Bringing FAIR Bioimage Data Management into Practice: the Information Infrastructure for BioImage Data (I3D:bio) Project – bottom-up Community Support for Microscopy Data Sharing and Preservation.

Christian Schmidt[1], Michele Bortolomeazzi[1], Tom Boissonnet[2], Julia Dohle[3], Tobias Wernet[4], Janina Hanne[5], Roland Nitschke[4], Susanne Kunis[3], Karen Bernhardt[3], Stefanie Weidtkamp-Peters[2], Elisa Ferrando-May[1]

[1]Deutsches Krebsforschungszentrum, Heidelberg;
[2]Heinrich-Heine-Universität Düsseldorf;
[3]Universität Osnabrückk;
[4]Albert-Ludwigs-Universität Freiburg;
[5]German BioImaging – Gesellschaft für Mikroskopie und Bildanalyse e.V.

The practical adoption of the widely acknowledged FAIR (Findable, Accessible, Interoperable, Reusable) principles for research data and data stewardship requires researchers and infrastructure providers to work hand-in-hand. Microscopy-driven and image-analysis-driven research projects are particularly challenging concerning FAIR data handling, data sharing, and data preservation. The multitude of imaging modalities, the large number of vendor-specific proprietary file formats, the enormous sizes of high-dimensional bioimaging data files, and the vast array of software products for processing and analysis have led to the stigma of bioimaging data being cumbersome to handle. Community-driven solutions for image data management systems, file format translation libraries, and open-source software exist in the field of bioimaging. However, many solutions are only applicable in practice if researchers have access to institutional resources for hardware and technical support for implementation, maintenance, and training. To help overcome this hurdle to the practical implementation of FAIR data management in bioimaging, the Information Infrastructure for BioImage Data project (I3D:bio) started in 2022 as a DFG-funded, collaborative effort to provide bioimaging research data management (RDM) support at universities and research institutions. The work program is based on the exchange of experience among members of German BioImaging – Society for Microscopy and Image Analysis (GerBI-GMB), the Research Data Management for Microscopy (RDM4mic) group, and the NFDI4BIOIMAGE consortium participating in the National Research

Publiziert in: Vincent Heuveline, Nina Bisheh und Philipp Kling (Hg.): E-Science-Tage 2023. Empower Your Research – Preserve Your Data. Heidelberg: heiBOOKS, 2023. DOI: https://doi.org/10.11588/heibooks.1288.c18090

Data Infrastructure (NFDI). The I3D:bio project focuses on data handling leveraging the open-source software OME Remote Objects (OMERO), offering direct and on-premise support for the implementation of bioimage data management, including storage concepts for long-term archiving, data sharing capabilities, and metadata enrichment. We are developing tailored training material and workshops. The I3D:bio project, moreover, aims to create a public resource for reusable metrology data in microscopy. Here we present the project progress, our support offers, and the lessons learned from use cases at the co-applicant sites and partner sites.

1 Introduction

Research across various disciplines relies on microscopy as a collection of techniques allowing insight into living or non-living matter with high spatial and temporal resolution. Data generated by modern microscopy is often characterized by large file sizes and considerable complexity regarding both the performed experiment and the file structure in which the acquired data is stored (Ouyang and Zimmer 2017). The lack of a standard file format in microscopy impedes reusability and accessibility for third-party users, while non-standardized or even lacking metadata about the technical setup and the experimental protocol are obstacles to the findability and interoperability of microscopy data. At present, the Open Microscopy Consortium's OME.TIFF format and the Bio-Formats translation library offer a partial solution in practice (Linkert et al. 2010). However, these approaches to data handling cannot scale with the increasing variability of imaging modalities and proprietary file formats by microscope vendors (OME 2019). Moreover, classical binary file formats are not well suited for modern microscopy and image analysis workflows that include remote storage, frequent access, and processing (Moore et al. 2021).

