Building an Image Management Core at Harvard Medical School using OMERO

Douglas Russell, Harvard Medical School

Background with OMERO

- 1999: Early OME Development in Sorger Lab at MIT
- 2010: HMS NIH LINCS
  - OMERO 4 Development with GS: HCS and Public Data Sharing
- Harvard Program In Therapeutic Science (HiTS) and Laboratory For Systems Pharmacology (LSP)
  - Perkin Elmer Columbus

HMS Facts and Figures

HMS Pre-Clinical Research Departments Producing Microscope Images

- Biological Chemistry and Molecular Pharmacology
- Cell Biology
- Genetics
- Global Health and Social Medicine
- Microbiology and Immunobiology
- Neurobiology
- Systems Biology

- Tenured and tenure-track faculty on the HMS campus in 10 preclinical departments 151
- Full-time faculty on campus and at affiliated hospitals 9,443
- Total PhD students 799
- Trainees (residents and postdoctoral fellows) 9,071

Objectives

- Store and organize microscopy data and associated metadata
- Reduce redundancy
- Data sharing for collaboration (e.g. IDAC or even PI oversight)
- Exploration and visualization
- Common interfaces to heterogeneous data
- Figure production
- Analysis
- Publication (including NIH conformance)
- Archiving (including NIH conformance)

Getting Started

- Support from PIs and core facilities
- Partnership with Research Computing (RC)
- Support contract with Glencoe Software
- Partnership with Research IT Solutions (RITS)
  - Mandated to deliver unconventional services
- Funding for the Image Management Core from the Tools-N-Technology committee
- Funding from HiTS through LINCS
- Funding from HMS IT
- Douglas comes to HMS

HARVARD UNIVERSITY

HARVARD.EDU

HARVARD MEDICAL SCHOOL

Search

Image Management Core

A service to help HMS researchers store, manage, and share images

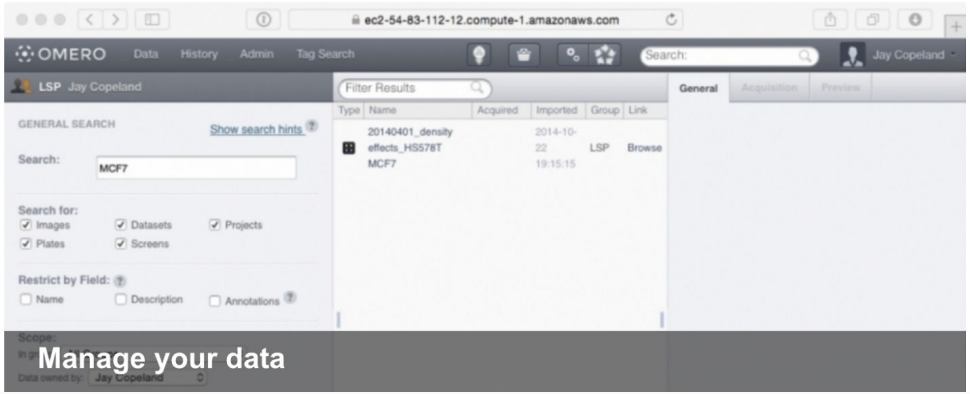
HOME | About | Software | Support | FAQ

Image Management Core

The Harvard Medical School Image Management Core is a new service help researchers manage image data and metadata. The IMC was created in response to the growing complexity and difficulty of managing research image data. Read more.

The Challenge of Managing Research Images

We can help you...



IMC Web page and introduction

Enterprise Deployment

Trials and Tribulations

- Provisioned twin sets of identical hardware for dev/production environment and to mitigate hardware failure
- Installed and managed by HMS Research Computing
  - Configured with configuration management tool Ansible
- Supported by Glencoe Software
- Project managed by Research IT Solutions
  - Planning, user requirements, process, timelines, accountability, documentation

- LDAP to Active Directory (in future will have full group information)
- Storage (very large scalable storage on EMC Isilon)
- OMERO Client software on compute cluster and distributed to users

Configuration Management with Ansible

Configuration management (CM) is a systems engineering process for establishing and maintaining consistency of a product's performance, functional, and physical attributes with its requirements, design, and operational information throughout its life… *Wikipedia*

```
lineinfile: dest=/etc/example_config_file line="foo: bar"
```

Idempotence is a funky word that often hooks people!

Idempotence is the property of certain operations in mathematics and computer science, that can be applied multiple times without changing the result beyond the initial application.

