be captured. Because the Metadata Exchange Hub was not ready, processes and systems and an interim database (nicknamed Leonard) had to be developed to collect data about research activity. Knowledge gained from this interim exercise proved very useful for designing the Hub processes in the later project and also informed thinking about skills and organisational structural changes. The following diagram (Figure 4) depicts the processes and systems put in place to capture the original set of data.

A more detailed analysis of the technical aspects of the implementation was presented at the 2010 VIVO conference [Rebollo, DeVine, & Porter, 2010].
**Utilisation of the Metadata Exchange Hub**

Although the Metadata Exchange Hub is in pilot, metadata collected to date has been harvested by both Research Data Australia and the National Library of Australia's Trove resource discovery system. In addition it is currently being interrogated internally by University researchers. University funds have been allocated as a high priority to move this system into production. Work is underway to finalise the automated updating of research activity data from enterprise systems with an anticipated rollout by mid 2011.

Because the Hub is based on linked open data, the metadata feeds expose the relationships among researchers, their research groups, their projects and their research outputs, including datasets. This means that research information is available for publishing in a “profile”. Therefore the Hub creates individual “Researcher Profiles”, which provide a history of research undertaken by a respective researcher. Similarly a “Research Group Profile” provides a history of research undertaken by a respective research group, e.g. research centre. Both have links to the actual research data, which supports the ANDS’ objectives outlined previously. These “profiles” will be uploaded to both RDA and Trove, which will increase research exposure at three levels: individual researcher, research group, and the university. For the postgraduate student, for example, a “profile” can be used as a tool to identify seminal research undertaken by experts within a particular group in order to select a potential supervisor.

The Hub also plays an important role through feeds into the University discovery services. For example, Griffith has recently deployed the Serials Solutions’ Summon web-based discovery service as the library search / discovery tool. It is now possible to utilise the Metadata Exchange Hub to push key research information through to the Summon library search tool, making it another resource available for scholarly purposes. The first stage of the Summon service ingestion includes research data collections published to the Hub as well as content within Griffith Research Online, the University’s institutional repository. As a result, previously siloed research content across the University is being made discoverable through the aggregation of data from a range of different systems.

**Discussion**

The projects recently funded by the Australian National Data Service are part of an overall response from the Australian federal government to emerging national research needs. Universities have been encouraged to participate by suggesting projects which will contribute to a national framework for collecting and exposing research data. Because Griffith University library services are integrated in a division along with IT and learning and teaching support, the Division of Information Services was nicely positioned to respond to the ANDS funding opportunities. There were already people with the required skills in the eResearch Services Unit and the Academic Services Unit who could come together to work on the projects. Where necessary, additional skilled people were drawn from other central IT units as well as the Office for Research to work on the project, e.g. arrange access rights to corporate systems and map ontologies to fields in the corporate research administration database.

Critical to the success of the EIF project was the availability of staff with a deep understanding of ontology development, metadata and semantic web technology. Library professionals and information science professionals were key resources in the project. Critical to the success of the NCRIS project was having a metadata specialist allocated to the project plus ready access to Academic Services Librarians. The latter already had a strong relationship with the researchers who had been selected for interviewing. In summary these projects utilised information professionals with specialised areas in information management, programming and database design, librarianship, and technical architecture design.

The development of a national standard ontology--based on core Vitro ontology--involved many information professionals from key stakeholder groups around Australia, e.g. universities, ANDS, National Library of Australia, and the Commonwealth Scientific and Industrial Research Organisation. By working with professionals from other Australian and overseas institutions, divisional staff have become members of networks of professionals who are discussing and collaborating on various topics, particularly standards. This has not only produced better
outcomes but also built a strong and broad knowledgebase upon which to move forward, thereby ensuring sustainable progress. An author of this paper, for example, has partnered with ANDS, VeRSI and Monash University to produce a guide entitled *What is Research Data?* [Henty, Searle, Winton, Rebollo, & Burton 2011]

The library is seen as the logical place which understands about information. Therefore it is ideally positioned to pull together the various stakeholders to develop institutional strategies to respond to new drivers [Luce 2010]. MacColl [2010] observes, “The library should be knowledgeable about knowledge, and should be the main authority on the campus about the ways knowledge is generated and transmitted through all of the disciplines it contains”. The Division has capitalised on the development of the Hub service to work with key stakeholders on a research data management proposal for the University and to initiate the development of data archives and related systems. In addition to further developing relationships with the researchers involved, the two projects have resulted in an improved communication network within the Division itself as well as within the University as a whole, especially with the Office for Research and with major research centres.

Although the project was undertaken with a view of long term benefits, some immediate outcomes have been to expose more information about research activity at Griffith, enabling discovery locally, nationally and internationally. While this helps lift the brand profile and research impact of the University, there is now increased potential for Griffith researchers to be approached to collaborate on projects, to share data, and consequently to be cited on any data made available. It has also lifted the University’s knowledge and capability to provide more “author services” [Borgman 2010] through the networks developed with research centres and the Office for Research, as well as through the integrated library and IT units within the Division. Academic Services Librarians now have new roles in data consultation and collection; the Division has a new enterprise system to collect, aggregate and publish research information; a more wholistic approach is currently being taken to the concept of a “resource discovery layer” within the University; and various stakeholder groups within the University are not only interested in using the Hub services but also wish to contribute data to the aggregator.

**Conclusion**

In 2010 Griffith University received national funding to develop a semantic web-based service for collecting research data as part of a national collaborative research information infrastructure strategy. Development of the service was undertaken by the Division of Information Services, which integrates e-learning, e-research, library, information and communication technology services, systems and infrastructure. By adopting some of the innovative strategies discussed earlier in the paper in regard to university libraries, the Division has demonstrated “value within the context of overarching institutions” [Oakleaf 2010].

Within Griffith the implementation of the Metadata Exchange Hub has highlighted the value that can be obtained by re-using and exposing research data already held in University repositories and databases, especially by making accessible information about the activities of Griffith researchers. Previously siloed research content across the University is being made discoverable through the aggregation of data from a range of different systems. The end result is a major contribution to the enhancement of both brand profile for the University and of research impact for not only the University but also individual researchers and research groups.

In addition to benefiting the University, this initiative provides a rich internal discovery environment which is also an important component of the national Australian research discovery layer.
References


