ReDBox 2.0 / Provisioner

Gavin Kennedy, QCIF
Peter Sefton, University of Technology Sydney
ReDBox

ReDBox is a Research Data Management Platform that assists researchers and institutions to **plan, create and publish** their research data assets.

- Open source (github.com/redbox-mint)
- Managed and supported by QCIF
- Fully customisable
- **Integration and interoperation**
- Cloudified
- Community Driven
ReDBox History
ReDBox 1.x

1. Institutional Data Catalogue ++
2. Metadata Store
3. Mint Name Space Authority
4. DMPt
5. Curator
6. Harvester
7. Transformers
8. API
9. OAI/PMH Publisher
ReDBox 2.0

- Technology refresh
- Single web application framework
- Easy to configure forms
- Modular/plug-in architecture
- Multi-institutional RBaaS (e.g. RAPortal.org.au)
- DLC (Data LifeCycle) support via RAiDs
- maDMPs
- Workspace concept to manage resources and datasets
- Flexible storage provisioning with the UTS Provisioner Framework
- Dataset harvesting, archiving and publishing based on BagIt
- Comprehensive search/discovery tool
Data Life Cycle

Collaborators

Data Stores

Datasets

Provisioning

Harvesting

Group Id

Activity ID (RAID)

RDMP

Metadata & Archive

Search

Researcher

Institutional RIMS and REPOs

Global Data Catalogues
ReDBox
maDMPs
RAP

- www.raportal.org.au
- Configurable SaaS Platform
- Proto-DMP
- RAiD Support in ReDBox
- Institutional Views
Provisioner - principles
What is Provisioner
Give the people what they (should want)

Improved ability to do high-integrity high-impact research
Small Pieces, loosely joined

Avoid point to point integration, use standards for integration
An army of bots

To automate obvious things, like provisioning lab notebooks to everyone in a lab, git repository of every software PhD
Cite this work:
Chambers, Deborah; Liston, Carol; Wieneke, Christine (2015) Farms to Freeways Example Dataset. Western Sydney University. Datacrate. http://dx.doi.org/10.4225/35/555d661071c76
Download a zip file

This catalog file describes a dataset. It uses the Draft DataCrate Packaging format v0.2.
This file was made with Calcyte.
A machine-readable version of this page is available: CATALOG.json

<table>
<thead>
<tr>
<th>@type</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>@name</td>
<td>Farms to Freeways Example Dataset</td>
</tr>
<tr>
<td>@description</td>
<td>This data set was exported from an Omeka Repository as an example of a DataCrate. It contains the Collections and Items from the repository but does NOT have the exhibitions. The DOI resolves to an archive of the data elsewhere</td>
</tr>
<tr>
<td>@datePublished</td>
<td>2015</td>
</tr>
<tr>
<td>@creator</td>
<td>Deborah Chambers, Carol Liston, Christine Wieneke</td>
</tr>
<tr>
<td>@path</td>
<td>data</td>
</tr>
<tr>
<td>@publisher</td>
<td>Western Sydney University</td>
</tr>
<tr>
<td>@hasPart</td>
<td>Interview Transcripts, Interview Audio Recordings, Photographs, Letters and Notes, Project Materials, Interviewees</td>
</tr>
</tbody>
</table>
DataCrate Specification version 0.2.1

This is a draft work in progress, any and all details of this spec could change, including its very existence.

As this spec is still under development, v0.2 will be updated and committed to github here: https://github.com/UTS-eResearch/datacrate/tree/master/spec/0.2. The version number above will be incremented if there are breaking changes from 0.2.1, edits for clarity and more detail will not be considered a new version.

Please give feedback via github issues or a pull request.

The samples and examples here have not been exhaustively verified, and there may be some inconsistencies between this spec and the associated files, such as the DataCrate JSON-LD Context. Where there is a discrepancy, the file is the authority rather than the spec (this doc).
Adding a workspace
Workspaces

Associated workspaces

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
<th>Location</th>
</tr>
</thead>
</table>

Pre-Provisioning

Pre Provisioning (*)

eResearch storage platforms (e.g. OMERo, GitLab)

Create or link a workspace

Select from the list the desired app

- GitLab
Researchers use the Service Catalogue to create, share and link workspaces in research apps, and to publish datasets.

Mint is a component of Stash which provides contextual metadata: for eg, a researcher's current active projects; PhD student-supervisor links etc.

For some apps, provisioning can go through Service Connect for approval.
Future possibilities

- Generic long tail data repositories
- Integration with preservation systems
- More provisioning plugins (e.g., Github, OSF, Figshare)
- National data catalogue
- Actionable Service Catalogue
- (inter)National DMP
- Open data workflows
- Generate project portals
Sustainability

- Yes, it is FREE OPEN SOURCE SOFTWARE, but...
- QCIF is Not For Profit
- Support subscriptions
- Project funding
- Managed hosting
- SaaS model
- Communication
- Community driven development