

MyTardis: FAIR data management for instrument data

Amr Hassan

eResearch Delivery Leader

Monash eResearch Centre

Wojtek Goscinski

Associate Director

Monash eResearch Centre

In collaboration with:

Wojtek Goscinski, Lance Wilson, James Wettenhall, Manish Kumar, Samitha Amarapathy, Keith Schulze, Grischa Meyer, Steve Androulakis, Jason Rigby, Andrew Perry, Anitha Kannan, Andrew Mehnert, Andrew Janke



Today's schedule

MyTardis: FAIR Data Management for Instrument Data

Wojtek J. Goscinski, Amr Hassan, Andrew Janke, Andrew Mehnert, Aswin Narayanan, Dean Taylor, James M. Wettenhall, Jonathan Knispel, Keith E. Schulze, Lance Wilson, Manish Kumar, Samitha Amarapathy

Schedule: <https://goo.gl/knft4D>

Time	Topic	Speaker
9:00 AM - 9:30 AM	Introduction to MyTardis and instrument integration at Monash	Wojtek Goscinski (Monash)
9:30 AM - 10:00 AM	Developing and Operating MyTardis at the University of Queensland and NIF	Aswin Narayanan (UQ)
10:00 AM - 10:30 AM	Developing and Operating MyTardis at the University of Western Australia NIF Trusted Data Repositories	Andrew Mehnert (UWA)
10:30 AM - 11:00 AM	Morning Tea (Provided)	
11:00 AM - 12:00 PM	MyTardis Architecture and Operations	James Wettenhall Manish Kumar (Monash)
12:00 PM - 12:30 PM	Future Plans and Architecture for MyTardis	Amr Hassan (Monash)
12:30 PM - 1:30 PM	Lunch	
1:30 PM - 3:00 PM	MyTardis Demo: <ul style="list-style-type: none">• Docker Deployment of MyTardis• Deployment using Salt Stack (NeCTAR Cloud)• How to contribute to MyTardis?	James Wettenhall Manish Kumar UQ Team UWA Team
3:00 PM - 3:30 PM	Afternoon Tea (Provided)	
3:30 PM - 4:00 PM	MyTardis Demo (Cont.) <ul style="list-style-type: none">• Configuring for scalability	James Wettenhall Manish Kumar UQ Team UWATeam
4:00 PM - 5:00 PM	Discussion - FAIR Instrument Integration Future: Challenges and planning	All workshop participants

MyTardis

Open Source Data Management for Instruments

Reliable and Secure Web-based Solution to address the challenges of data storage, data access, collaboration and data publication.

Deployments

- Store.Monash
- ImageTrove (National Imaging Facility)
- University of West Australia (CMCA)
- University of Queensland
- University of New South Wales
- University of Newcastle
- Others

Store.Monash Data Store

Please login to see your experiment data.

The 8 most recent public experiments

- X-ray diffraction images of trypsin crystals soaked with SFTI-TCTR(N12,N14) cyclopeptide inhibitor**
Blake T. Riley, Xingchen Chen, David E. Hoke, Ashley M. Buckle, Jonathan M. Harris
15th December 2017 | 1 | 90
[Download data as .tar](#)
- CT dose reduction factors in the thousands using X-ray phase contrast**
Marcus J. Kitchen, Genevieve A. Buckley, Timur E. Gureyev, Megan J. Wallace, Nico Andres-Thio, Kentaro Uesugi, Naoto Yagi, and Stuart B. Hooper
22nd November 2017 | 25 | 122
[Download data as .tar](#)
- Parameters for calculating absolute U, B, V, R, I, J, K, u, g, r, i, z magnitudes from single observed colors based on apparent Bw, R, I, J, Ks, [3.6 micron] and [4.5 micron] magnitudes at redshifts z<1.2 using the method of Beare et al. (2014).**
Richard Beare, Michael Brown, Kevin Pimblett
30th January 2017 | 1 | 46
[Download data as .tar](#)
- Parameters for calculating absolute Sloan u, g, r, i, z magnitudes from single observed Sloan colors at redshifts z<0.5 using the method of Beare et al. (2014).**
Richard Beare, Michael Brown, Kevin Pimblett
30th January 2017 | 1 | 11
[Download data as .tar](#)
- Jamadar, Thienel, Karayanidis 2014 ALE Meta-Analysis Methods and Results**
Sharna Jamadar, Renate Thienel, Frini Karayanidis
30th January 2017 | 1 | 1
[Download data as .tar](#)
- An Atlas of Galaxy Spectral Energy Distributions From The UV to the Mid-Infrared**
Michael J. I. Brown, John Moustakas, J.-D. T. Smith, Elisabete da Cunha, T. H. Jarrett, Masatoshi Imanishi, Lee Armus, Bernhard R. Brandl, J. E. G. Peek
30th January 2017 | 130 | 3199
[Download data as .tar](#)

MyTardis Home About My Data Public Data Stats Help Register Log In

Experiment

An Atlas of Galaxy Spectral Energy Distributions from the Ultraviolet to The Mid-Infrared (BETA)

Michael J. I. Brown, John Moustakas, J.-D. T. Smith, Elisabete da Cunha, T. H. Jarrett, Masatoshi Imanishi, Lee Armus, Bernhard R. Brandl, J. E. G. Peek

130 Datasets

Just start typing to filter datasets based on descriptions

- An Atlas of Galaxy SEDs** | 130 | 10.1 MB
- Arp_118** | 10 | 200.4 MB
- Arp_256_N** | 23

Toggle Full Description

This is the Beta version of "An Atlas of Galaxy Spectral Energy Distributions From The UV to the Mid-Infrared" and will be updated once the paper is accepted for publication. The first folder contains all the

Description Metadata Sharing Transfer Datasets

Institution Monash University

Licensing This experiment data is licensed under Creative Commons Attribution 3.0 Australia (CC BY 3.0)

Administrators Michael Brown

Download All [A Tar](#)

Facility Overview

Test Facility

Filter by: user name, experiment, instrument, X Clear filter

Owner	Group	Experiment	Dataset description	Instrument	Created	Download	View
testfacility	testfacility	Test Microscope - Test User1	Charybdomones	Test Microscope	2015-03-12 11:21PM	Download	View
testfacility	testfacility	Test Microscope - Test User1	Biological TEM	Test Microscope	2015-03-12 11:21PM	Download	View
testfacility	testfacility	Test Microscope - Test User1	Bone marrow TEM	Test Microscope	2015-03-12 11:21PM	Download	View
testfacility	testfacility	Test Microscope - Test User1	Blood cells SEM	Test Microscope	2015-03-12 11:21PM	Download	View
testfacility	testfacility	Test Microscope - Test User1	Amorphophallus stemon SEM	Test Microscope	2015-03-10 8:24PM	Download	View

File name | Size | Created | Verified?