Configuration Management with Ansible

- Research Computing currently manage infrastructure with Ansible
- Ansible role from Glencoe Software to install, update and configure OMERO
- Allows us to rebuild from scratch and upgrade quickly
- We can test configuration code on development for fixes, enhancements or upgrades, before deployment to production

OMERO and Ansible

- Not ideal to have to specify defaults (which might change)

```
omero config set omero.web.ui.top_links '[["Data", "webindex", {"title": "Browse Data via Projects, Tags etc"}],["History", "history", {"title": "History"}],["Help", "http://help.openmicroscopy.org/", {"title": "Open OMERO user guide in a new tab", "target": "new"}],["Figure", "figure_index"]]'
```

- `omero config append` is not idempotent

```
omero config append omero.web.ui.top_links '["Figure", "figure_index"]'
omero config append omero.web.ui.top_links '["Figure", "figure_index"]'
# omero.web.ui.top_links=[["Data", "webindex", {"title": "Browse Data via Projects, Tags etc"}],["History", "history", {"title": "History"}],["Help", "http://help.openmicroscopy.org/", {"title": "Open OMERO user guide in a new tab", "target": "new"}],["Figure", "figure_index"], ["Figure", "figure_index"]]
```

OMERO and Ansible

- Idempotent `omero config` commands would be great

```
omero config present omero.web.ui.top_links '["Figure", "figure_index"]'
omero config present omero.web.ui.top_links '["Figure", "figure_index"]'
# omero.web.ui.top_links=[["Data", "webindex", {"title": "Browse Data via Projects, Tags etc"}],["History",
"history", {"title": "History"}],["Help", "http://help.openmicroscopy.org/", {"title": "Open OMERO user guide
in a new tab", "target": "new"}],["Figure", "figure_index"]]
```

- OMERO Ansible Module might be ideal

```
# Now
command: "{{ omero_base }}'/bin/omero config set {{ item.key }} {{ item.value }}"

# With an OMERO Ansible module
omero_config: key={{ item.key }} value={{ item.value }} state=present
```

User Support

Research IT Solutions

- RITS/HMS IT Customer Service Representatives trained to triage basic queries and escalate to technical staff where appropriate
  - Phone support
  - Email/ticket-based
  - Expert consultant on figure creation trained on OMERO.figure
  - Live support with Slaask
- Introductory classes

Static/Live Demo of Slaask

User Requirements

Mandatory Scientific Metadata

- Mandatory scientific metadata to be attached to imported datasets
- Varies by group and/or dataset type
- Tracking of changes
- User friendly adding/editing
- Reporting of non-conformance

OMERO.forms

- New OMERO.web plugin
- JSON powered forms
- Form submitted data stored as MapAnnotations attached to a “form master” super-user
  - Historical as well as current
  - Record of when and by whom a form was submitted
- Supports moving of underlying data between groups
- Tamperproof
- Also attaches a MapAnnotation to the dataset with the key-values extracted from the form submission, but owned by the submitting user for convenience for analysis or similar

Static/Live Demo of OMERO.forms

1984

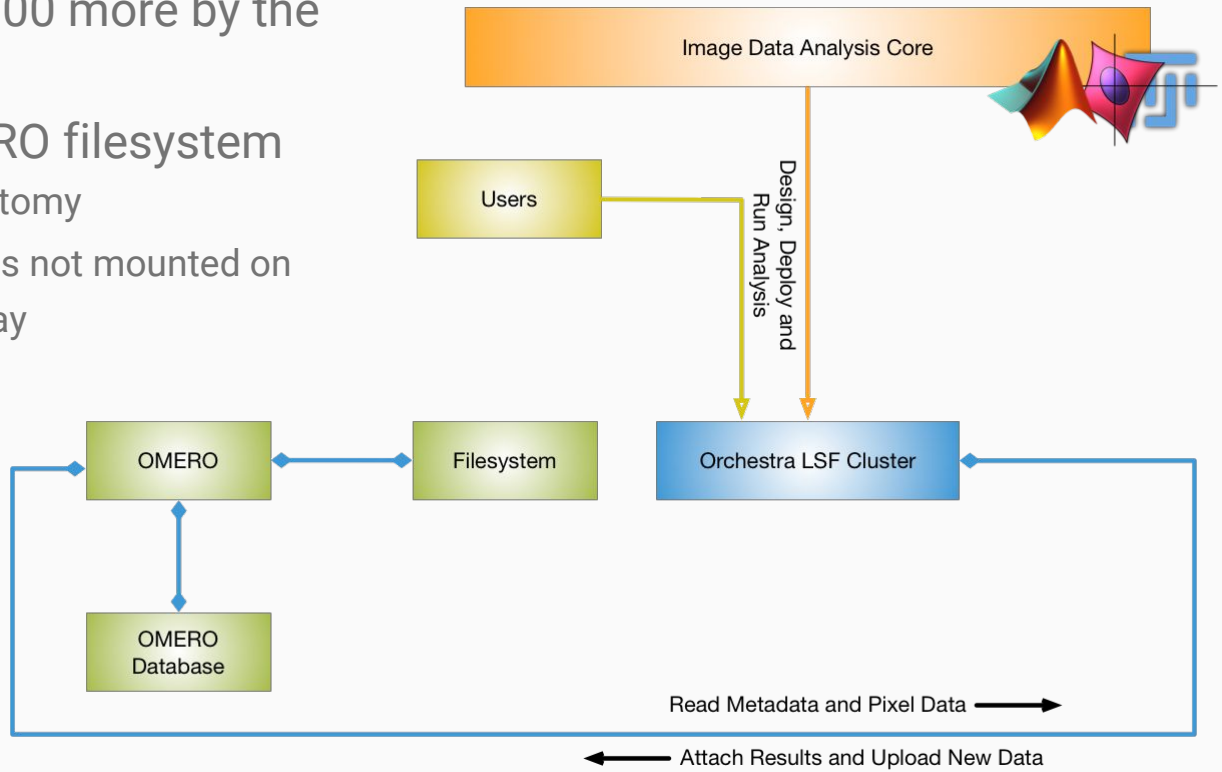
- Scheduled generation of reports
- Group specific requirements
  - A user in group A has no committment to fill any forms
  - A user in group B must have form X for all datasets
  - A user in group C must have one of the group C designated forms for all datasets
- Group specific notifications
  - If a user from group B does not fulfill requirements they get notified
  - If a user from group C has more than 3 datasets that do not fulfill requirements, the owner of group C gets notified

