NFDI4BIOIMAGE: Perspective for a national bioimaging standard

Josh Moore[1,†], Susanne Kunis[2,†]

[1]German BioImaging e.V, Society for Microscopy and Image Analysis, Konstanz, Germany
[2]Department of Biology/Chemistry and Center for Cellular Nanoanalytics, University Osnabrück, Germany

Keywords
Bioimaging, FAIR Data Objects (FDO), FAIR Image Objects (FAIR-IO), LinkML, Resource Description Framework (RDF), Next-generation file formats (NGFF), Open Microscopy Environment (OME)

Internationally, countless petabytes of biological imaging data are collected each year. Only a fraction can be shared in public domain archives like EBI's BioImage Archive or the Image Data Resource (IDR). Much of the remaining data is in want of approachable FAIR infrastructure to be shared. Exacerbating this situation, imaging data tends to be particularly opaque, due to a combination of proprietary file formats, unspecified metadata, and monolithic files which must be downloaded *in toto*. Yet this vast untapped pool of knowledge could foster education, training models, and new scientific discoveries.

Numerous groups are actively working on similar challenges across the research landscape. Within Germany, the National Research Data Infrastructure (NFDI) was founded to link all national research outputs across all domains. Funding was recently approved for NFDI4BIOIMAGE [1], including a team of research software engineers and data stewards, to represent the needs of the bioimaging community starting March 2023. Its first objective is to "champion the standardization of the bioimage data type." One task area led by the authors seeks to make such data open and web-accessible by combining the "FAIR Data Object" (FDO) concept with previous and ongoing efforts within bioimaging. This effort has been tentatively titled "FAIR Image Objects (FAIR-IO)" in the application, as seen in Figure 1.

To become a FAIR data object, imaging data first needs an open and accessible metadata representation. Two previous efforts have independently translated the widely-used XSD-based OME metadata model into OWL [2] [3]. However, though RDF is an ideal fundament, it has achieved limited uptake due to the perceived accessibility by the user and developer communities. To increase usability and drive adoption, LinkML is now being investigated as a platform-layer on top of RDF. A higher-level representation will help to harmonize data descriptions and provide more powerful tools such as code generators. Both for NFDI and internationally, data's expressive power will be determined by the degree to which common abstractions can be agreed upon.

SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences

[†]These authors contributed equally.

josh@openmicroscopy.org (J. Moore); sukunis@uos.de (S. Kunis)
0000-0003-4028-811X (J. Moore); 0000-0001-6523-7496 (S. Kunis)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3415/paper-27.pdf

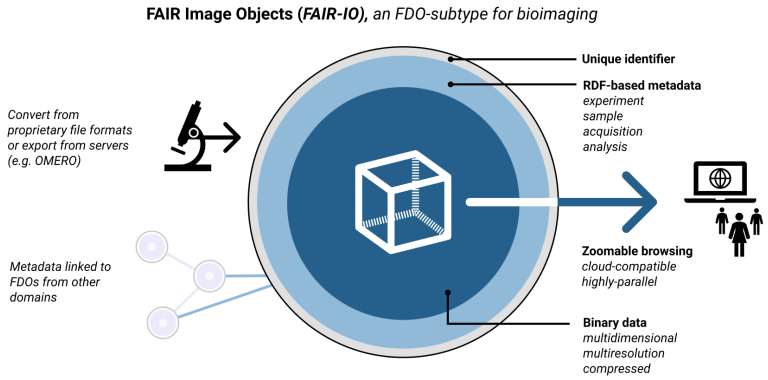

Figure 1: A FAIR Image Object combines the necessary acquisition and provenance metadata together with multi-resolution, chunked binary data in a single cloud-compatible format for simplified sharing and re-use. CC-BY: *NFDI4BIOIMAGE Consortium. (2021). Zenodo. https://doi.org/10.5281/zenodo.7394675*

However, even if described by powerful open metadata, access to the binary data remains a struggle. OME is working on a next-generation file format (NGFF) [4] to provide a scalable and web-compatible way to share bioimaging binary data. NGFF splits large, n-dimensional arrays into individidually referencable chunks stored as separate files. Packaging NGFF together with a graph of metadata creates a FAIR Image Object.

FAIR-IO enables sharing terabyte-scale data with a single URL with full context for downstream analysis into the linked data ecosystem. Within NFDI4BIOIMAGE, datasets will span multiple acquisition modalities, e.g. single-cell analyses, while linkages to other NFDI consortium will be established for larger, e.g. biomedical, goals. However, we are interested in bridging to other, international efforts and invite feedback from and collaboration with SWAT4HCLS participants throughout the soon-to-begin 5 year funding period.

References

[1] C. Schmidt, E. Ferrando-May, NFDI4BIOIMAGE - An Initiative for a National Research Data Infrastructure for Microscopy Data, 2021. URL: https://doi.org/10.11588/heidok.00029489, [Online; accessed 3. Dec. 2022].

[2] S. Samuel, B. König-Ries, Reproduce-me: Ontology-based data access for reproducibility of microscopy experiments, in: E. Blomqvist, K. Hose, H. Paulheim, A. Ławrynowicz, F. Ciravegna, O. Hartig (Eds.), The Semantic Web: ESWC 2017 Satellite Events, Springer International Publishing, Cham, 2017, pp. 17–20.

[3] J. Moore, N. Kobayashi, S. Kunis, S. Onami, J. R. Swedlow, On bringing bioimaging data into the open (world), in: Proceedings of the 12th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences (SWAT4HCLS), 2019, pp. 44–53. URL: http://ceur-ws.org/Vol-2849/#paper-06.

[4] J. Moore, et al., OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies, Nat Methods 18 (2021) 1496–1498. URL: https://www.ncbi.nlm.nih.gov/pubmed/34845388. doi:10.1038/s41592-021-01326-w.