01_AMS1_Seq2011.1 (1).jpg	6 MB	2015-03-10 12:04PM	Yes
01_AMS1_Seq2011.1.jpg	6 MB	2015-03-10 12:04PM	Yes
01_AMS1_Seq2011.2.jpg	5 MB	2015-03-10 12:04PM	Yes
01_AMS1_Seq2011.3.jpg	4 MB	2015-03-10 12:04PM	Yes
01_AMS1_Seq2011.4.jpg	6 MB	2015-03-10 12:04PM	Yes
01_AMS1_Seq2011.2.jpg	5 MB	2015-03-10 12:04PM	Yes

Load more (showing 5 of 5)

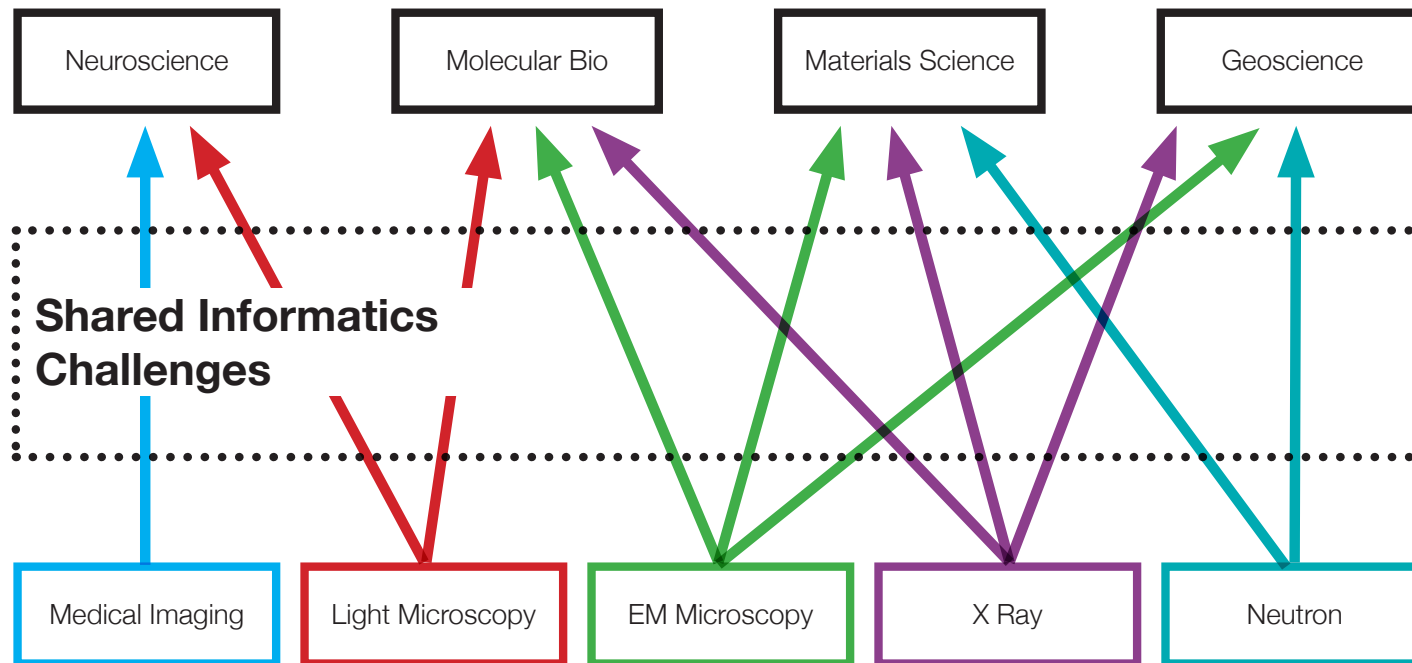
“Here is your CD of data...”

to

“Your data is moving up to a data management system in the cloud where you have access to a range of tools and services to start your data analysis”

A Collaborative Australian Characterisation Informatics Strategy

Characterisation has become a capability where informatics infrastructure, expertise and best practice is essential to turning data into new discoveries.



The Australian characterisation community provides a wide range of techniques that are applied across a variety of scientific domains. Common across these are a set of shared informatics challenges.

A Collaborative Australian Characterisation Informatics Strategy

Challenge:

Scale and complexity

- Data volumes are increasing with new detector technology
- Processing requirements are increasing – the “science is being affected by compute”
- Opportunity offered by multi-site, Australia-scale data capture introduces specific challenges
- Custom and specialised instrumentation requires custom workflows
- Increasing need to perform analysis “in-experiment”
- Analysis of data from multiple instruments requires specialised skills and familiarity with the data

Requires:

A national infrastructure program that supports:

- **Community driven instrument integration and data management initiatives** to capture data from the point of generation
- **Rich online environments for characterisation in the cloud and on HPC platforms**
- **Simple and seamless access across instruments, repositories and analysis environments**
- Programs for **specialised and big data producing instruments**

A Collaborative Australian Characterisation Informatics Strategy

Challenge:

Working with digital objects

- Data remains unpublished, is difficult to reuse, and it is often unclear whether it can be trusted
- Data curation is a second priority to publication and data is often non reusable
- Research software is often closed source and impossible to validate, can be challenging to newcomers and is often very specific to particular problems
- Research outcomes are difficult to validate and are often unreproducible
- To increase return on investment research outputs need to be machine readable

Requires:

Making Characterisation digital objects Findable, Accessible, Interoperable, and Reusable (FAIR)

To achieve this requires:

- **Community** efforts to increase application of FAIR principles
- **Coordination** across Australia to provide leadership and organisation
- **Commitment** by data producers, in partnership with research communities and tools developers to increase uptake of FAIR principles

A Collaborative Australian Characterisation Informatics Strategy

Challenge:

Expertise is rare

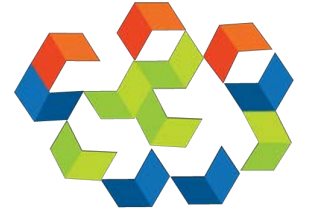
- The characterisation community is increasingly reliant on data-science skills
- Digital expertise coupled with applied characterisation knowledge is rare
- Cross modality analysis requires multiple areas of expertise to facilitate new insights and discoveries
- Where knowledge is available it is often in accessible beyond a local node or institution

