There and back again: RDF as a bridge to domain-platforms like OMERO

Josh Moore[1,†], Andra Waagmeester[2,†], Kristina Hettne[3], Katherine Wolstencroft[4] and Susanne Kunis[5]

[1]*German BioImaging e.V, Society for Microscopy and Image Analysis, Konstanz, Germany*

[2]*Micelio, Belgium*

[3]*Centre for Digital Scholarship, Leiden University Libraries, The Netherlands*

[4]*Leiden Institute of Advanced Computer Science, The Netherlands*

[5]*Department of Biology/Chemistry and Center for Cellular Nanoanalytics, University Osnabrück, Germany*

Keywords

Open Microscopy Environment (OME), Bioimaging, Resource Description Framework (RDF)

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[1]. The ever-increasing volume of bioimages and the granularity of metadata and annotations complicate data management. OMERO[2] is an open-source platform built to manage information from the over 160 bioimaging file formats[3] produced by microscopes using the Open Microscopy Environment (OME) data model[4]. Beginning in 2005, the history of OMERO reflects the evolution of bioimaging, adapting to the growing complexity of data. Interestingly, OMERO's development predates the formalization of FAIR[5] principles and even the widespread use of JSON. An early attempt was made to base OMERO on RDF, but just a year after the 1.0 release of RDF, performance considerations led to the development of a more traditional SQL approach. Growing demands for annotations, though, have led OMERO to incorporate a system of structured annotations, weaker than the expressivity of RDF but still based on the same concepts.

While users of OMERO are appreciative of the ability to manage, annotate, and critically view their images, they must also link these results into wider FAIR management platforms. **Generic RDM** tools like iRODS and RO-Crate[6] exactly allow researchers to link disparate products of their investigations, aiding accessibility and interoperability. Other platforms like Fair Data Points[7] (FDP) improve the findability of datasets. In each of these cases, a domain-agnostic representation of OME data is needed to make images FAIR. Even within the **life sciences**, more general models such as the Investigation-Study-Assay model[8] (ISA) require mapping out of the conceptual bioimaging space. This is of interest for national level research

projects such as DataPlant within Germany's National Research Data Infrastructure (NFDI) with their Annotated Research Context[9] (ARC) as well as the Dutch FAIR Cell Observatory, a high content screening project in the Netherlands Bioimaging network with their Minimum Information for High Content Screening Microscopy Experiments[10] (MIHCSME) which in turn is based on the REMBI minimum information model[11]. Work with **patient data** has also shown that an intermediary is needed to bridge OMERO data with the highly tabular world of patient records. Whether dedicated platforms such as RedCap[12] or custom spreadsheets, the domain-specific model should and really cannot be extended to encompass the many existing biomedical ontologies that exist today. Instead, they must be made to work together.

In these use cases, RDF emerges as the most robust mechanism for capturing the complexity and diversity of metadata in bioimaging. To this end, we are currently developing omero-rdf[13] to provide consistent bridging between the domain-specific and agnostic worlds. From there, we will explore higher-level frameworks such as LinkML, YARRML, and ShExML frameworks for data modeling and transformation.

References

[1] C. Schmied, H. K. Jambor, Effective image visualization for publications - a workflow using open access tools and concepts, F1000Res. 9 (2020) 1373.

[2] C. Allan, et al., OMERO: flexible, model-driven data management for experimental biology, Nat. Methods 9 (2012) 245–253.

[3] M. Linkert, et al., Metadata matters: access to image data in the real world, J. Cell Biol. 189 (2010) 777–782.

[4] J. R. Swedlow, et al., Informatics and quantitative analysis in biological imaging, Science 300 (2003) 100–102.

[5] M. D. Wilkinson, et al., The FAIR guiding principles for scientific data management and stewardship, Sci Data 3 (2016) 160018.

[6] S. Soiland-Reyes, et al., Packaging research artefacts with RO-Crate, Data Sci. 5 (2022) 97–138.

[7] da Silva Santos, et al., FAIR data point: A FAIR-oriented approach for metadata publication, Data Intell. 5 (2023) 163–183.

[8] S.-A. Sansone, et al., Toward interoperable bioscience data, Nat. Genet. 44 (2012) 121–126.

[9] D. community, nfdi4plants/arc-specification: 1.2, 2023. URL: https://doi.org/10.5281/zenodo.10091038. doi:10.5281/zenodo.10091038.

[10] R. Hosseini, et al., FAIR high content screening in bioimaging, Sci Data 10 (2023) 462.

[11] U. Sarkans, et al., Rembi: Recommended metadata for biological images—enabling reuse of microscopy data in biology, Nature Methods 18 (2021) 1418–1422. URL: http://dx.doi.org/10.1038/s41592-021-01166-8. doi:10.1038/s41592-021-01166-8.

[12] P. A. Harris, et al., The REDCap consortium: Building an international community of software platform partners, J. Biomed. Inform. 95 (2019) 103208.

[13] J. Moore, German-bioimaging/omero-rdf: v0.1.5, 2023. URL: https://doi.org/10.5281/zenodo.10359688. doi:10.5281/zenodo.10359688.