Researcher Centred Data Repository Workflows

Potential Workflows

1. **DMP-derived project description, then project-end deposit of results**
   A Data Management Plan (DMP) is created/refined on award and shared with Research Data Services (RDS). RDS creates a “container” describing the research project based on the DMP which may be made public. The outputs of the project (both publications and data collections) are added later and linked to the project when offered by the researcher.

2. **Ad hoc deposits without project context**
   Researcher deposits an individual data collection without a project description, possibly in response to a publisher policy on data availability. Data underlying one or more publications can be linked to them when available either in advance of or after publication. A project container can be added at a later date. If this information becomes available.

3. **Publication deposit, with data as support**
   Researcher deposits a publication in the publications repository (ROAR). They then add the data underlying this publication into data.uel, recording the association between the two.

4. **Publication deposit, with data deposited elsewhere**
   Researcher deposits a publication, which details the availability of data at an external data archive/service. Using data.uel as a registry, the latter is recorded in a blank data collection linking to the external site.

**Abstract**

Supporting the management of research data in universities requires a range of services and components, including a long-term vehicle for archiving and sharing data. Such data repositories can serve multiple functions, and ensuring they engage with researchers through appropriate workflows is a key challenge. Datauel, a new data repository at the University of East London, seeks to provide workflows for four use cases.

Repositories can serve multiple functions, and ensuring they engage with researchers through appropriate workflows is a key challenge. Data.uel, a new data repository at the University of East London, seeks to provide workflows for four use cases.

Authors:
Stephen Grace
Research Services Librarian
s.grace@uel.ac.uk

David McElroy
RDM Officer
d.mcelroy@uel.ac.uk

Research Data Services
University of East London

Acknowledgements:
We would like to thank Rory McNicholl and Timothy Miles from the Office of University of London Computer Centre for the excellent work developing and supporting our repositories.

Technical

Both ROAR and data.uel are hosted and maintained by the University of London Computer Centre (ULCC). UEL created the specification and designed the look and feel, while ULCC built the back-end repository. ULCC created the linking system to allow for the workflows discussed above. This involved two different types of linking between ePrint records, and one external link. As the two types of ePrint to ePrint link do not exist in stock EPrints, some customisation was needed.

There are 3 types of links deployed:

1. **Project to Datasets**
   This was accomplished using the Containers plugin for EPrints. The plugin allows a record to be created which can contain any number of dataset records in the repository. The linking is the first part of the Project workflow.

2. **Datasets to link to the Publications which they support**
   The depositor can search the repository by dataset title, linking as many as needed. There are three types of links deployed:

   These links are then stored in the Container ePrint.

3. **ePrints to datasets - data.uel to ROAR**
   This is a custom plugin written by ULCC. It allows datasets to link to the publications which they support in ROAR. The links are stored in the data.uel workspace. The depositor searches for a title in ROAR, which supplies the dataset from the ROAR repository.

This is a many to many relationship, so the depositor can link to multiple publications in ROAR.

The link is only stored in dataset, so ROAR has to look up dataset to get the links. ROAR sends a GET request for a JSON, to see if a link has been made. If the link is to a ROAR that has been defined, it gets an empty file.

If a link is present, the details are returned in the JSON.

The link is displayed on the summary page, with the colour of the collection repository determining where it is going to. This information is also in the machine-readable metadata as:

   "meta:internal_link_type": "http://data.uel.ac.uk/id/app/355/"

**3. Data to external datasets (Figshare)**

Links to external resources are handled by the existing EPrints field linked_resource, with minor customisation.

It allows for externally stored datasets (UK Data Archive, Figshare, etc) to link to an ePrint. This means it is possible to have a metadata-only record (registry) which links to the actual data.

The depositor inputs the URL, the title and a description of what the object is. These are then displayed in the same way as the repository links above but with a different colour to help differentiate them.