



MONASH University

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

Steve Androulakis, Philip Bertling, **David Groenewegen** and Andrew Harrison

Store.Synchrotron.org.au

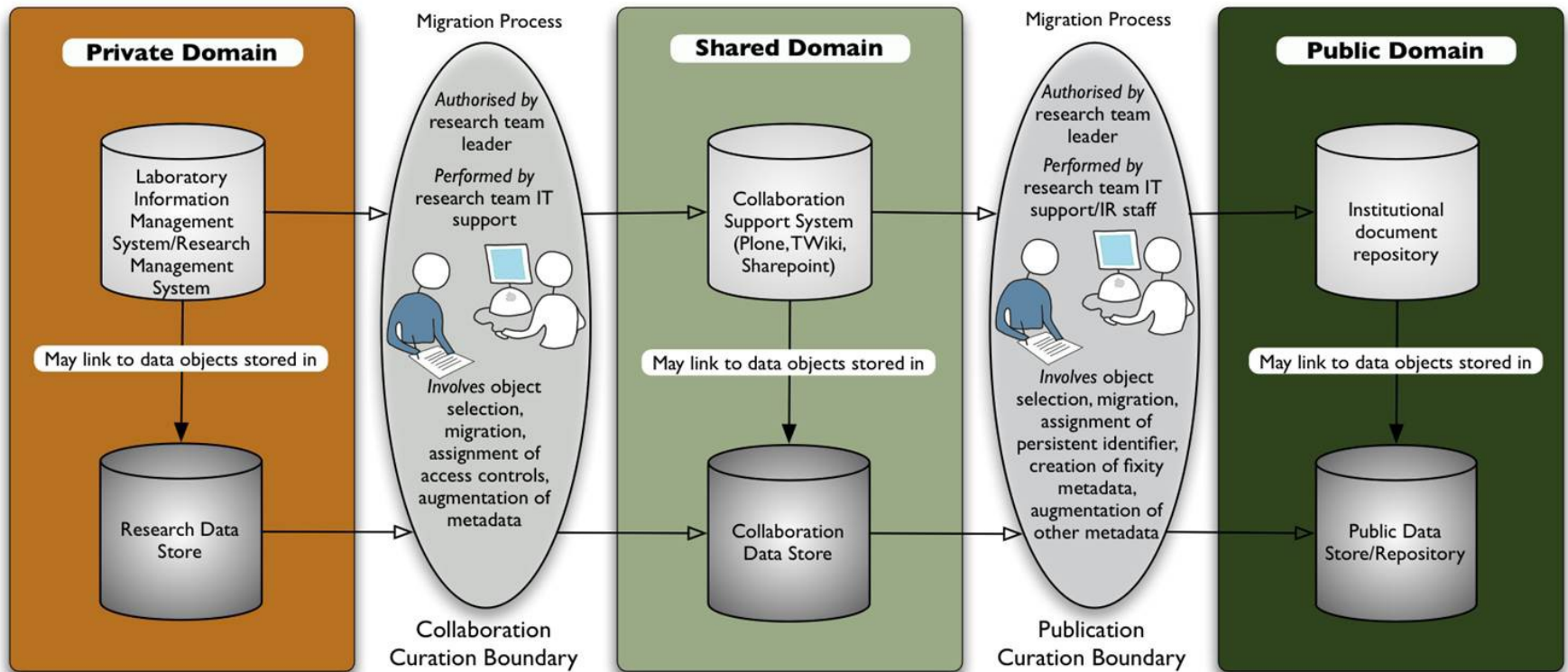
- Store.Synchrotron is a system that captures all macromolecular beamline data, available online to all non-commercial Australian Synchrotron users. It was developed by Monash University in a strategic, ongoing partnership.
- Data is **immediately shareable** by the researcher on the web and able to be **published**.
- The service operates on the Australian NeCTAR Research compute cloud in a scalable setup able to withstand load. (<http://nectar.org.au/>)
- 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



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.

