Bringing Together Researchers, Collaborators, Organisations and Data Resources: a University Model for Seeding the Data Commons

Robyn Rebollo¹, Cheryl Stevens² Kim Keogh³
¹Griffith University, Brisbane, Australia, r.rebollo@griffith.edu.au
²Griffith University Brisbane, Australia, c.stevens@griffith.edu.au
³Griffith University Brisbane, Australia, k.keogh@griffith.edu.au

INTRODUCTION:
The Australian government has made a significant investment in organisations, such as the Australian National Data Service, that provide educational outreach on data management and curation, and encourage institutions in the pursuit of proper data capture, cataloguing and discovery. Information Services (INS) at Griffith University embarked on the important ANDS endeavour, Seeding the Data Commons, through grant funding from the National Collaborative Research Infrastructure Scheme (NCRIS).

The primary goals of Seeding the Data Commons are

- To improve the state of data capture and management across the research sector
- To improve the fabric for data management and the amount of content in the data commons [1]

This abstract examines the diversified and modern approach Information Services (INS) at Griffith University has taken for Seeding the Data Commons. Our Seeding the Data Commons (NCRIS) project gathered information professionals with specialised areas in information management, innovative librarianship and technical architecture, to develop a uniform approach of identifying, capturing and exposing research data, while honouring, the key principles set forth in the Griffith University Code for the Responsible Conduct of Research. Specific actions conducted by the Seeding the Data Commons (NCRIS) project team included:

I. Identified and described research data associated with Griffith research projects in order to meet the criteria set out by the Australian Research Data Commons and the National Collaborative Research Infrastructure Strategy.

II. Assessed related datasets and identified appropriate access requirements (open, mediated, controlled, meta-data only, restricted) in line with Federal, State and Griffith’s data management policies.

III. Catalogued appropriate metadata pertaining to research metadata stores using the The Registry Interchange Format - Collections and Services (RIF-CS) Schema, the data interchange format supporting Research Data Australia. [2]

IV. Migrated Griffith’s RIF-CS records directly from a Griffith institutional repository to the Australian Research Data Commons (ARDC) using cross walking techniques, and via XML push technology and the use of the OAI-PMH harvest ‘protocol’.

V. Accommodated any legal, ethical, security or other constraints over the lifecycle of research data identified via direct communication with the Office of Research and research participants at Griffith University.

ARCHITECTURE:
The design, development and implementation of a Research Data Audit portal was produced to support the project’s deliverables. This solution, known by the project team as Leonard, included a robust content management system (e.g. Drupal) for capturing metadata associated with collections, projects, researchers, organisations and services. Initial data regarding researchers (e.g. contact details and affiliations) and projects (e.g. grant activities), was sourced from enterprise systems including our Research Administration Database and the HR Peoplesoft system. All RIF-CS ready data included in Leonard is routinely migrated to the Griffith Metadata Exchange Hub (via Vitro) and harvested (via OAI-Cat) by Research Data Australia in RIF-CS format using the OAI-PMH protocol. Other essential functionality incorporated into Leonard were an online questionnaire for research data audit interviews, and a comprehensive reporting system for auditing purposes. The reporting system was instrumental in conveying significant findings regarding the scalability of managing research data at a university level.
DATA INTERVIEWS:
As part of the INS restructure, late in 2009 the new Director of Scholarly Information and Research sought expressions of interest from the ex-Faculty Librarians to be involved in data collection work as part of the NCRIS project. Four information professionals volunteered to participate. Not only were these Contact Librarians responsible for collecting research data but this work was identified as a way of building deeper relationships with the Griffith academy. Datasets created from ARC, NHMRC and Arts Council funded projects since 2002 were to be identified by the Contacts from interviews with researchers. To retrieve this data, a questionnaire was developed by the eResearch Team. This online questionnaire was available via Leonard and collected information about researcher background, research practices, and data identification, volume, storage and management. Part 3 of the questionnaire included several 'library focused'-questions that addressed current awareness practices and web resource preferences. In February 2010, armed with this questionnaire, the four Contact Librarians proactively began making appointments with researchers in the various disciplines across the six campuses of Griffith. Three months later the majority of researchers had been interviewed and research data identified, described and uploaded into Leonard.

CATALOGUING:
Progressive cataloguers understand that describing relationships between records or objects through textual notes [e.g. MARC records] is no longer sufficient in the digital age of discovery. [3] The adoption of RIF-CS has enabled (INS) at Griffith University to expose reciprocal relationships between the following registry objects: collections, parties, activities and services. We established reciprocal relationships for our registry object records by using linkable identifiers (URIs / PIDs). It is recommended that uniform resource identifiers (URIs) or persistent identifiers (PIDs) are resolvable (they can be used to refer to a resource on a continuing basis, to find information about it, and to detect it).

CONCLUSION:
The collaboration of both technical and information professionals has significantly helped (INS) at Griffith University gain a greater understanding of how the availability of research data can contribute to Griffith’s overall research program and National discovery (Research Data Australia), agendas, and develop a continuing culture, capability and capacity to manage research data well into the 21st Century research realm.

REFERENCES: