

RDF AS A BRIDGE TO DOMAIN-PLATFORMS LIKE OMERO

or, There and back again

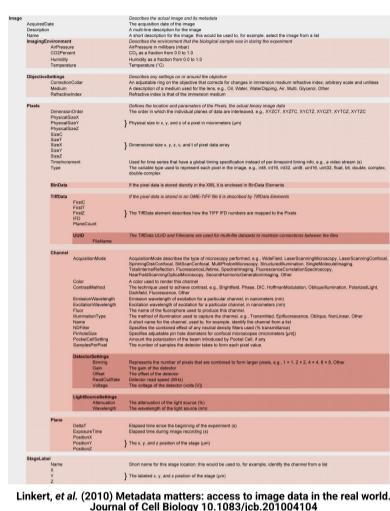
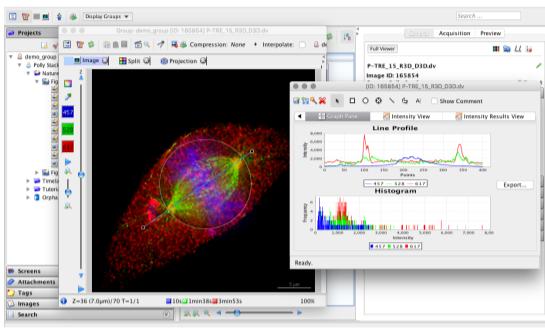
In 2005, the first version of OMERO stored RDF natively. However, just a year after the 1.0 release of RDF, performance considerations led to the development of a more traditional SQL approach for OMERO. A binary protocol makes it possible to query and retrieve metadata but the resulting information cannot immediately be combined with other sources. This is the adventure of rediscovering the benefit of RDF triples as a -- if not the -- common exchange mechanism.

Josh Moore, Andra Waagmeester, Kristina Hettne, Katherine Wolstencroft, Susanne Kunis
 0000-0003-4028-811X 0000-0001-9773-4008 0000-0002-4182-7560 0000-0002-1279-5133 0000-0001-6523-7496
 German Biomaging e.V. Micelio Centre for Digital Scholarship Leiden University Libraries The Netherlands Advanced Computer Science Center for Cellular Nanoscopy University Osnabrück, Germany



The production of digital images in biomedical research is a field in rapid evolution, with 500 newly indexed articles on PubMed each day in the life-sciences alone. The ever-increasing volume of bioimages and the granularity of metadata and annotations complicate data management.

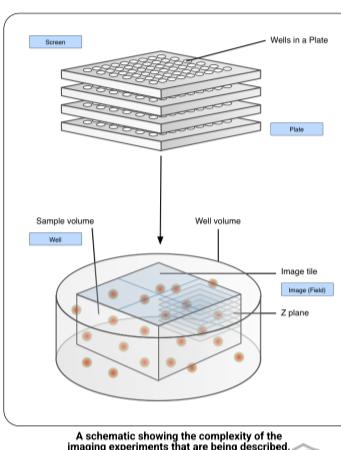
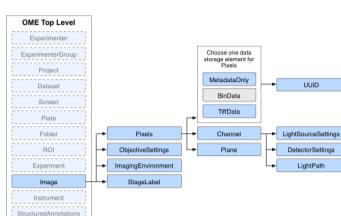
OMERO is an open-source platform built to manage information from the over 160 bioimaging file formats produced by microscopes using the Open Microscopy Environment (OME) data model.



Linker, et al. (2010) Metadata matters: access to image data in the real world. Journal of Cell Biology 10.1083/jcb.201004104

The **OME Model**, defined in XML Schema, consists of over 160 classes representing a common vocabulary for describing multi-dimensional imaging experiments. In OMERO, these values are stored in PostgreSQL tables with numerous foreign key constraints and made available through a bespoke API.

Extracting all information about an imaging experiment can be time-consuming and error-prone. By providing a mapping via the **omero-rdf** library, integration is made possible with a host of other platforms.



A schematic showing the complexity of the Imaging experiments that are being described.