Large N-dimensional arrays within the image files require high network transfer rates and sufficiently large computer memory (RAM). The heterogeneity of microscopy modalities and file types is particularly challenging for individual researchers. While specialized staff at imaging core facilities train researchers for the proper use of microscope systems or even assist with the preparation and acquisition of bioimaging data, data management after the acquisition is often a sole user responsibility. Dedicated funding to support RDM for microscopy is often lacking in imaging core facilities and individual research laboratories. As a result, bioimaging data today is often hard to preserve and share in compliance with the FAIR principles (Wilkinson et al. 2016). To overcome these hurdles toward FAIRification of microscopy data, several members of the German bioimaging community, firmly grounded in the imaging core facility network German BioImaging – Society for Microscopy and Image Analysis (GerBI-GMB), initiated the "Information Infrastructure for BioImage Data" (I3D:bio), as a bottom-up project promoting the direct benefit of community-tested image data handling software as well as metadata annotation tools and guidelines.[1]

1 `https://www.i3dbio.de`

2 Resources

I3D:bio was proposed by imaging core facility managers as a collaboration of four German institutions. Regular exchange in the open community group "Research Data Management for Microscopy" (RDM4mic) identified common issues and pitfalls in microscopy data handling[2]. Several core facilities have implemented centralized instances of the image data management software OME Remote Objects (OMERO) between 2017 and 2022 (Burel et al. 2015; Zobel, Weischer, and Wendt 2022; Kunis, Bernhardt, and Hensel 2023).

The 2021 NFDI4BIOIMAGE community survey furthermore confirmed that OMERO was the so far best-known and most widely used image data management software in the (German) bioimaging community, evincing OMERO as common ground for starting a bioimage RDM initiative (Schmidt et al. 2022). While OMERO greatly facilitates structuring, sharing, and annotating imaging data as compared with classical file folder hierarchies, running OMERO as a facility requires investing in personnel capacity and IT resources that are not readily available for many core facilities in Germany. Therefore, a goal of I3D:bio is to capacitate core facilities for implementing and managing OMERO instances by offering support and guidance for this process. Additionally, core facility users must be trained to become proficient in the use of OMERO, which is based on object storage instead of file hierarchies. Questions of data ownership, storage security, or access to OMERO-hosted data for image processing and analysis were observed as impediments to the practical adoption by users. I3D:bio intends to provide best-practice, peer-reviewed training material as a resource for core-facility-centric user training in addition to the OMERO guides[3]. Additionally, annotation tools and guidance to comply with community-driven metadata enrichment standards are being tested in use cases and at the applicant sites. These efforts are orchestrated to align with the activities of the international bioimaging community. To reach these goals, the three work packages of I3D:bio focus, first, on the deployment of OMERO instances according to best practices as well as building a public OMERO-based database for microscopy metrology data, second, the identification of the suitable technical infrastructure requirements for image data storage in OMERO and metadata annotation software, and, third, the coordination with (international) partners, communication and training (Figure 1). The project duration is three years; a prolongation is possible

3 Support

I3D:bio lays groundwork and complements the goals of the collaborating NFDI4BIOIMAGE consortium[4], which is part of the German National Research Data Infrastructure (NFDI[5]). Several NFDI consortia from the life sciences support the I3D:bio project and have stated their intention to implement I3D:bio guidelines for bioimaging RDM work-

2 `https://german-bioimaging.github.io/RDM4mic.github.io`
3 `https://omero-guides.readthedocs.io/en/latest`
4 `https://nfdi4bioimage.de`
5 `https://nfdi.de`

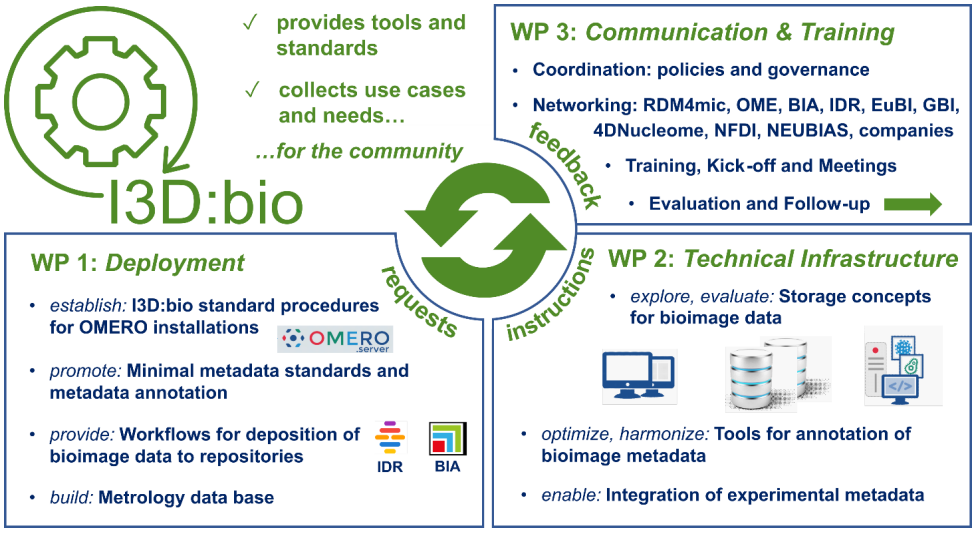

Figure 1: I3D:bio work packages. Taken from the I3D:bio project proposal.

