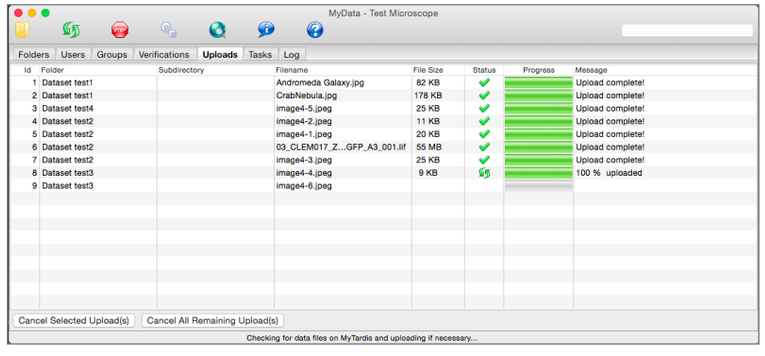
Managing instrument data in a scalable way

Grischa Meyer - Monash eResearch Centre

Data capture with MyTardis

- At the instrument, MyData captures data automatically



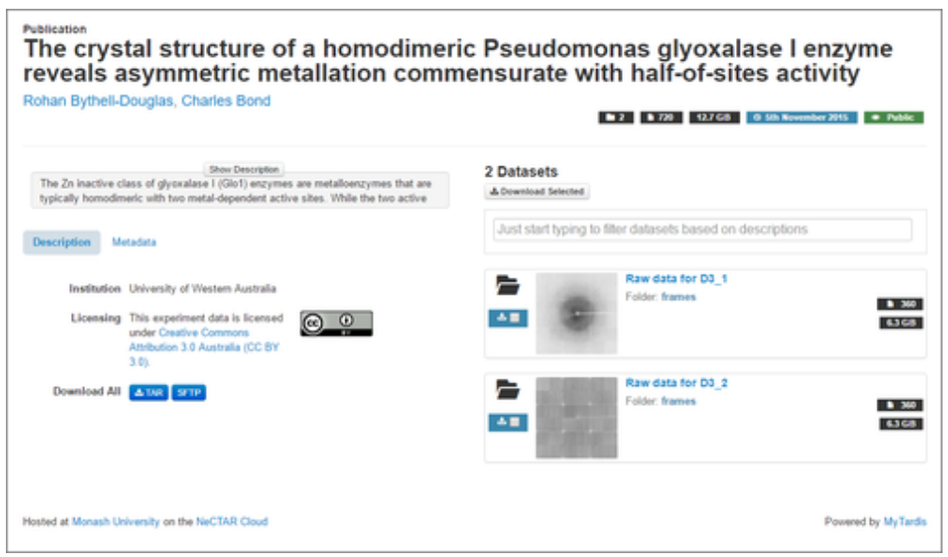
- To be transferred and stored securely

- 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