Requires:

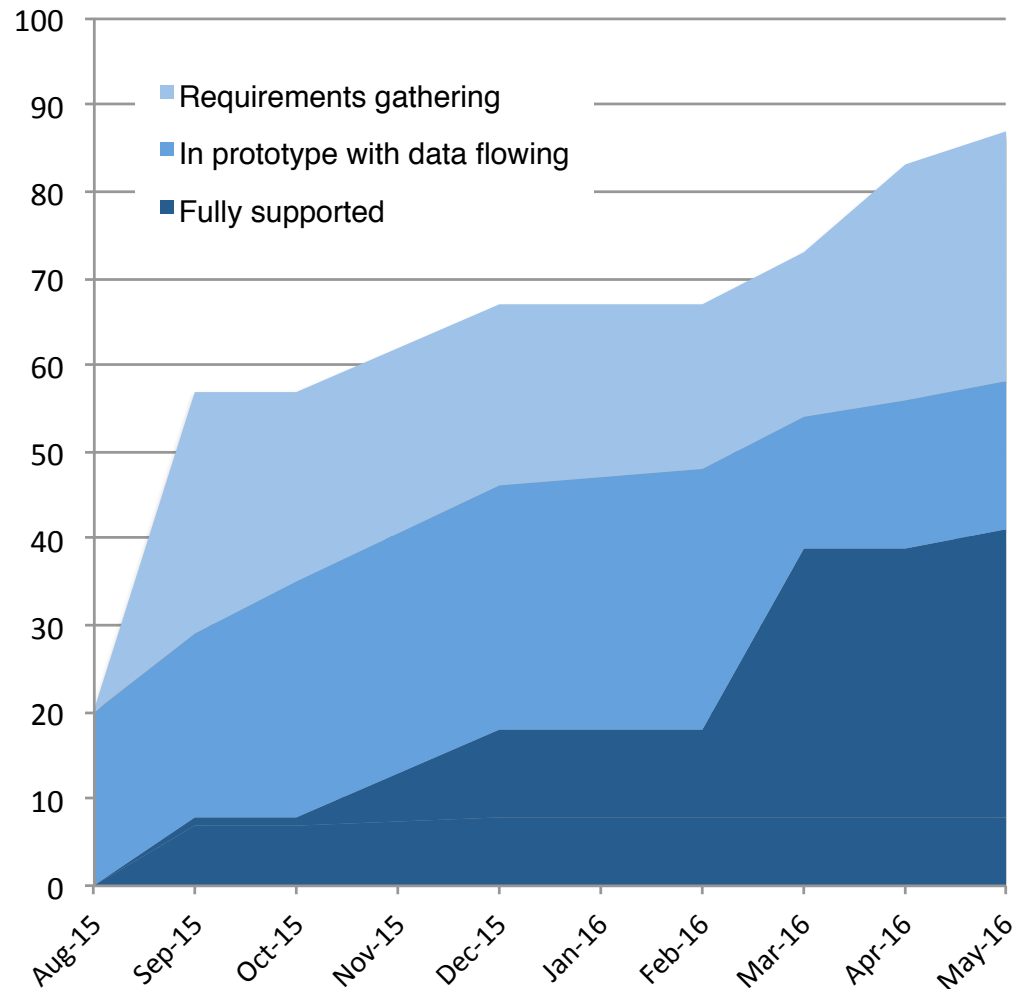
A national program to spread knowledge and underpin change, which includes:

- **National training** to uplift data skills across characterisation users
- A **national network of characterisation informatics experts** with expertise in research software engineering, and specialist skills in specific modalities, as part of an overarching Australian characterisation experts network

National Data Instrument Integration



Instruments integrated into research cloud for neuroscience research
Leveraging: XNAT, MyTardis, Loris, Omero, and others



Animal MRI Facility,
Florey Neuroscience Institutes
Australian Centre for Microscopy & Microanalysis, USydney
Australian Synchrotron,
Biological Optical Microscope Platform (MDHS),
UoM Biological Resources Imaging Laboratory,
UNSW Australian Centre for Neutron Scattering,
ANSTO
Center for Advanced Imaging, UQ
Centre for Microscopy and Microanalysis UQ
Centre for Microscopy, Characterisation and Analysis, UWA
Florey, Melbourne Brain Centre
FlowCore, Monash University
Melbourne Brain Centre Imaging Unit, UoM

MicroNano Research Facility, RMIT
Monash Biomedical Imaging
Monash Biomedical Proteomics Facility
Monash Injury Research Institute
Monash Micro Imaging
Monash Micro Imaging (AMREP)
University of Newcastle, Light Sheet Microscopy
Queensland Brain Institute
Single Molecule Science, UNSW
Royal Children's Hospital
Royal Melbourne Hospital
St Vincents Hospital
The Clive and Vera Ramaciotti Centre for Structural Cryo-Electron Microscopy
X-ray Microscopy Facility for Imaging Geomaterials (XMFIG), Monash

The total estimated capital value of the instruments identified by this project is **\$242M+** across **100 instruments** at **22 facilities** using a variety of tools and services

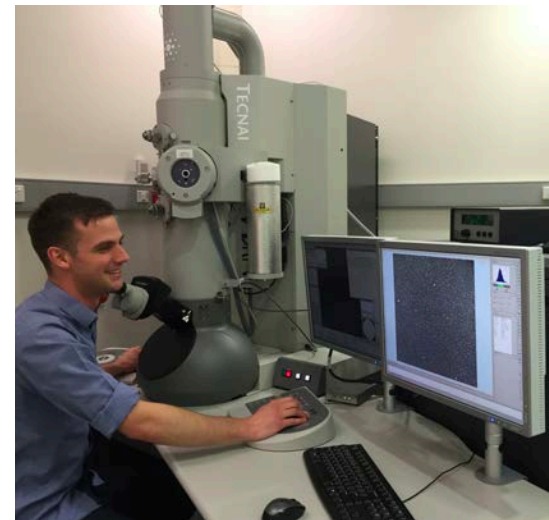
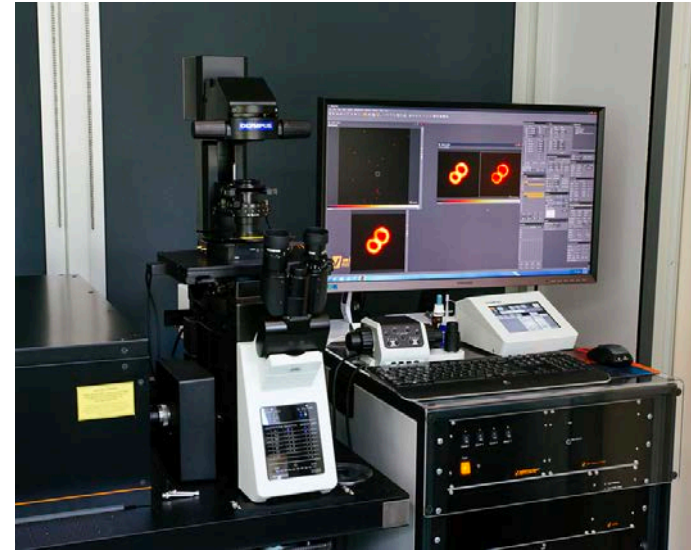
Collected **petabytes** of research data

