Scalable strategies for a next-generation of FAIR bioimaging

Josh Moore[1,2]
[1]German BioImaging-Gesellschaft für Mikroskopie und Bildanalyse e.V., Constance, Germany,
[2]Open Microscopy Environment (OME) Consortium

For over 20 years, the Open Microscopy Environment (OME) has worked to provide infrastructure for the sharing of bioimaging data. Tools, APIs, and specifications like Bio-Formats[1], OMERO[2], and OME-TIFF[3] have allowed the community to grapple with much of the growing size, heterogeneity, acquisition rates, and general complexity of image data published in some 500 articles per day in the life-sciences alone[4]. The rising significance of research data management following, e.g., the Nelson memo[5] in the United States, and changes in storage paradigms like cloud computing, however, have created new demands within the bioimaging community.

In this presentation, the author as a representative of national and international initiatives will present strategies for scaling our cyberinfrastructure to meet the challenges of a FAIR future. This includes chunked, cloud-native formats[6] for sharing terabyte- and petabyte-size datasets and mechanisms for working with microscopy vendors to accurately capture all metadata related to a dataset in community-defined representations[7]. Vital, though, will be the ability of all stakeholders to evolve consensually the infrastructure itself to scale to meet future needs.

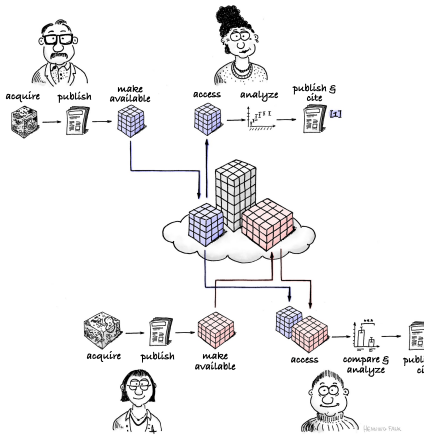

Fig. 1 FAIR re-use

FAIR sharing of data is beneficial for both data producers and consumers. Consumers gain access to interesting datasets that would otherwise be out of reach. Producers get citations to their work, when consumers publish their derivative work. OME-Zarr is a technological basis for enabling effective FAIR sharing of large image datasets.

References:

1. Melissa Linkert, et al. Metadata matters: access to image data in the real world. J Cell Biol (2010) https://doi.org/10.1083/jcb.201004104
2. Allan, et al. OMERO: flexible, model-driven data management for experimental biology. Nat Methods (2012) https://doi.org/10.1038/nmeth.1896
3. Besson, S. et al. Bringing Open Data to Whole Slide Imaging. Digital Pathology. ECDP (2019) https://doi.org/10.1007/978-3-030-23937-4_1
4. Schmied C and Jambor HK. Effective image visualization for publications – a workflow using open access tools and concepts. F1000Research (2021) https://doi.org/10.12688/f1000research.27140.2
5. Nelson, A. Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf Published Aug 25, 2022
6. Moore, J., et al. OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies. Nat Methods (2021) https://doi.org/10.1038/s41592-021-01326-w
7. Sarkans U., et al. REMBI: Recommended Metadata for Biological Images-enabling reuse of microscopy data in biology. Nat Methods (2021) https://doi.org/10.1038/s41592-021-01166-8

Keywords: FAIR; community; bioimaging, data sharing; data; cloud; format

Supported by: J.M. is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 501864659 as part of NFDI4BIOIMAGE, by EU Horizon grant 101130216 as part of FoundingGIDE, and by Chan Zuckerberg Initiative DAF for work on OME-NGFF by grant numbers 2019-207272 and 2022-310144 and on Zarr by grant numbers 2019-207338 and 2021-237467.