

Managing instrument data in a scalable way

Grischa Meyer - Monash eResearch Centre

Data capture with MyTardis

- At the instrument, MyData captures data automatically
- To be transferred and stored securely
- It can then be accessed by authorised researchers immediately
- It can be shared with collaborators around the world
- And published
- And integrated with other publication services



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



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 Datasets
Download Selected

Just start typing to filter datasets based on descriptions

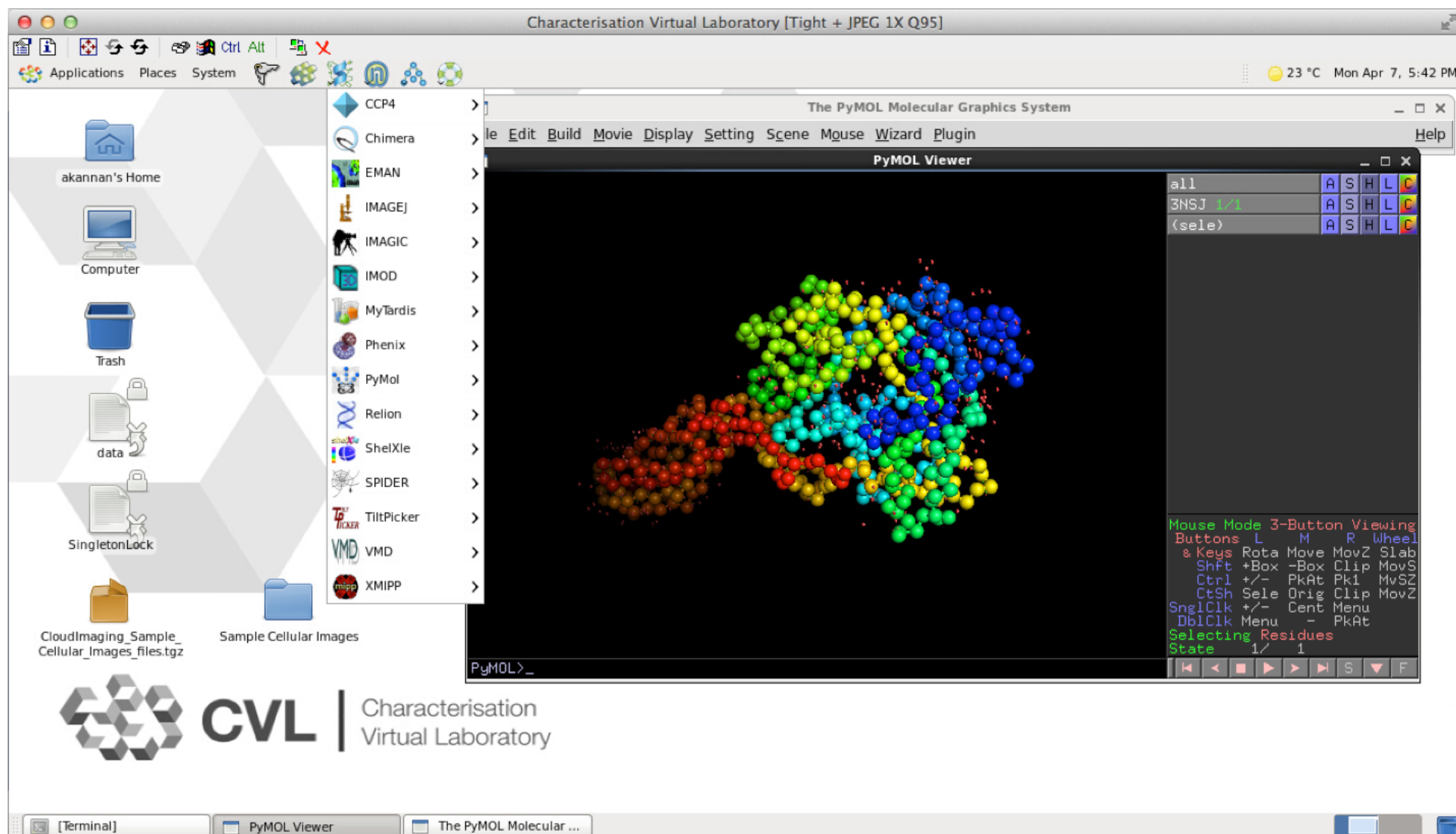
Raw data for D3_1
Folder: frames
360 frames
6.3 GB