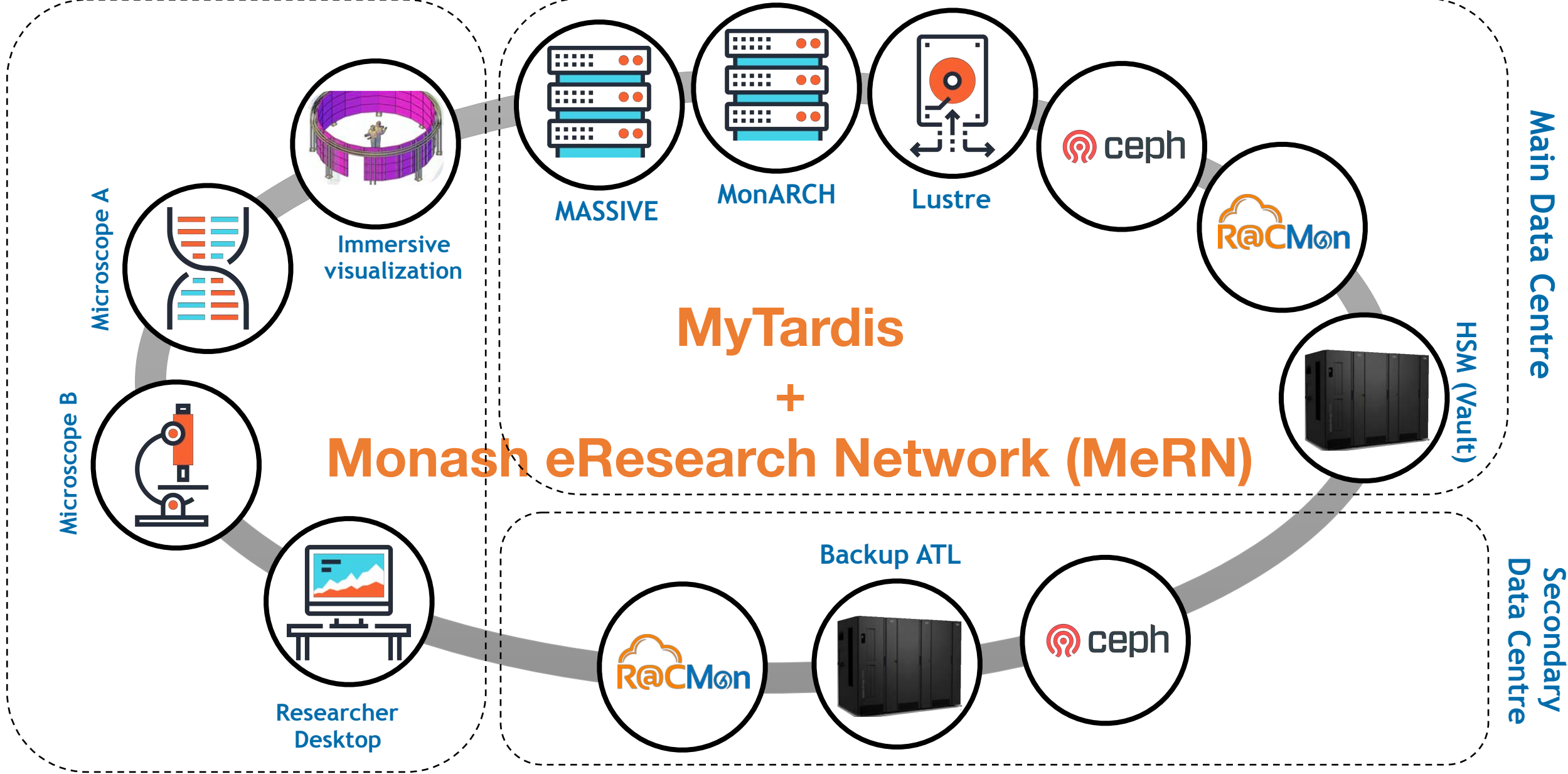
Monash Clayton Precinct

Largest Australian Scientific Data Producing Precinct.

Globally unique imaging infrastructure - from an atom to a whole animal:

- Imaging and Medical Beamline (ANSTO Synchrotron)
- X-ray diffraction beamline (ANSTO Synchrotron)
- Monash Biomedical Imaging
- Monash Centre for Electron Microscopy
- Monash Micro Imaging
- Ramaciotti Centre for Cryo-Electron Microscopy





MyTARDIS

A data management system for Microscopes

Capture data at its source:

Reliable Data Management Solution to ensure the data is Findable and Accessible

Add value to your Microscope data:

Extensible data processing architecture
Partner with users to develop unique views and filtering of data

Id	Folder	Subdirectory	Filename	File Size	Status	Progress	Message
1	Dataset test1		Andromeda Galaxy.jpg	82 KB	✓	100%	Upload complete!
2	Dataset test1		CrabNebula.jpg	178 KB	✓	100%	Upload complete!
3	Dataset test4		image4-5.jpeg	25 KB	✓	100%	Upload complete!
4	Dataset test2		image4-2.jpeg	11 KB	✓	100%	Upload complete!
5	Dataset test2		image4-1.jpeg	20 KB	✓	100%	Upload complete!
6	Dataset test2		03_CLEM017_Z...GFP_A3_001.tif	55 MB	✓	100%	Upload complete!
7	Dataset test2		image4-3.jpeg	25 KB	✓	100%	Upload complete!
8	Dataset test3		image4-4.jpeg	9 KB	✓	100%	Upload complete!
9	Dataset test3		image4-6.jpeg		✓	100%	100 % uploaded

Raw Read Statistics

Sample name	Index	Lane	Read	Number of reads	Percent gc	Read length
QQInputF2	AACCAG	2	1	2,000	51.0	51
QQH4K4F2	TGGTGA	2	1	2,000	51.0	51
QQH4K9F2	AGTGAG	2	1	2,000	51.0	51

FastQC Summary

Please note: failed QC checks don't always indicate a problem with the data! Please read the [FastQC documentation](#).