\$ omero rdf Image:165854

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix efo: <http://www.ebi.ac.uk/efo/> .
@prefix fhir: <http://hl7.org/fhir/> .
@prefix jerm: <http://jermontology.org/ontology/JERMOntology#> .
@prefix obi: <http://purl.obolibrary.org/obo/OBI> .
@prefix ome: <http://www.openmicroscopy.org/Schemas/OME/2016-06#> .
@prefix rocrate: <https://w3id.org/ro/crate/1.1/context> .
@prefix sio: <http://semanticscience.org/resource/> .
@prefix sioc: <http://rdfs.org/sioc/ns#> .
```

#1 Take-home: By translating domain-specific OMERO metadata to RDF, it is more readily combinable with other domains.

The journey out of the domain-specific realms begins!

Service integrations

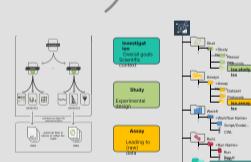


Description: RO-Crate is a lightweight approach to packaging research objects with structured metadata based on schema.org annotations in JSON-LD.

Bioimaging: Exporting OMERO datasets to RO-Crate provides a standardized interface around a collection of domain-specific files. Bioimaging brings complex data formats which, like a puzzle, describe the spatial relationship between independent files representing regions of an image.

Challenge: How best to bridge to these domain-specific protocols is still unclear.

#2 Take-home: Rather than being "hidden" in a domain-specific OME-XML file, extracted triples can be made discoverable in RO-Crate's JSON-LD.



Description: The Annotated Research Context (ARC) by DataPLANT is a self-contained data format originally designed for plant research, including assay or measurement data, workflow and computation results accompanied by metadata. It combines the Investigations, Studies, Assays (ISA) model, Common Workflow Language (CWL), and RO-Crate standards.

Bioimaging: OMERO-based data can be transformed into or from this format.

Challenge: The need to return to a file-based view on the data when users are accustomed to an image-centric view.

#3 Take-home: An ARC extends the RO-Crate model providing versioned metadata and ISA support which aligns closely with OMERO.



Description: Provides a flexible yet expressive standard for describing many kinds of data models from value sets and flat, checklist-style standards to complex normalized data structures that use polymorphism and inheritance (<https://f1000research.com/posters/12-1101>).

Bioimaging: Many groups in the bioimaging community are now moving to describing their models with LinkML with the hope of creating an extensible network of schemas

#4 Take-home: LinkML provides a modeling language that can capture the richness of OMERO while providing URIs for each Class and Predicate.

A stable protocol allows mapping the many chunks of a multidimensional image into a consistent representation.



Description: "REDCap (Research Electronic Data Capture) is a browser-based, metadata-driven EDC software and workflow methodology for designing clinical and translational research databases." (<https://en.wikipedia.org/wiki/REDCap>)

Bioimaging: Thousands of groups use REDCap to capture detailed information about critical human diseases, such as tuberculosis in our current work. Images of the diseased tissue, however, along with annotations from pathologists are stored in OMERO, making it difficult to get a complete view of all research materials.

Challenge: To construct a single knowledge graph, it is necessary to implement an automated process for mapping REDCap tables to RDF using the referenced URIs provided by OMERO.

#5 Take-home: By embedding an R script in REDCap, the relational tables making up the clinical records can be exported to RDF

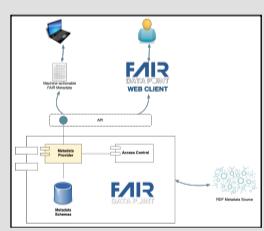


Description: The FAIRDOMHub is an ELIXIR-approved public resource for sharing heterogeneous research assets - datasets, models, processes, workflows and research outcomes. It is based on the FAIRDOM-SEEK platform, where research assets are available in RDF. They can be aggregated and shared as RO-Crate Research Objects and published using DOIs.

Bioimaging: The LiSym Cancer Project has integrated FAIRDOM-SEEK with OMERO by extracting metadata into SEEK and providing external links to the OMERO server.

Challenge: A more comprehensive solution would make the images directly accessible as FAIR objects.

#6 Take-home: Exported triples can be exposed in FAIR platforms with no need for further transformation or encoding.



<https://spec.fairdatapoint.org/fdp-specs-v1.2.html>

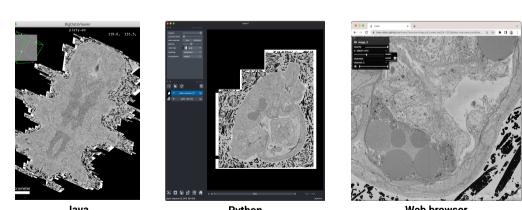
#7 Take-home: FDP offers a registry for OMERO datasets into which exported triples can be loaded.

Join us at the hackathon to help make it happen!
<https://swat4hcls2024-hack-1-publishing-fair-datasets-from-bioimaging-repositories>

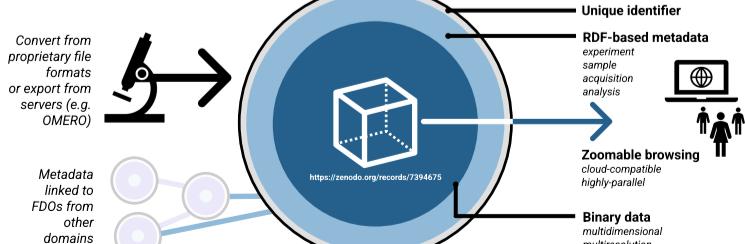


FAIR Image Objects

Description: FDO-compatible datatype for bioimaging to ensure data is open and web-accessible. Terabytes of pixels as well as analytical results can be made shareable, linkable, browsable, re-usable, archivable. A pyramidal structure allows Google Maps-style zooming. A cloud-optimized ("chunked") format – **OME-Zarr** – allows referencing individual regions of an image in parallel.



Exported objects can be further enriched with annotations which were not originally present in OMERO.



#8 Take-home: To establish links to other domains, it's essential to leverage more than just custom APIs.

8-TB electron microscopy volume of a 6 day old Platynereis larva from Vergara et al. 2020 (doi: 10.1101/2020.02.26.961037)
 Available at: <https://s3.embl.de/2k-2020/platy-raw.ome.zarr>



Have you been on a similar journey in your domain? We'd love to hear.

This poster can be found under

DOI: 10.5281/zenodo.1068769

