Breaking down the boundaries to storing, sharing and publishing research data

Steve Androulakis, Philip Bertling, David Groenewegen and Andrew Harrison
Store.Synchrotron.org.au

• Store.Synchrotron is a system that captures all macromolecular beamline data, available online to all non-commercial Australian Synchrotron users. It was developed by Monash University in a strategic, ongoing partnership.

• Data is **immediately shareable** by the researcher on the web and able to be **published**.

• The service operates on the Australian NeCTAR Research compute cloud in a scalable setup able to withstand load. ([http://nectar.org.au/](http://nectar.org.au/))

• We’re actively opening access to raw data behind high-impact research publications under CC BY licenses. Six institutions have opened data so far.

• Built on MyTardis – an open source, Australian made data management platform used all over Australia in proteomics, genomics, electron microscopy, medical imaging, astrophysics, quantum physics and more.
  • Generalised so can be used for other data capture processes
  • Growing community of users and instruments (40 at Monash)
Why do it?

- To meet the challenge of diffraction data deposition
- Provide a model for other synchrotrons globally
- To ensure that satisfactory evidence is provided to support the interpretation of structural experiments
Data Curation Continuum

Private Research, Shared Research, Publication, and the Boundary Transitions

**Private Domain**
- Laboratory Information Management System/Research Management System
  - May link to data objects stored in Research Data Store

**Shared Domain**
- Collaboration Support System (Plone, TWiki, Sharepoint)
  - May link to data objects stored in Collaboration Data Store
  - Authorised by research team leader
  - Performed by research team IT support
  - Involves object selection, migration, assignment of access controls, augmentation of metadata

**Public Domain**
- Institutional document repository
  - May link to data objects stored in Public Data Store/Repository
  - Authorised by research team leader
  - Performed by research team IT support/IR staff
  - Involves object selection, migration, assignment of persistent identifier, creation of fixity metadata, augmentation of other metadata

This domain involves the core research team as they undertake the research, usually within a single institution. Access is often tightly controlled as hypotheses and analyses are developed.

This domain involves researchers outside the core team as they collaborate with colleagues, often across institutions. Access is more open, but not everything is shared.

This domain involves the public sphere (publication in the sense of making public). Access usually open to all.

Previous state

- Most users were taking portable hard drives to the beam lines
- Problems:
  - There is little standardisation in the storage and filing of data.
  - Portable drives may be the lab’s sole resource for raw data.
  - The metadata, essential to identify which raw frames match a given structure, may be lost.
  - Several instances of users contacting the AS for access to raw images relating to datasets that they have lost or when their own media got corrupted.
The Operation of Store.Synchrotron

Collector

Beamline

Store.Synchrotron

Collector

Collaborators

Public
What about the metadata?

- Use external sources to derive metadata as early as possible in process
  - Instrument metadata at time of capture
  - User proposal information for descriptive metadata

- On publication work with researchers directly
  - To enhance descriptions
  - Connect to related publications
  - Identify other outside sources of metadata eg PBD

- Add additional discovery information
  - Citation for dataset
  - Digital Object Identifier
Capture began June 2013. As of June 2014, it has captured over 26 terabytes of data in over 1.5 million raw diffraction images.

Source: http://bdp-aaf-dev.dyndns.org/graphtime.html
X-ray crystal structure of the streptococcal specific phage lysin PlyC

Sheena McGowan, Ashley Buckle, James Whisstock

Experiment

Describe
Share & Publish
Browse
Download
Integrate.. Provide..
Sharing

Users
Users who have a share in this experiment:

<table>
<thead>
<tr>
<th>Username</th>
<th>Name</th>
<th>Permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>mytardis</td>
<td>Steve Ancirculakis</td>
<td>Read Edit Owner</td>
</tr>
<tr>
<td>oded</td>
<td>Oded Kleinfeld</td>
<td>Read</td>
</tr>
<tr>
<td>bosco</td>
<td>Bosco Ho</td>
<td>Read</td>
</tr>
<tr>
<td>synchrotron</td>
<td>Synchrotron Test</td>
<td>Read Edit Owner</td>
</tr>
</tbody>
</table>

Change User Sharing

Groups
Groups who have a share in this experiment:

There are currently no groups with access to this experiment.

