

A service architecture for bio-image data management at France BioImaging

G. Gay^{1,2}, J.F. Guillaume³, G. Maucort⁴, M. Mongy³, E. Faure², P. Paul-Gilloteaux³

¹ BioCampus, UAR 3426 CNRS - US 09 INSERM - UM, Montpellier, France; ² LIRMM, UMR 5506 CNRS, INRIA, UM, Montpellier, France; ³ Nantes Université, CHU Nantes, CNRS, Inserm, BioCore, US16 UAR 3556, SFR Bonamy, Nantes, France; ⁴ University of Bordeaux, CNRS, INSERM, Bordeaux Imaging Center, UMS 3420, US 4, Bordeaux, France.

Abstract

Microscopy data has dramatically increased both in volume and complexity in the last two decades. At the same time, institutions, funding bodies and publishers insist on the adoption of Open Science practices. Thankfully, the need to properly manage huge and complex data and the technical requirements for open science can be met synchronously and benefit each other, as big data volumes and sophisticated analysis pipelines provide researchers with an incentive to adopt new data management workflows, and FAIR (Findable, Accessible, Interoperable and Reusable) principles answer the technical requirements of such a workflow.

Following the French Research Ministry Open Science Roadmap, France BioImaging has initiated in 2021 its "data" initiative to deploy a new data management infrastructure in each of its regional nodes, switching from hosting microscopy data in the lab to mutual storage in University managed data-centers. The mission objectives are to optimize infrastructure and maintenance cost, ease data sharing and access, and work towards FAIR data practices.

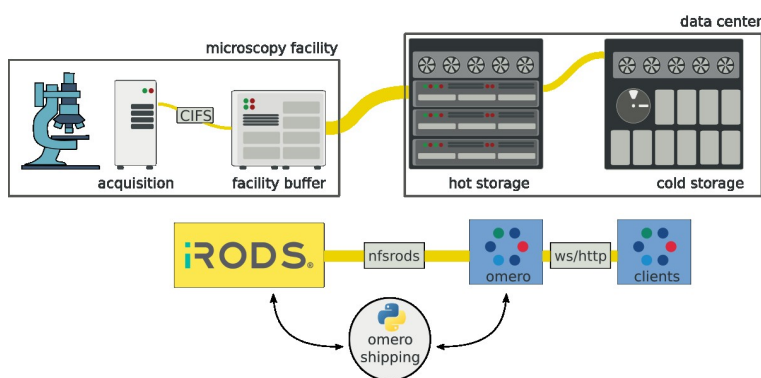
We choose OMERO (openmicroscopy.org) as the core platform for data annotation and access, with iRODS (irods.org) as the data transport and management layer.

In this contribution, we first give technical details on the chosen architecture and workflows: user management, data import, export and analysis. We aim to provide an Infrastructure as a Service architecture that is easy to maintain and adaptable to each context by relying on "infrastructure as code" and continuous integration methods.

We then discuss our efforts to ease users onboarding and training. In particular, we present our recommendations for data organization, following the Investigation Study Assay framework, and our efforts to ease metadata management from the users.

We conclude with the future of the service and its integration to other Open Science tools in the community.

This work is supported by The National Infrastructure France-BioImaging supported by the French National Research Agency (ANR-10-INBS-04)



Data acquired on a microscope is uploaded to a storage unit local to the facility through a SAMBA / CIFS mount. iRODS ingests this data. The omero-shipping package is used to transport the data in an iRODS collection and into OMERO. In OMERO, data is imported in place through a NFS mount of the iRODS collection, so that no copy occurs. Data management (e.g. transfer to cold storage) is managed by iRODS transparently to OMERO and the users. Clients can access the data through the OMERO API or directly through iRODS