Optical Microscopy Image Data Management BoF

Jason Swedlow [1], Andrew Lewis[2], Kim Linton[3]
[1] University of Dundee, Dundee, UK, j.swedlow@dundee.ac.uk
[2]Griffith University, Brisbane, Australia, a.lewis@griffith.edu.au
[3]Monash University, Melbourne, Australia, Kim.Linton@monash.edu

DESCRIPTION

Microscopy images of animal and/or human tissues and cells are taken in many research projects at most Universities. Cellular and systems biology, for example, have become increasingly data-intensive fields of research. Current generation imaging instruments are now automated, high-throughput devices capable of generating Terabytes of data daily. With multiple microscope types and vendors producing many different proprietary microscopy image data formats, a major issue lies in how to extract, store and query the machine-generated metadata; also how to annotate and associate experimental metadata with the raw data whilst ensuring the integrity of the association and provenance of the data.

On the 1st July 2012, the NHMRC Policy on Dissemination of Research Findings will change, making it a requirement that any publications arising from an NHMRC supported research project must be deposited into an open access institutional repository within twelve months of publication. This is in keeping with the fundamental belief that data sharing is essential for expedited translation of research results into knowledge, products, and procedures to improve human health.

At the institutional level a data management system (or similar software platform) is more desirable than manual processes conventionally used, to effectively store and manage metadata relating to microscopy image and associated experimental data. The system must also provide interfaces for the storage and management of derived or additionally annotated metadata (e.g. relating to derived image analysis) and automate where possible the generation, capture and/or association of metadata relating to microscopy image collections and experiments.

OUTLINE

This Bird of a Feather session will provide an overview of OMERO and two approaches for enhancing this platform to enable data collections to be shared through ANDS Australian data commons, Research Data Australia.

1. Overview of OMERO - Data-intensive research depends on tools that manage multi-dimensional, heterogeneous datasets. Open Microscopy Environment Remote Objects (OMERO) is an open-source software platform that enables access to and use of a wide range of biological data. OMERO uses a server-based middleware application to provide a unified interface for images, matrices and tables. OMERO's design and flexibility have enabled its use for light-microscopy, high-content-screening, electron-microscopy and even non-image-genotype data. This presentation will provide an overview of OMERO.

    This presentation will be delivered by Jason Swedlow. (*10 minutes*)

2. **Approach #1 -** The National Centre for Adult Stem Cell Research (NCASCR), at Grifith University undertook a project to address the key issues of image importing and archiving, metadata collation, conversion, standardisation, sharing, publication-readiness and secure enterprise-wide access and querying. The resulting system

enables sharing and collaboration by enabling image creators to share their datasets with other labs and researchers. Of greatest immediate impact however is the ability to allow users to search for historical images according to metadata attributes.

The system was built using a freely available Open Source microscopy software suite [1] (including: OMERO, OME Web, BioFormats and related components). Processes were developed for metadata capture, annotation and management. Image collections (subject to publication status, copyright, ethics and all legal requirements) are fed to the Australian Research Data Commons (ARDC) using Griffith University's Research Hub. The system also houses experimental and non-published image collections for microscopy users to manage their day-to-day investigations and provides a mechanism (publication readiness tagging) to help facilitate the automated / semi-automated workflows which lead to image collections being published to the ARDC. Additional functionalities and interfaces to meet specific project requirements were developed, and project source code is publically available through the SourceForge project [2].

This presentation will be jointly delivered by the two developers, Heidi Perrett and Amanda Miotto. (*10 minutes*)

3. **Approach #2** – Optical Microscopy Researchers, associated with Monash Univerisity's Micro Imaging 'MMI' facility required a solution to import, securely store, search and analyse images. They also needed to be able to collaborate with other researchers within their communities.

After evaluating various options, the project team consisting of researchers, MMI and e-Research personel selected OMERO. This product offers both a web and desktop client. To meet the ANDS requirements of being able to register data collections with Research Data Australia, the desktop client was extended, as it offered more functionality and operated within the researchers' existing workflow.

