Data, CRIS and other systems

Presented By James Toon – Customer Consultant Pure
Date November 2016
Introduction

- Mission of a CRIS – system for managing data about research activity and the relationships between outcomes
- Focus on providing information for evidence based decision making
- Providing functionality to support workflow management for compliance and reporting – a constantly changing space
- Can provide a variety of services in support of research outcomes management *including* that of a repository (data and publication)
- Functionality in CRIS is provided to support both research and administrative requirements for institutions, but these needs are often very different – and challenging

https://dx.doi.org/10.6084/m9.figshare.873617.v4
https://dx.doi.org/10.6084/m9.figshare.873617.v4
Retrieved: 10 19, Nov 21, 2016 (GMT)
Where we are now

• Ability to capture data on datasets in Pure since at least 2010 (and relate these to other outcomes)
• Datasets working group established following Herts user group meeting in May 2012 to advise on desired/required functionality. Driven heavily by the compliance to EPSRC data policy
• Significant changes made to Pure to provide dedicated content type and associated functionality based on user group input
• Focus has been on Pure as a registry rather than a data store, but tries to offer options in both scenarios
• Dataset info available on portals and via Pure web services for integration into other services as required
• Recent changes (from v5.6) to allow synchronisation and import of dataset metadata via xml (with supporting workflow)
• Mendeley data integration on roadmap as delivery candidate in v5.8
Choices

• We have seen a variety of approaches taken across the Pure user community… some institutions want to use Pure as register only, some as data repository, some both (and some none).

• Important to note that both scenarios are supported (and will continue to be)
1. Stored (existing in some form)
2. Preserved (long-term & format-independent)
3. Accessible
4. Discoverable (data is indexed or data is linked from article)
5. Citable
6. Comprehensible (description / method is available)
7. Trusted (e.g. reviewed)
8. Reproducible
9. Re-usable (allow tools to run on it)

Reproducibility Strategy
Accessibility Strategy
Discoverability Strategy
Reusability Strategy

Integrate upstream and downstream - make metadata to serve use (inc. reportability and compliance).

https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data
De Waard, Cousijn, Aalbersberg
Configurable Dataset Synchronisation

This job synchronises datasets from an XML file. Below you can find the XML schema files that describes the XML format of the source file. The source file must be valid when validated against the schema files.

There are also 2 example files, one containing two simple examples and one containing an advanced example with different text in multiple languages.

**NOTICE!** You cannot synchronise the examples into Pure without correcting person and organisation lookups in the examples to ids of persons and organisations that already exist in your Pure.
Meeting expectations

• Institutional administrative requirements for managing information on research data distinct from the needs of researchers.

• CRIS mostly focusing on the administrative requirement

• The use cases for academic use can be problematic for whole host of reasons. Layers of complexity on top of complexity.

• Trying to meet as wide a set of use cases as possible but not able to provide ALL solutions to ALL scenarios. Customer expectation management needed.
1. Stored (existing in some form)
2. Preserved (long-term & format-independent)
3. Accessible
4. Discoverable (data is indexed or data is linked from article)
5. Citable
6. Comprehensible (description / method is available)
7. Trusted (e.g. reviewed)
8. Reproducible
9. Re-usable (allow tools to run on it)

Reproducibility Strategy
Reusability Strategy
Discoverability Strategy
Accessibility Strategy

https://www.elsevier.com/connect/10-aspects-of-highly-effective-research-data
De Waard, Cousijn, Aalbersberg
Multiple components in academic workflow

Integration into researcher workflow (scalability)

Capture and annotate Research Data

Publish, Manage and Share Research Data

Compliance/Reporting

Track all your research outputs

Quality

Function

API function

API function
Workflow

• Researchers use Mendeley Data for their daily work on Data sets
  • Update their profile to include at which institution they work
• Import job setup in Pure that will on regular intervals pull in Data sets belonging to their institution into Pure
  • If researchers Mendeley ID is registered in Pure, import is fully automatic
  • If researcher does not have Mendeley ID registered in Pure we do “author match” like for publications and present as candidate
• All imported Data sets will be in the “For validation” workflow step to allow for local validation/enhancement workflow
• Mendeley Data is our first integration for data sets, we are expecting to make more integrations in the future with other solutions
Summary

- Different needs for academics and research institutions drive development
- Academic needs are likely best served via purpose led tools (such as Mendeley Data) but local storage supported too.
- Believe that the institution administrative needs best served by the CRIS i.e. registration, linking, reporting, compliance and showcasing
- Combined offering where data seamlessly travels between services to CRIS designed to serve both academic and administrative use cases
- Mendeley Data is our first integration for data sets – more integrations are expected
Thank you & do you have questions?

j.toon@Elsevier.com