FastQC version: 0.11.4

Sam...	Basic Statistics	Per base sequence quality	Per sequence quality scores	Per base sequence content	Per sequence GC content	Per base N content	Sequence Length Distribution	Sequence Duplication Levels	Overrepresented sequences	Adapter Content	Kmer Content
QQInputF2 (L2, R1)	✓	✓	⚠	✓	✗	✓	✓	✓	✓	✓	✓
QQH4K4F2 (L2, R1)	✓	✓	⚠	✓	✗	✓	✓	✓	✓	✓	✓
QQH4K9F2	✓	✓	⚠	✓	✗	✓	✓	✓	✓	✓	✓

MyTARDIS

A data management system for Microscopes

Distribute data without a fuss:

Securely and automatically shifting data from instrument to repository, accessible by the right users or user groups.

Support for   Coming in Version 4.0

Data publications:

Facilities can open selected data to the world, adding key information relating to research publications, derivative data and downstream analysis.

Sharing

Users

Users who have a share in this experiment:

Username	Name	Permissions
bloem	BioEM Facility	Read Edit Owner
wettenhj	James Wettenhall	Read Owner
steveand	Steve Androulakis	Read

[Change User Sharing](#)

Publication

The crystal structure of a homodimeric Pseudomonas glyoxalase I enzyme reveals asymmetric metallation commensurate with half-of-sites activity

Rohan Bythell-Douglas, Charles Bond


2 | 720 | 12.7 GB | 5th November 2015 | Public

[Show Description](#)

The Zn inactive class of glyoxalase I (Glo1) enzymes are metalloenzymes that are typically homodimeric with two metal-dependent active sites. While the two active

[Description](#) [Metadata](#)

Institution University of Western Australia



Licensing This experiment data is licensed under Creative Commons Attribution 3.0 Australia (CC BY 3.0) 

Download All [TAR](#) [SFTP](#)

2 Datasets

[Download Selected](#)

Just start typing to filter datasets based on descriptions

-  **Raw data for D3_1**
Folder: frames 360 6.3 GB
-  **Raw data for D3_2**
Folder: frames 360 6.3 GB

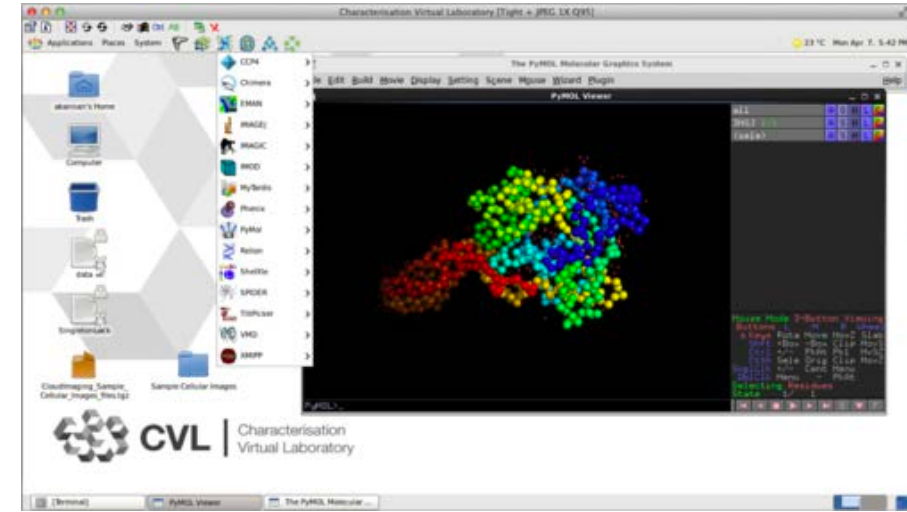
Hosted at Monash University on the NeCTAR Cloud Powered by MyTardis

MyTARDIS

A data management system for Microscopes

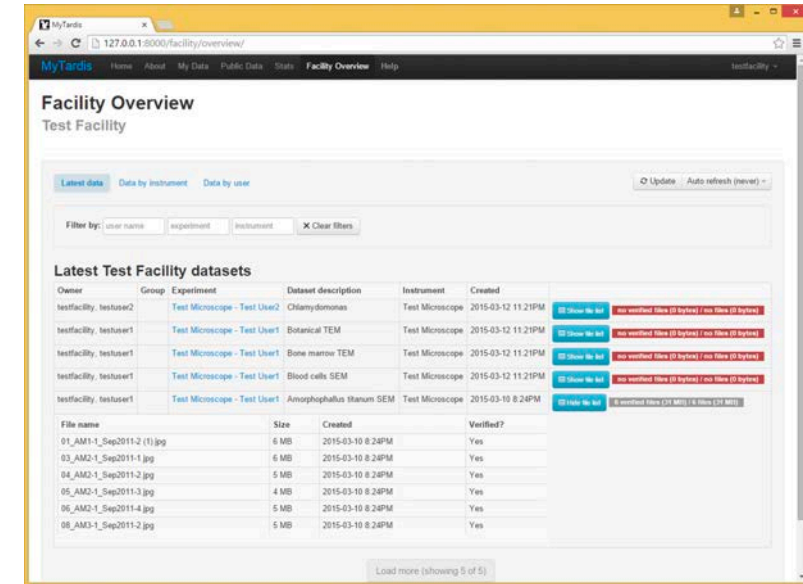
Push data to Compute:

'Push To' feature allows users to seamlessly transfer data directly from MyTardis to HPC, or virtual laboratories for further analysis and processing.