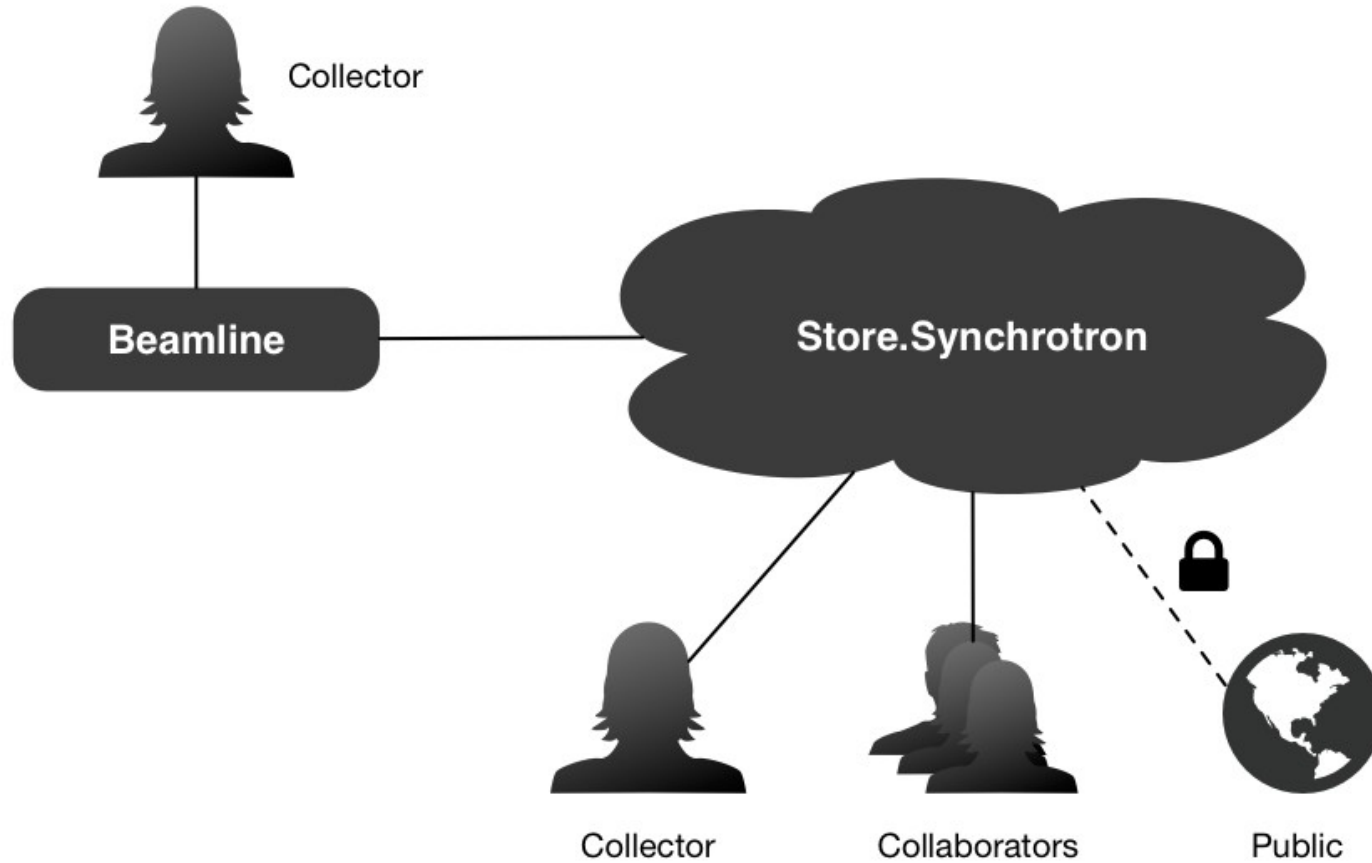
Version 1.4.3, <http://andrew.treloar.net/>, 19 Mar 2012



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.