Raw data for D3_2
Folder: frames
360 frames
6.3 GB

Hosted at Monash University on the NeCTAR Cloud
Powered by MyTardis

Data processing

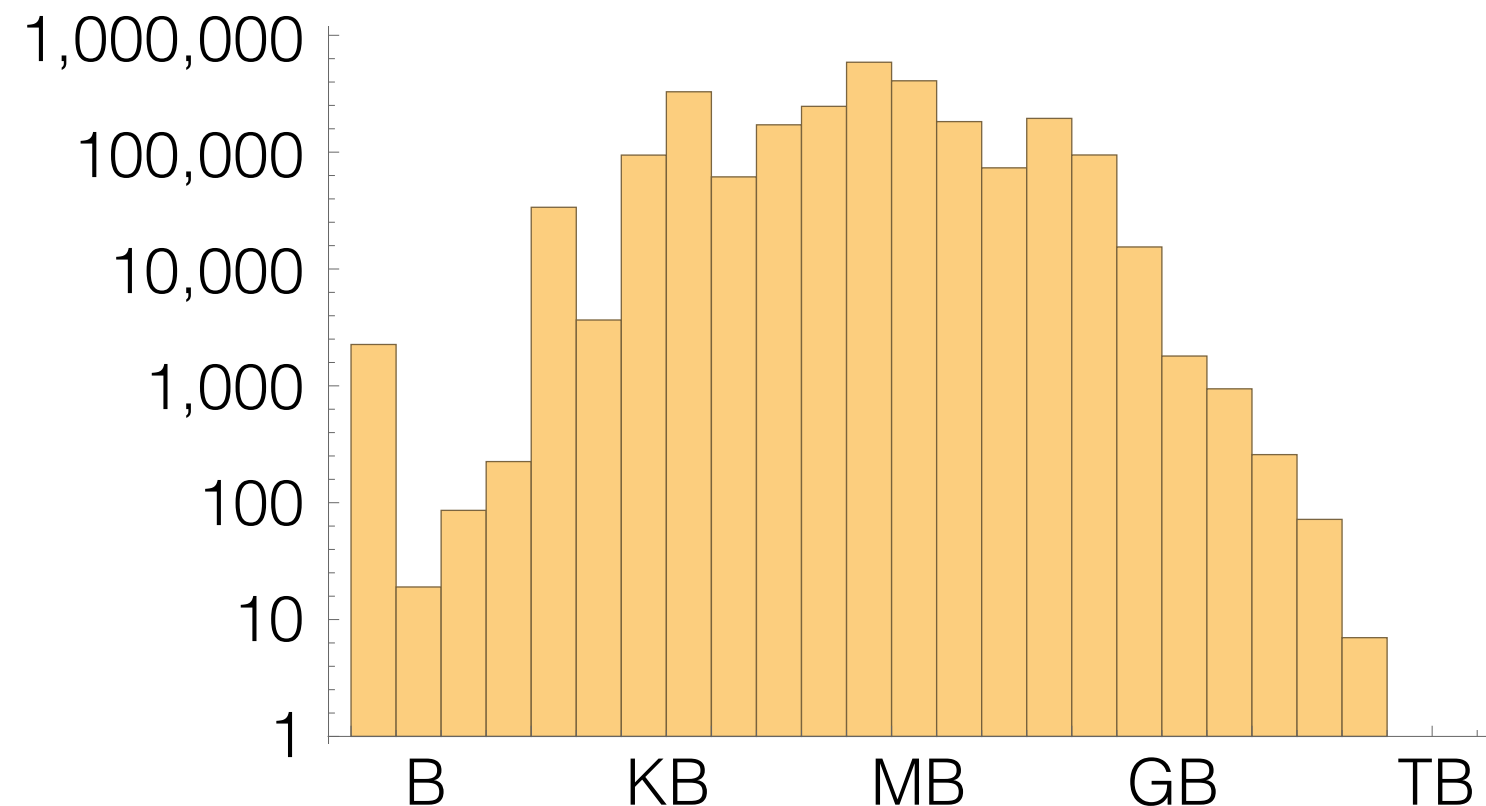
- MyTardis at Monash is integrated with the MASSIVE HPC facility and the CVL virtual desktops
- Selected data can be pushed to these resources with the press of a button on the web interface