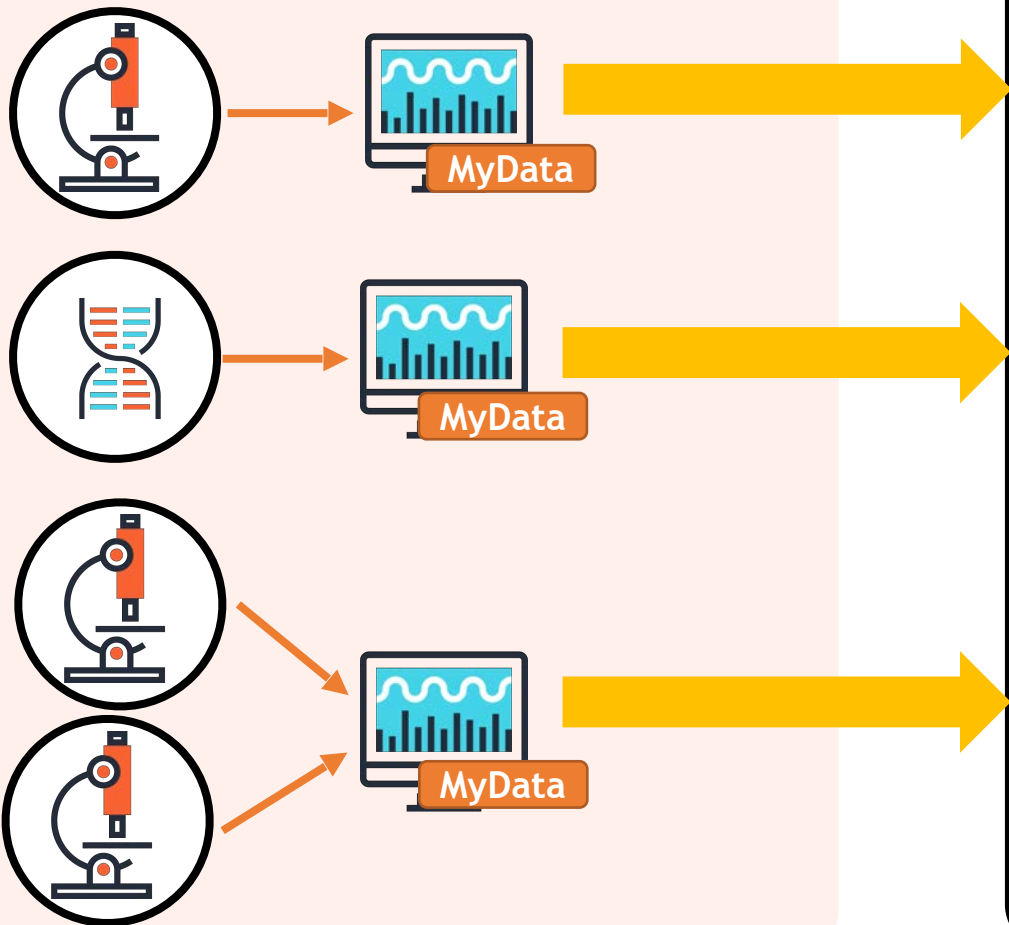
Facility View:

provides facility staff with a total view of data generation from all instruments. Filter by user, instrument or browse the summary page for the ultimate view of data activity within the facility. It enables facility staff to fix ownership issues for un-owned data, and share data with external researchers.



MyTARDIS

Microscope Facility



HPC Facility, and/or Public/Private Cloud

MyTARDIS



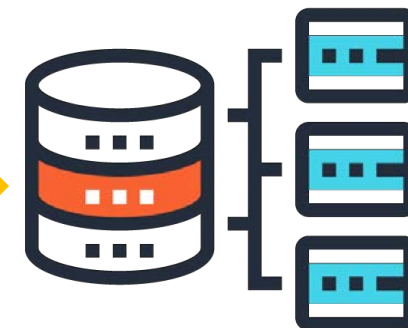
Protect



Share



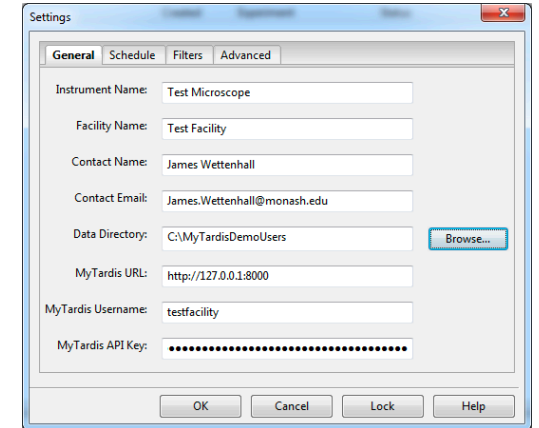
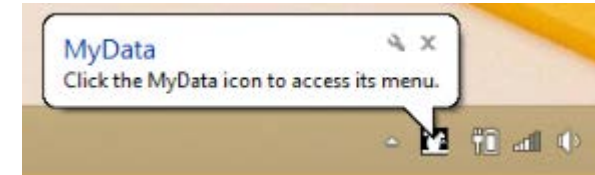
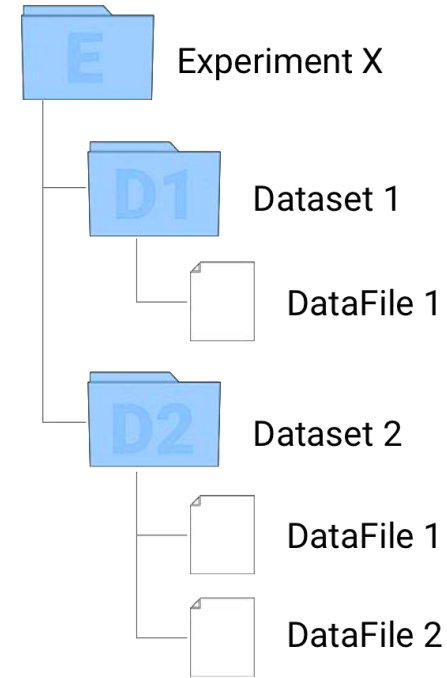
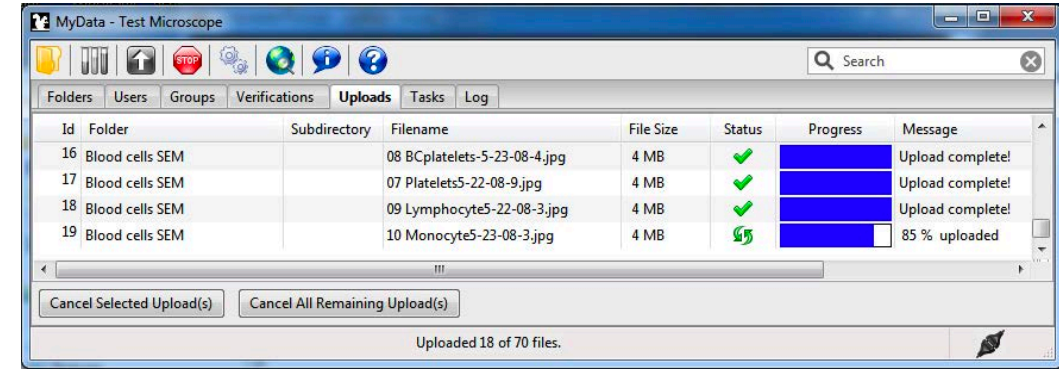
Publish