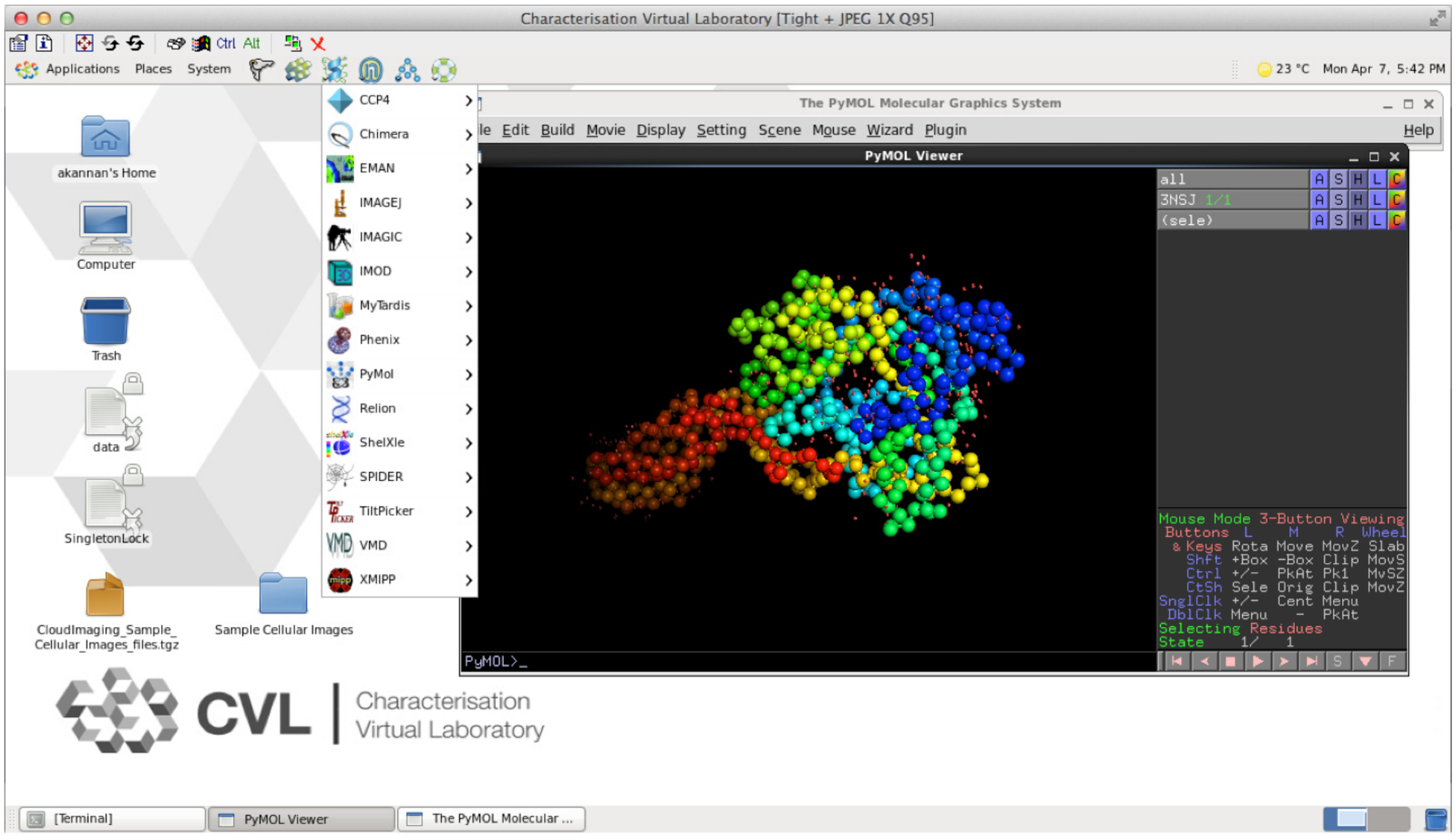


figshare

Data processing

- MyTardis at Monash is integrated with the MASSIVE HPC facility and the CVL virtual desktops

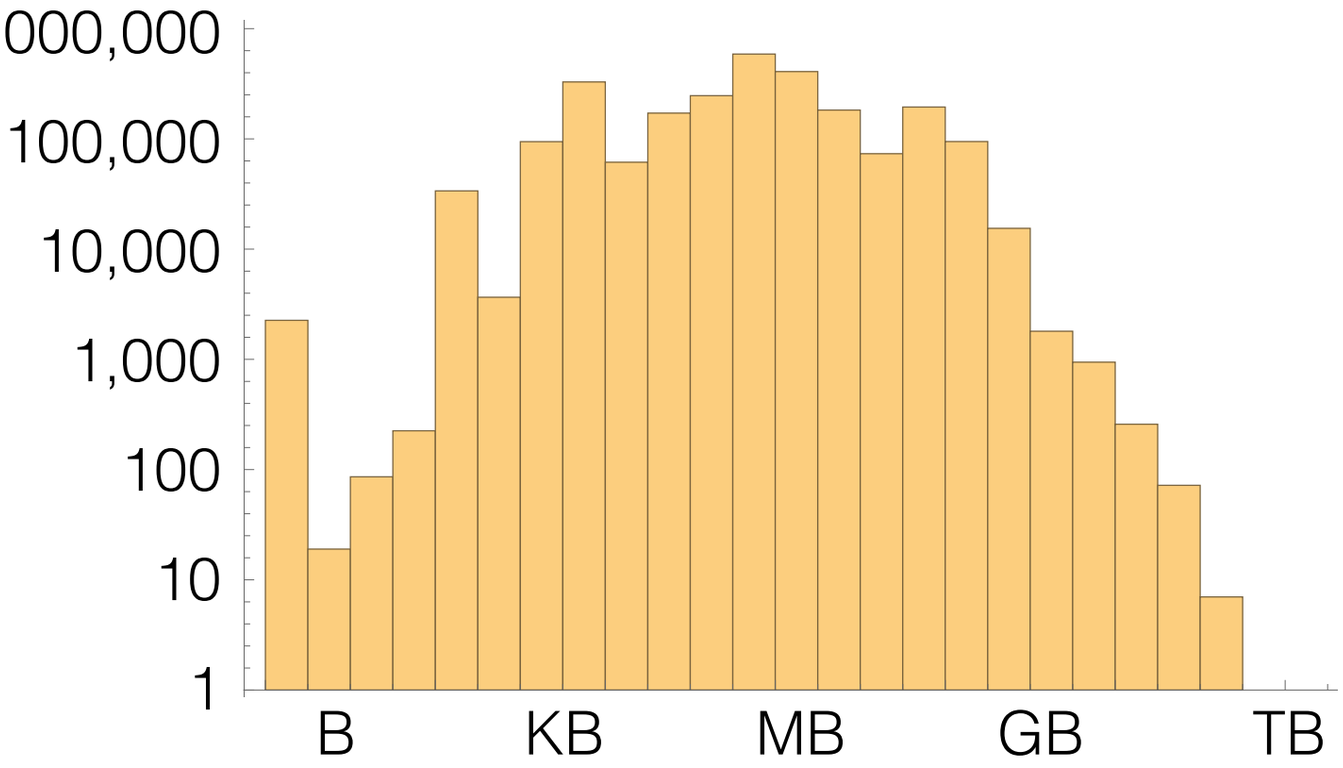
- Selected data can be pushed to these resources with the press of a button on the web interface



