

Meeting the next-generation instrument data challenge with MyTardis

Schulze, K.¹, Wettenhall, J.¹, Kumar, M.¹, Meyer, G.¹, Androulakis, S.¹, Rigby, J.¹, Perry, A.², Kannan, A.³, Mehnert, A.⁴, Janke, A.⁵, Goscinski, W.⁶ and Hassan, A.⁷

¹ Monash eResearch Centre, Monash University, Melbourne, VIC, Australia, ² Monash Bioinformatics Platform, Monash University, Melbourne, VIC, Australia, ³ Office of the Vice-Provost (Research and Research Infrastructure), Monash University, Melbourne, VIC, Australia, ⁴ Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, Perth, Australia, ⁵ National Imaging Facility, The University of Queensland, Brisbane, Australia, ⁶ Monash eResearch Centre, Monash University, Melbourne, Australia, ⁷ eResearch Delivery Leader, Monash University, Australia

New microscopy technologies are advancing characterisation capabilities faster than ever before. This creates exciting new challenges in the handling and storage of next-generation instrument data. Moreover, there is now strong consensus among the research community that we need to make research data Findable, Accessible, Interoperable and Reusable (FAIR). The FAIR data principles are recognised as key to deriving the most impact from research data, while also helping to ensure research integrity. One approach to meeting the challenges of handling large volumes of data while simultaneously making it FAIR is by capturing data and important metadata in a centralised data management system. Several data management solutions have emerged in recent years including MyTardis (<https://mytardis.org>), an open source research data management platform and ecosystem that is seeing adoption both in Australia and internationally.

MyTardis was originally developed to handle data from macromolecular crystallography¹, but through national Australian projects like the Characterisation Virtual Laboratory, ImageTrove and Australian National Data Service Trusted Data projects, its capabilities have been vastly expanded. Across its many deployments in Australia, MyTardis and the ecosystem of tools around it now capture and manage data from a broad selection of high-end research instruments, including optical and electron microscopes, pre-clinical and clinical medical imaging instruments and next-generation sequencing instruments.

MyTardis is a scalable web framework capable of securely ingesting, managing and storing 10s of Terabytes of data per week from multiple instruments and facilities. It also provides a unique set of features to enhance the ability of both researchers and core facilities/research platforms to manage data from different instruments. These features include:

- A well-defined data model - Experiment, Dataset, Datafile.
- The ability to capture essential metadata at all levels of the data hierarchy.
- A focus on providing streamlined capture of primary data at the source i.e., instrument integration.
- The ability to securely share data with collaborators.
- User-controlled data publishing.
- The ability to push data into cloud and high-performance compute environments for analysis and visualisation.
- Tools for facility managers to monitor data flow from facilities.

In this talk we will present our experience in developing MyTardis and operating a large deployment of MyTardis called Store.Monash, with particular emphasis on how it helps to securely store and manage imaging data from the core imaging facilities at Monash University. We will also outline the short- to medium-term roadmap for MyTardis development and our plan to engage the wider community to help us build the next-generation platform for instrument data management.

Acknowledgements

Contributors from the MyTardis open source community.

References

Meyer, G. R. *et al.* Operation of the Australian Store.Synchrotron for macromolecular crystallography. *Acta Cryst. D* **70**, 2510 - 2519 (2014).