Data Management Engagement Award Submission

Gavin Kennedy, Queensland Cyber Infrastructure Foundation, gavin.kennedy@qcif.edu.au;
Peter Sefton, University of Technology Sydney, peter.sefton@uts.edu.au

What we propose to do - summary

This proposal is to host an openly available online version of ReDBox, an innovative open source Research Data Management platform created using Australian government infrastructure grants and co-investment from Australian universities. The proposed evaluation platform will allow researchers to:

- Register a Research Data Management Plan (RDMP). The RDMP can be designed according to a preferred schema identified in consultation with SPARC/JISC. The RDMP can then be exported/published to research institutions and funding organisations.
- From the RDMP request a workspace in which to provision storage and manage data. The storage would be a free auto-allocation and provided by a SPARC/JISC infrastructure partner.

The ReDBox RDMP is more than a planning tool because it can be used to provision services, such as storage, using the ReDBox Provisioner. The ReDBox RDMP service also integrates with institutional repositories and registries to link in vital data, and generates research data collection metadata used for data publishing.

The ReDBox Provisioner is used to created structured, templated workspaces, such as pre-formatted notebooks, or well-laid-out git projects, and inject metadata into those workspaces, including a precis of the RDMP. This means that even in an evaluation platform, where the system is eventually made unavailable, the provisioner leaves useful metadata inside the workspace that can be accessed by researchers and institutional data managers later on.

This award would assist QCIF in creating an evaluation platform that would provide:
- A testing-ground for organisations (including JISC) which might wish to deploy the complete ReDBox 2 platform.
- Exportable PDF data management plans which can be submitted to funders.
- Data management records available to the infrastructure providers.

Using the award funds, plus in-kind effort from our collaborators, QCIF would:
1. Deploy the ReDBox platform on cloud infrastructure.
2. Implement a DMP based on standard UK/EU DMP elements.
3. Integrate the ReDBox platform with an appropriate authentication platform (e.g. OpenID Connect).

4. Implement two new provisioner modules: (a) NextCloud storage on JISC shared infrastructure and (b) github, as an exemplar of a commercial service.

ReDBox Background

ReDBox (Research Data Box) is a mature open source software application used by leading Universities for the planning, management and storage of research data assets. ReDBox development is managed by the Queensland Cyber Infrastructure Foundation (QCIF), a not-for-profit eResearch organisation. In use in Australian institutions since 2010, ReDBox was initially funded through the Australian National Data Service (ANDS) Metadata Stores project and other eResearch infrastructure grants. Now ReDBox is sustained by QCIF through a combination of a support based subscription service and client funded customisation projects. Growing our client base beyond Australia and adapting to the needs of a rapidly developing Research Data Management community is critical to the ongoing success of ReDBox.

QCIF is collaborating with the Australian Research Data Services (RDS) and the University of Technology Sydney (UTS) in the development of ReDBox 2. ReDBox 2 is not only a software stack renewal, it incorporates new research data life cycle supporting services and functions.

Two innovative features of this development are the integrated Research Data Management Planning tool and the ReDBox Provisioner.

The RDMP tool allows researchers to document how they plan to collect, store, secure and share their research data. ReDBox provides customisable forms and workflows to assist researchers, supervisors and administrators in creating, tracking and approving RDMPs. Using the Mint name-authority and vocabulary service that complements ReDBox, the RDMP tool readily integrates with institutional sources of truth such as Research Information Management Systems and HR/Identity management platforms to index and make data available to support the RDMP completion (e.g. personnel or grant details).

Using ReDBox researchers can then access infrastructure through the DMP process using the ReDBox Provisioner, a service catalogue from which researchers can provision services. Developed by UTS and integrated by QCIF, the ReDBox Provisioner provides a modular architecture for the provisioning and tracking of research workspaces, so researchers can request services such as storage space allocations, or a git repository and associated issue management in a system such as gitlab, or a new live working repository collection in which to store data. ReDBox not only facilitates the creation of workspaces, but can identify and link to existing workspaces.

As ReDBox 2 is further developed it will support harvesting of data and file level metadata from the provisioned resources back into the ReDBox repository and subsequent curation and publication of the data collection.