Multiple Storage back-ends

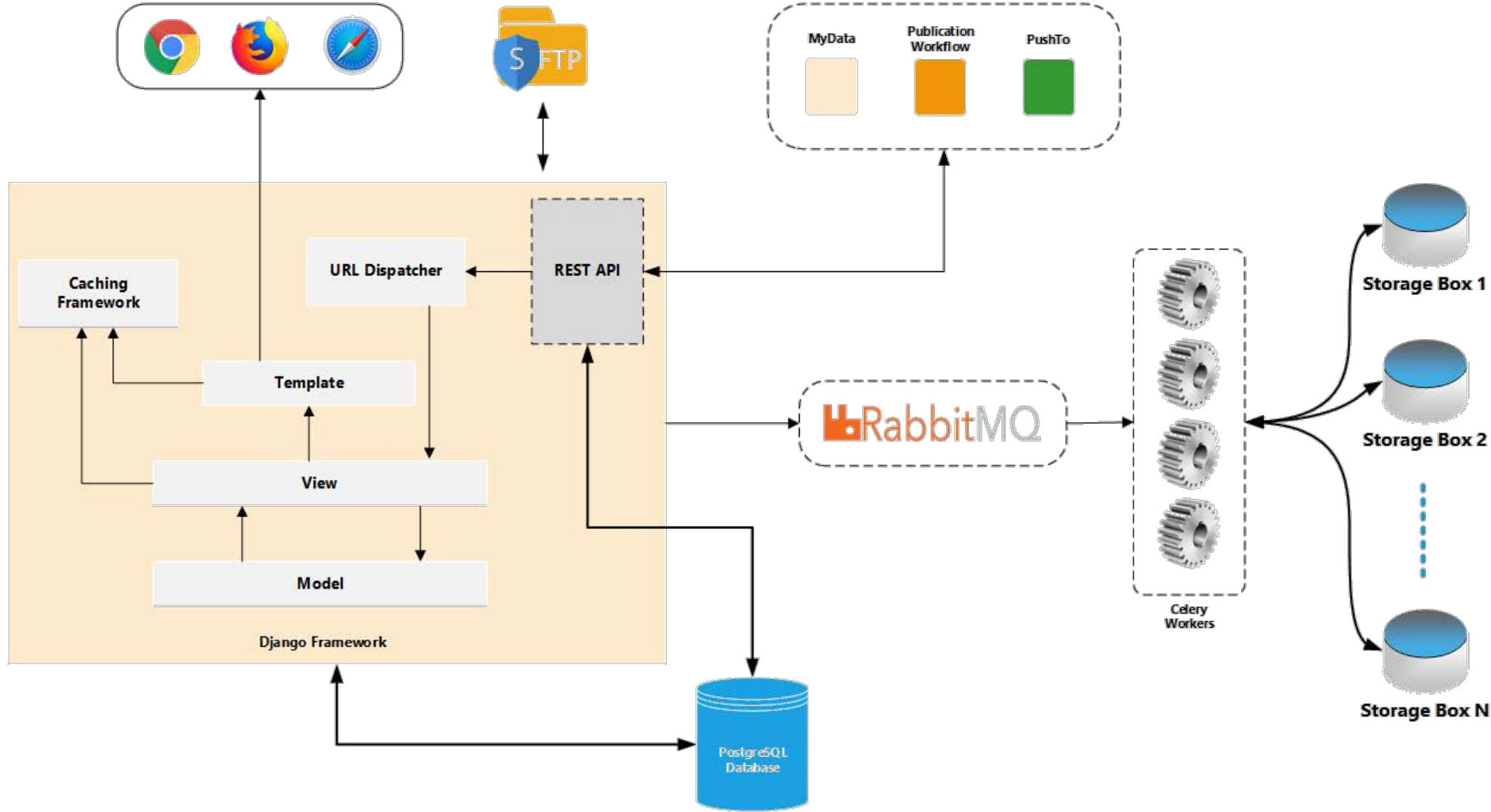
MyData

- Microscope integration in the hands of the facility / instrument
- An easy mechanism for Microscope data to be captured, stored, managed etc.
- Work with partners to create a range of configurable options



<https://mydata.readthedocs.io/>

MyTardis Architecture



- Local disk storage
- Mounted network drives (NFS, CIFS, SSHFS)
- S3 object storage
- Swift object storage
- SFTP accessible locations