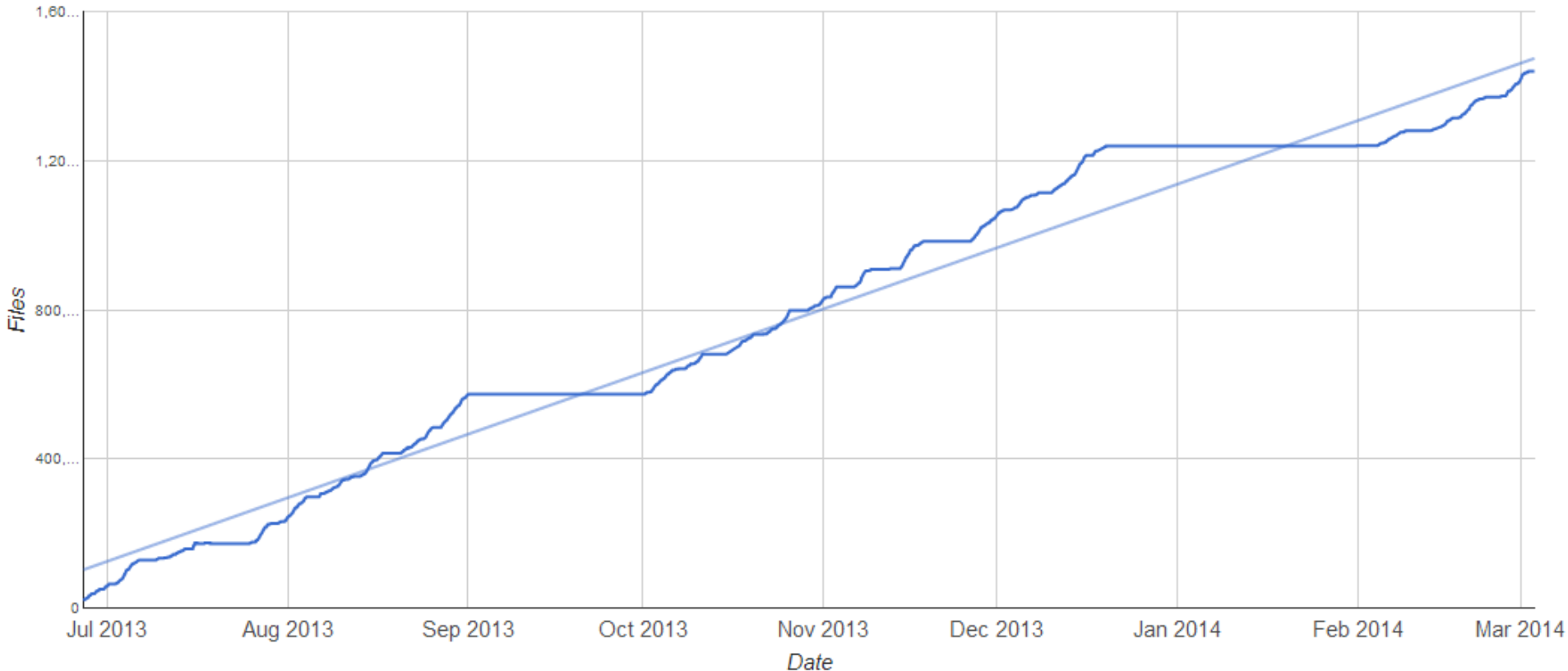
The Operation of Store.Synchrotron



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

Real-time instrument data capture



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>

Dataset

2K3e_5

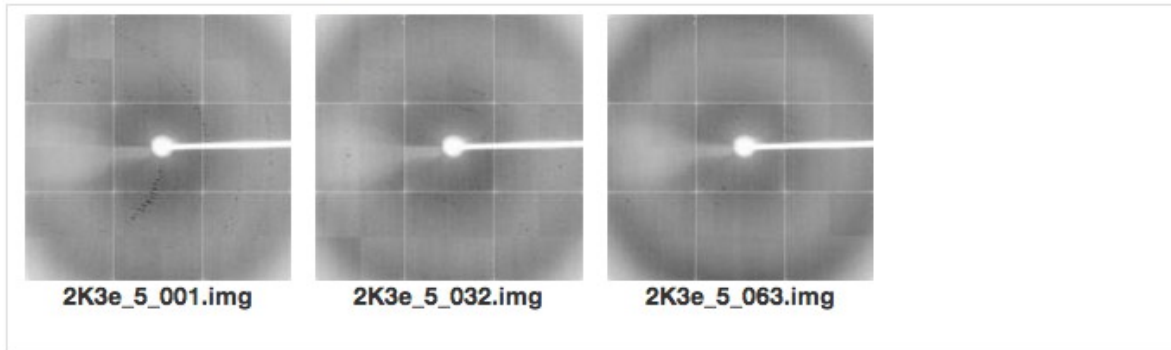
[Edit Dataset Description](#)

2

94

1.7 GB

A few diffractions from your dataset



Datafile Information

Click the button next to any file to view details and actions.

Dataset Metadata

[+ Add Dataset Metadata](#)

Synchrotron MX Data

Other Experiments

This dataset also appears in these experiments.