A workflow utilising tags, interfaces to Creative Commons and Monash University's Research Master, and an automated interface directly to Research Data Australia (ANDS) from the OMERO client was developed and deployed. The solution is now being utilised by many researchers to store, retrieve, reuse and share their data.

ANDS funded this activity as one of its Data Capture projects. The University of Dundee also supported Monash University in the development effort by providing guidance and helping them to extend their existing code base. OMERO project source code can be found in Github[3]. Metadata Registration server source can be found in Google Code[4].

This presentation will be delivered by Kim Linton. (*10 minutes*)

4. **Discussion -** Engage the audience to participate in the discussions. (**30 minutes)**

REFERENCES

1. *The Open Microscopy Environment*. Available from:
   http://www.openmicroscopy.org/site, accessed 9 May 2012.
2. *The GIIAF Microscopy Library.* Available from: http://sourceforge.net/projects/giiaf-micro-lib/, accessed 9 May 2012.
3. Github. Available from: https://github.com/emilda/openmicroscopy. Accessed 29 May 2012
4. Google Code. Available from: http://code.google.com/p/monashmie/. Accessed 29 May 2012

ABOUT THE SPEAKERS

Jason Swedlow earned a BA in Chemistry from Brandeis University in 1982 and PhD in Biophysics from UCSF in 1994. After a postdoctoral fellowship with Dr T. J. Mitchison at UCSF and then Harvard Medical School, Dr Swedlow established his own laboratory in 1998 at the Wellcome Trust Biocentre, University of Dundee, as a Wellcome Trust Career Development Fellow. He was awarded a Wellcome Trust Senior Research Fellowship in 2002 and named Professor of Quantitative Cell Biology in 2007. His lab focuses on studies of mitotic chromosome structure and dynamics and has published numerous leading papers in the field. He is co-founder of the Open Microscopy Environment (OME), a community-led open source software project that develops specifications and tools for biological imaging. In 2005, he founded Glencoe Software, Inc., a commercial start-up that provides commercial licenses and customization for OME software. In 2011, Prof Swedlow and the OME Consortium were named BBSRC's Social Innovator of the Year and Overall Innovator of the Year. In 2012, He was named Fellow of the Royal Society of Edinburgh. Prof Swedlow is Co-Director of the Analytical and Quantitative Microscopy Course, an annual course that covers the latest developments in advanced quantitative light microscopy at Marine Biological Laboratory, Woods Hole, USA.

Dr Andrew Lewis is Senior Research Specialist in Griffith University's eResearch Services group. He has over 25 years experience in computational research, including a decade in industrial applied research. He has been providing support to Griffith University researchers in High Performance Computing and eResearch for over 15 years and has a joint appointment in the School of Information and Communication Technology pursuing research and teaching in HPC, parallel and distributed computing, scientific data visualisation, and computational optimisation and its applications.

Ms Heidi Perrett is an eResearch Web and Software Developer in Griffith University's eResearch Services group. She graduated with a double degree in Science and IT majoring in Biological Sciences and Information Systems. She has been a lead developer on a number of systems utilising open source software platforms for eResearch support of researchers in all disciplines. She has developed discipline-specific portals and tools using Drupal, Wordpress, CiviCRM, OMERO and a variety of scripting languages.

Ms Amanda Miotto is an eResearch Support Specialist in Griffith University's eResearch Services group. She graduated with a Bachelor of Science majoring in Bioinformatics before gaining hands on experience developing software systems in support of a range of projects in genomics, systems biology and microbiology. She writes in php, python, sql, and R primarily and has worked on projects using tools such as UCSC Genome Browser mirror, AISRAP websites, Limesurvey, Cytoscape, Galaxy and Chipster before getting involved with OMERO for Griffith's Adult Stem Cell Data Capture Project.

Ms Kim Linton is the Senior Research Systems Facilitator at the Monash e-Research Centre. She is responsible for gathering requirements for proposed research systems undertaken by the Monash e-Research centre and then turning the requirements into a conceptual model for further development. She joined Monash University in October 2011. Prior to this appointment Kim was a Senior Business Analyst in the Telstra Enterprise Data Warehouse for a number of years where she integrated and reported on enterprise wide data.