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

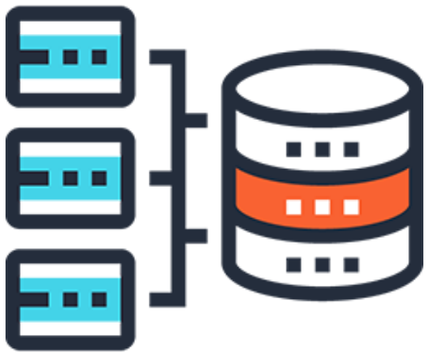


System architecture

- 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

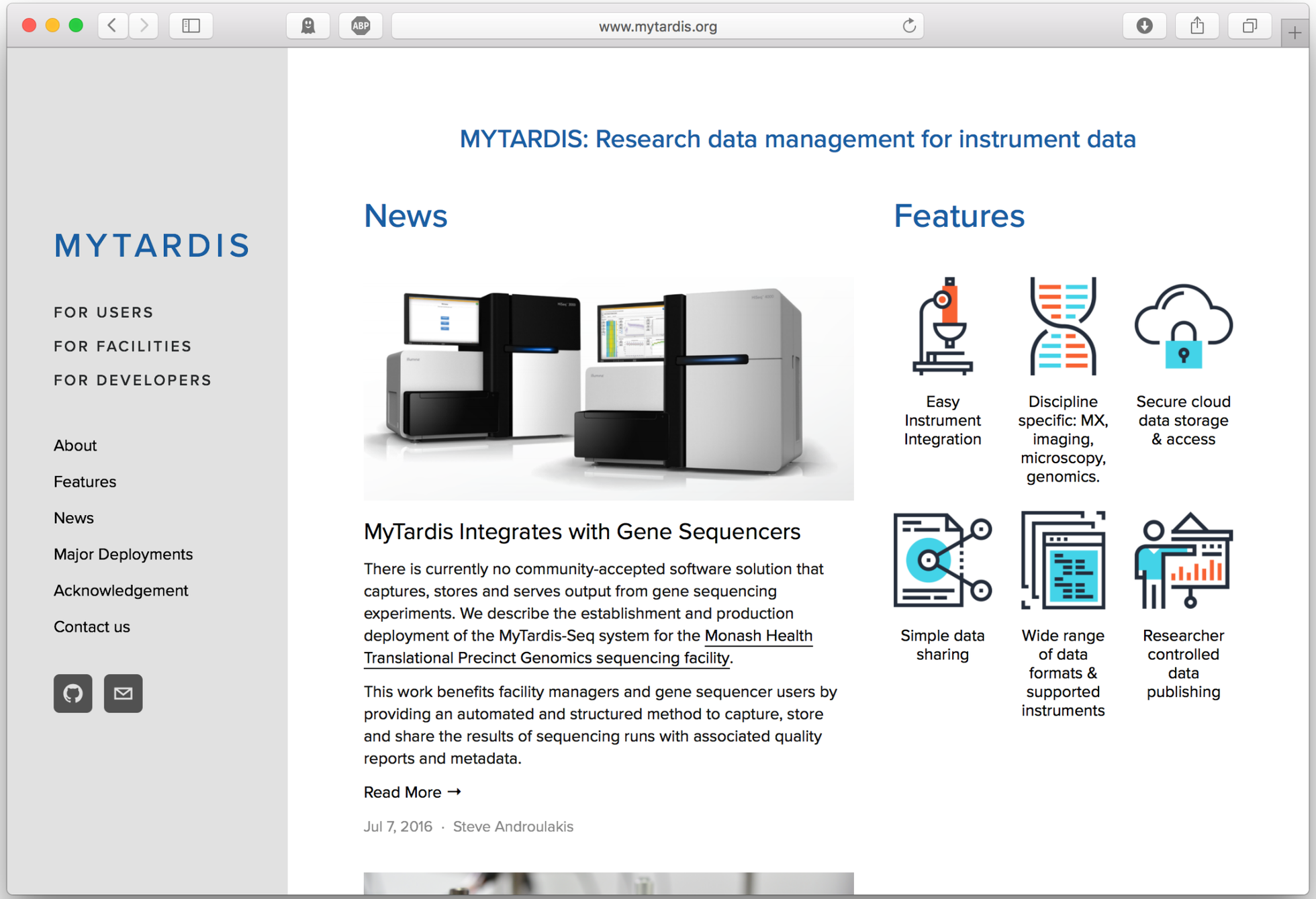
Compute & Storage

- 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



MONASH University · Australian Synchrotron · ands AUSTRALIAN NATIONAL DATA SERVICE · nectar · RDS www.rds.edu.au · National Imaging Facility · RMIT UNIVERSITY · ansto