Analysis at Scale

Orchestra Compute Cluster, OMERO and performance

Orchestra LSF Compute Cluster

- ~8000 cores (+10000 more by the end of the year)
- No access to OMERO filesystem
  - Permissions dichotomy
  - Storage filesystems not mounted on many nodes anyway



OMERO and Orchestra

- Use OMERO directly
  - Dummy analysis job to test performance. Just reads planes.

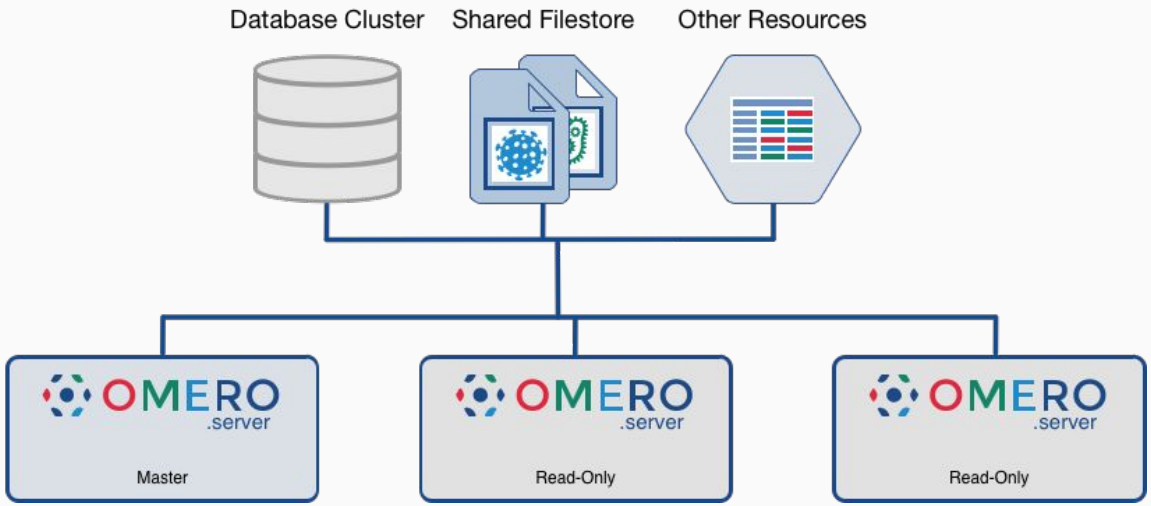
#Jobs	Mean time per job	Median time per job	Throughput
1	453.12	453.12	~100Mb/s
10	742.48	663.66	~900Mb/s
100*	1540.18	1539.37	~3Gb/s

*At X100, small, but noticeable impact on the client interface

Horizontal Scaling of OMERO

Proposal in conjunction with Glencoe Software

- Read-only OMERO instances for query
  - Read-only master or live replicated database access
  - Read-only Managed Repository/Files access



NIH LINCS BD2K Supplement

- Grant to fund this and some cloud based OMERO archiving work 99.999% (?) Approved!
- Hopefully a trend of NIH support of OMERO emerging

Conclusions

- Deploying services across a whole institution is very hard and we have only just begun!
- We were able to add features quickly that were required by some of our users within OMERO
- Analysis at scale will require some work

But…

- Facilities now available!
- Even in our Beta stage, users finding our site unbidden and getting started

Acknowledgements

- HMS Research Computing
- HMS IT
- HMS Research IT Solutions
- HMS Tools-N-Technology






Jason McDonald