Store.Monash

Microscope Facility



Data captured at the source with **MyData**

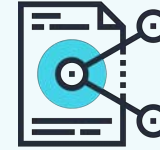


Monash Research Cloud

Store.Monash



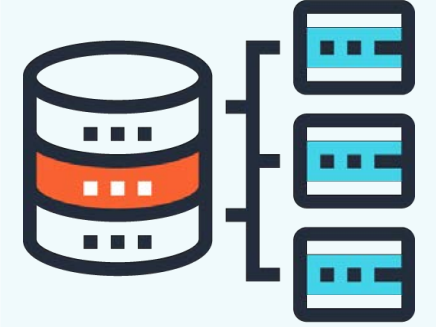
Protect



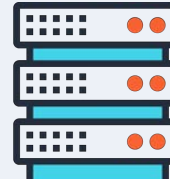
Share



Publish



Multiple Storage back-ends



Store.Monash

Monash University instance of the **MyTardis**



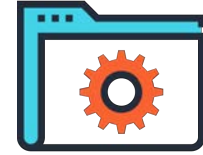
54

Instruments



0.6 PB

Managed Storage



55,000

Datasets



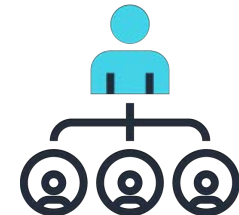
8 Million

Data Files



1568

Users



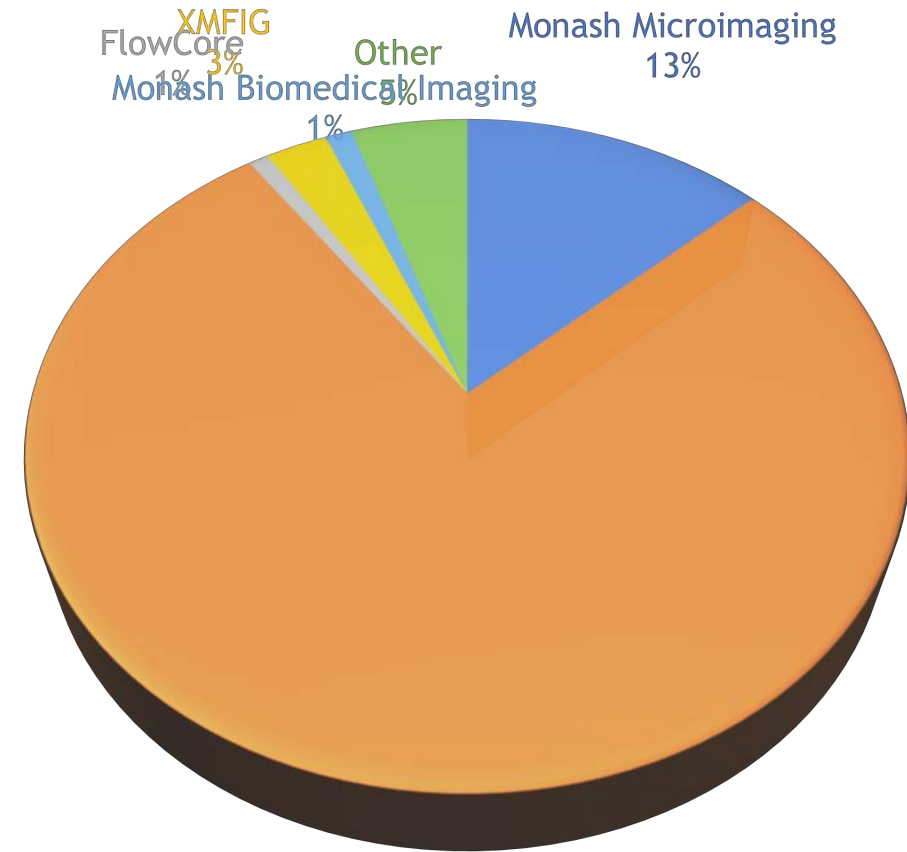
856

Groups

Store.Monash

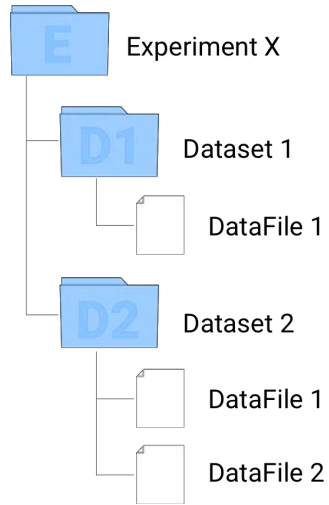
Types of instruments integrated:

- **Microscopy**
 - Advanced light microscope
 - Fluorescence and Confocal microscope, Multiphoton microscope
 - Cryo-electron microscope,
 - Scanning electron microscope
 - Transmission electron microscope
- **Flow Cytometry**
 - Cell Sorter, Analysers
- **Imaging**
 - X-ray microscope, MRI, PET/MRI, PET/CT
- **Proteomics**
 - Mass Spectrometer
- **Gene Sequencers**



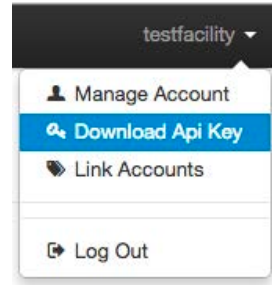
Data Usage

Managing your Microscope Data with MyTardis



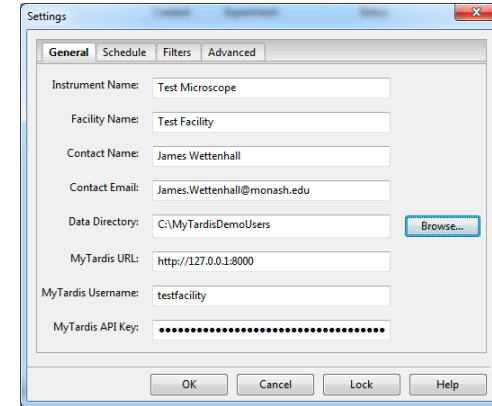
How your data is organized?

1



Add Instrument to MyTardis

2




Configure your MyData

3

 <http://www.mytardis.org/>

 <https://github.com/mytardis/mytardis>

 <https://mytardis.readthedocs.io>

 store.star.help@monash.edu

MyTardis has been supported by:



Today's schedule

MyTardis: FAIR Data Management for Instrument Data

Wojtek J. Goscinski, Amr Hassan, Andrew Janke, Andrew Mehnert, Aswin Narayanan, Dean Taylor, James M. Wettenhall, Jonathan Knispel, Keith E. Schulze, Lance Wilson, Manish Kumar, Samitha Amarapathy

Schedule: <https://goo.gl/knft4D>

Time	Topic	Speaker
9:00 AM - 9:30 AM	Introduction to MyTardis and instrument integration at Monash	Wojtek Goscinski (Monash)
9:30 AM - 10:00 AM	Developing and Operating MyTardis at the University of Queensland and NIF	Aswin Narayanan (UQ)
10:00 AM - 10:30 AM	Developing and Operating MyTardis at the University of Western Australia NIF Trusted Data Repositories	Andrew Mehnert (UWA)
10:30 AM - 11:00 AM	Morning Tea (Provided)	
11:00 AM - 12:00 PM	MyTardis Architecture and Operations	James Wettenhall Manish Kumar (Monash)
12:00 PM - 12:30 PM	Future Plans and Architecture for MyTardis	Amr Hassan (Monash)
12:30 PM - 1:30 PM	Lunch	
1:30 PM - 3:00 PM	MyTardis Demo: <ul style="list-style-type: none">• Docker Deployment of MyTardis• Deployment using Salt Stack (NeCTAR Cloud)• How to contribute to MyTardis?	James Wettenhall Manish Kumar UQ Team UWA Team
3:00 PM - 3:30 PM	Afternoon Tea (Provided)	
3:30 PM - 4:00 PM	MyTardis Demo (Cont.) <ul style="list-style-type: none">• Configuring for scalability	James Wettenhall Manish Kumar UQ Team UWATeam
4:00 PM - 5:00 PM	Discussion - FAIR Instrument Integration Future: Challenges and planning	All workshop participants