Change Group Sharing

Links
This experiment is private. A temporary link can be created by its owner(s) and privately shared for direct access.

Temporary access links provide full access to recipients regardless of an experiment’s public status.

<table>
<thead>
<tr>
<th>Temporary Link</th>
<th>Expiry</th>
<th>Granted By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right click...</td>
<td>15th November 2012</td>
<td>mytardis</td>
</tr>
<tr>
<td>copy link</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right click...</td>
<td>7th December 2012</td>
<td>mytardis</td>
</tr>
<tr>
<td>copy link</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Create New Temporary Link

OR2104: Breaking down the boundaries
Sharing

Users

Users who have a share in this experiment:

<table>
<thead>
<tr>
<th>Username</th>
<th>Name</th>
<th>Permissions</th>
</tr>
</thead>
<tbody>
<tr>
<td>mytardis</td>
<td>Steve Ancirculakis</td>
<td>Read, Edit, Owner</td>
</tr>
<tr>
<td>oded</td>
<td>Oded Kleinfeld</td>
<td>Read</td>
</tr>
<tr>
<td>bosco</td>
<td>Bosco Ho</td>
<td>Read</td>
</tr>
<tr>
<td>synchrotron</td>
<td>Synchrotron Test</td>
<td>Read, Edit</td>
</tr>
</tbody>
</table>

Groups

Groups who have a share in this experiment:

There are currently no groups with access to this experiment.

Change User Sharing
Change Group Sharing

Links

This experiment is private. A temporary link can be created by its owner(s) and private groups who have a share in this experiment.

Temporary access links provide full access to recipients regardless of an experiment's public access.

<table>
<thead>
<tr>
<th>Temporary Link</th>
<th>Expiry</th>
<th>Granted By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right click..</td>
<td>copy link</td>
<td>15th November 2012</td>
</tr>
<tr>
<td>Right click..</td>
<td>copy link</td>
<td>7th December 2012</td>
</tr>
</tbody>
</table>

Change User Sharing
Create New Temporary Link

Public Access

Step 1: Change Public Access:

Public access

Step 2: Select a license:

Creative Commons Attribution 3.0 Australia (CC BY 3.0)
This licence lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licences offered under Creative Commons.

Creative Commons Attribution-ShareAlike 3.0 Australia (CC BY-SA 3.0)
This licence lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and licence their new creations under the identical terms.
Challenges

- Researcher engagement
- Understanding researcher needs and domain
- Keeping up with changing requirements
- Making a project into a well supported service
- Currently have a complicated work flow to get data released
- Ongoing cost of keeping data available, maintaining DOIs
Presented on the Synchrotron’s Control PCs in a slideshow.
Discovery

- Currently feeding records to
  - Research Data Australia
  - Monash University Research Repository

- Future
  - Trove
Going forward

- Multi modal data
- Automating more workflows to increase diversity of publications and disciplines
- Incorporating connections to scientific tools and workflows eg connect to HPC to process results
- Increasing ‘instrument integration’ of this kind using the same underlying system as Store.Synchrotron.
  - This includes expansion to more beamlines at the Australian Synchrotron and likely the Australian neutron source ANSTO. Experiences here with different research communities, policies, practices, technology and culture should be broadly useful to this group.
Acknowledgments

- Auto-processing and the Store.Synchrotron was partially supported by the National eResearch Collaboration Tools and Resources (NeCTAR - grant “Monash Synchrotron - Storage Service”) project (www.nectar.org.au), and the VicNode/RDSI (grant 2013R2.1).

- Aspects of this work were funded by the Australian National Health and Medical Research Council (DA, APP1047004; CSB, APP513880 and APP1050585)
THANKS – QUESTIONS?

David.Groenewegen@monash.edu
@groenewegendave

THANKS – QUESTIONS?