[Experiment 3511](#)

94 Files

Download Dataset: [TAR](#)

Search: Enter part or all of a filename, then

[Download Selected Files](#)[Upload Files](#)Select: [All](#) / [None](#)

<input type="checkbox"/>	2K3e_5_001.img (18.0 MB)		Download	+	List
<input type="checkbox"/>	2K3e_5_002.img (18.0 MB)		Download	+	List
<input type="checkbox"/>	2K3e_5_003.img (18.0 MB)		Download	+	List
<input type="checkbox"/>	2K3e_5_004.img (18.0 MB)		Download	+	List
<input type="checkbox"/>	2K3e_5_005.img (18.0 MB)		Download	+	List
<input type="checkbox"/>	2K3e_5_006.img (18.0 MB)		Download	+	List
<input type="checkbox"/>	2K3e_5_007.img (18.0 MB)		Download	+	List
<input type="checkbox"/>	2K3e_5_008.img (18.0 MB)		Download	+	List

Experiment X-ray crystal structure of the streptococcal specific phage lysin PlyC

Sheena McGowan , Ashley Buckle , James Whisstock

2 631 10.3 GB 28th October 2012 Public

Describe



Show Description

DOI for this data: 10.5072/03/50171CC7EEC1DSample diffraction image (PlyCB: plycb0001.img):Publication (...)

Description Metadata Sharing Transfer Datasets

Institution Monash University

Licensing Unspecified

Download All ZIP TAR

Share & Publish

Download

2 Datasets

Download Selected

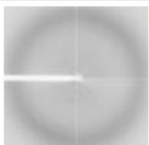
Just start typing to filter datasets based on descriptions

PlyCB



120 3.8 GB

PlyC



411 6.5 GB

Browse

Integrate..

Provide..

Sharing

Users

Users who have a share in this experiment:

Username	Name	Permissions
mytardis	Steve Androulakis	Read Edit Owner
oded	Oded Kleifeld	Read
bosco	Bosco Ho	Read
synchrotron	Synchrotron Test	Read Edit Owner

[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.

Temporary Link	Expiry	Granted By
Right click.. copy link	15th November 2012	mytardis
Right click.. copy link	7th December 2012	mytardis

[Create New Temporary Link](#)

Sharing

Users

Users who have a share in this experiment:

Username	Name	Permissions
mytardis	Steve Androulakis	Read Edit Owner
oded	Oded Kleifeld	Read
bosco	Bosco Ho	Read
synchrotron	Synchrotron Test	Read E

[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 pri

Temporary access links provide full access to recipients regardless of an experim

Temporary Link	Expiry	Granted By
Right click.. copy link	15th November 2012	mytardis
Right click.. copy link	7th December 2012	mytardis

[Create New Temporary Link](#)

Public Access

Step 1: Change Public Access:

Public access

Step 2: Select a license:

Use

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.

Selected

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

Publish your research, then publish your raw data with us.

The Australian Synchrotron is working with researchers to open access to datasets associated with publications. This process is entirely opt-in for researchers and gives permission for the Synchrotron to host your data publicly.

The screenshot shows the Store.Synchrotron website interface. At the top, there is a navigation bar with links for Home, About, My Data, Public Data, Stats, and Help, along with a Log In button. The main content area features the title 'Derivatives for structure solution of the peripheral stalk from T.thermophilus A-ATPase' by Alastair Stewart and Daniela Stock. Below the title, there are statistics: 4 datasets, 576 files, 10.1 GB, and a 'Public' status. A 'Hide Description' button is visible. The description text reads: 'Rotary ATPases couple ATP hydrolysis/synthesis with proton translocation across biological membranes and so are central components of the biological energy conversion machinery. Their peripheral stalks are essential components that counteract torque generated by rotation of the central stalk during ATP synthesis or hydrolysis. These datasets are derivatives of the peripheral stalk from T.thermophilus A-ATPase. Native crystals were soaked in Lutetium(III) acetate (2K7c_3_###.img) and Dysprosium(III) chloride (2K3#####.img). Resulting maps were used to create the pdb model 3V6I. The model was used to identify bending and twisting motions inherent within the structure that accommodate movements within the ATPase.' Below the text is a 3D ribbon diagram of the protein structure. To the right, a '4 Datasets' section is shown with a 'Download Selected' button and a search input field. Three dataset entries are listed: 2K3e_5 (94 files, 1.7 GB), 2K3d_4 (160 files, 3.2 GB), and 2K7c_3 (120 files, 2.1 GB). Each entry includes a folder icon, a download icon, and a thumbnail image of the data.

Data is made available via MyTardis at [Store.Synchrotron.org.au](https://store.synchrotron.org.au)

What do you need to do?

Contact
mxtickets@synchrotron.org.au

You'll be emailed a short form to fill in basic information about open access to your datasets associated with PDBs and publications.

Once data is made available via our service, you will receive a DOI to this data for citation.

Note: The data you're collecting today is also archived privately and securely via this system.

Log in to <https://store.synchrotron.org.au> with your Australian Synchrotron credentials to browse and download your data.

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)



David.Groenewegen@monash.edu

@groenewegendave

THANKS – QUESTIONS?