MyTardis at Monash

- Store.Monash
- So far 8 facilities, 38 instruments, e.g.:
 - Microscopy
 - Cryo-EM
 - MRIs
 - Bioinformatics
- Some more numbers:
 - Experiments: 2,815
 - # of files: 2,502,420
 - Data total: 53.1 TB
 - Average file: 21 MB

File size histogram



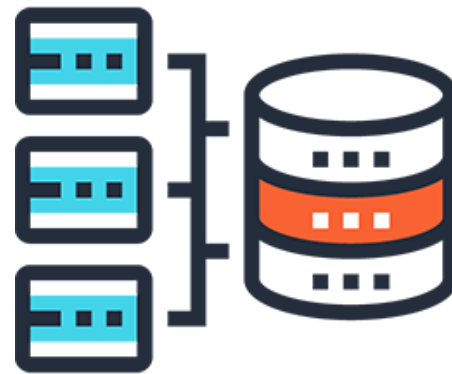
System architecture

- The components of MyTardis can be parallelised and distributed
- Using the Monash zones of the NeCTAR Cloud
- Orchestrated with Heat and SaltStack
- Configuration version controlled, changes reviewed
- Regularly scheduled maintenance windows



Compute & Storage

- 1 SaltStack master
- 2 database servers
- 1 database router
- 1 task queue
- 4+ task workers
- 4+ filter runners
- 8+ web workers
- 2+ SFTP servers
- 2 load balancers



- Dynamic storage handling
- Data migrations
- Currently active backends:
 - Mounted network storage, NFS, SSHFS
 - S3 compatible storage
 - SFTP storage

New <http://mytardis.org>

The screenshot shows a web browser window with the address bar displaying www.mytardis.org. The website has a clean, modern design with a light grey sidebar on the left and a white main content area. The sidebar contains the MYTARDIS logo, navigation links for users, facilities, and developers, and a list of menu items including About, Features, News, Major Deployments, Acknowledgement, and Contact us. The main content area features a header with the title 'MYTARDIS: Research data management for instrument data'. Below this, there are two columns: 'News' and 'Features'. The 'News' column includes a large image of two scientific instruments, a headline 'MyTardis Integrates with Gene Sequencers', a paragraph describing the integration, and a 'Read More' link. The 'Features' column lists six key features, each with an icon and a brief description: Easy Instrument Integration, Discipline specific: MX, imaging, microscopy, genomics, Secure cloud data storage & access, Simple data sharing, Wide range of data formats & supported instruments, and Researcher controlled data publishing.


MYTARDIS

FOR USERS
FOR FACILITIES
FOR DEVELOPERS

About
Features
News
Major Deployments
Acknowledgement
Contact us

MYTARDIS: Research data management for instrument data

News



MyTardis Integrates with Gene Sequencers







There is currently no community-accepted software solution that captures, stores and serves output from gene sequencing experiments. We describe the establishment and production deployment of the MyTardis-Seq system for the Monash Health Translational Precinct Genomics sequencing facility.

This work benefits facility managers and gene sequencer users by providing an automated and structured method to capture, store and share the results of sequencing runs with associated quality reports and metadata.

[Read More →](#)

Jul 7, 2016 · Steve Androulakis

Features

-  Easy Instrument Integration
-  Discipline specific: MX, imaging, microscopy, genomics.
-  Secure cloud data storage & access
-  Simple data sharing
-  Wide range of data formats & supported instruments
-  Researcher controlled data publishing