flows in their field of research. To test and refine I3D:bio's approach to implementing OMERO at research institutions and universities, two naïve sites have been selected for the de novo installation of OMERO. The applicant institution German Cancer Research Center supports the project with in-kind contributions for the institution-wide adoption of OMERO in the field of cancer research. The Technical University of Dresden's Medical Faculty serves as an external naïve site for testing and refining I3D:bio-guided image data management workflows. Moreover, I3D:bio works in close collaboration with members of the RDM4mic group, who provide input for the project and constitute the connection to the broader target audience whose members may become integrated into the project as use cases. At the international level, I3D:bio collaborates with community partners like Euro-BioImaging[6], BioImaging North America[7], the Quality Assessment and Reproducibility of Instruments and Images in Light Microscopy (QUAREP-LiMi) group[8], and others. For example, two joint workshops for metadata annotation for bioimaging were offered together with BINA at the European Light Microscopy Initiative (ELMI) conference in 2023[9].

4 Conclusions

Within its first year, I3D:bio has collected community input via RDM4mic and several use case partners in and beyond Germany. The website[10] was established as a central knowledge resource on bioimaging RDM, including an overview of different topics and solutions, links to training resources, and an introduction to the complex topic of bioimaging

6 https://www.eurobioimaging.eu
7 https://www.bioimagingnorthamerica.org
8 https://quarep.org
9 https://elmi2023.eu
10 https://www.i3dbio.de

metadata. As a live assessment of FAIR bioimaging data collection and annotation, the I3D:bio team has organized the management of microscopy data produced during practical workshops at the Trends in Microscopy 2023 Conference and invited participants and workshop providers to annotate imaging data based on the Recommended Metadata for Biological Images (REMBI) guidelines (Sarkans et al. 2021) for OMERO-based bioimage data handling and metadata annotation.

The I3D:bio website offers a Help Desk, and community stakeholders are invited to contact I3D:bio via the website or via the community forum image.sc to collaborate as a partner or use case.

Acknowledgements

We thank the I3D:bio project partners, in particular, J. Moore and T. Zobel, and the RDM4mic group for input and collaboration. German BioImaging is acknowledged for its contribution of resources and community integration. This work is funded by the Deutsche Forschungsgemeinschaft (German Research Foundation), project number 462231789.

References

Burel, Jean-Marie, Sébastien Besson, Colin Blackburn, Mark Carroll, Richard K. Ferguson, Helen Flynn, Kenneth Gillen, et al. 2015. "Publishing and sharing multi-dimensional image data with OMERO". *Mammalian Genome* 26 (9-10): 441–447. DOI: https://doi.org/10.1007/s00335-015-9587-6.

Kunis, Susanne, Karen Bernhardt, and Michael Hensel. 2023. "Setting up a data management infrastructure for bioimaging". *Biological Chemistry* 404 (5): 433–439. DOI: https://doi.org/10.1515/hsz-2022-0304.

Linkert, Melissa, Curtis T. Rueden, Chris Allan, Jean-Marie Burel, Will Moore, Andrew Patterson, Brian Loranger, et al. 2010. "Metadata matters: access to image data in the real world". *Journal of Cell Biology* 189 (5): 777–782. DOI: https://doi.org/10.1083/jcb.201004104.

Moore, Josh, Chris Allan, Sébastien Besson, Jean-Marie Burel, Erin Diel, David Gault, Kevin Kozlowski, et al. 2021. "OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies". *Nature Methods* 18 (12): 1496–1498. DOI: https://doi.org/10.1038/s41592-021-01326-w.

OME. 2019. "OME's position regarding file formats (Blog post)". Visited on May 9, 2023. https://www.openmicroscopy.org/2019/06/25/formats.html.

Ouyang, Wei, and Christophe Zimmer. 2017. "The imaging tsunami: Computational opportunities and challenges". *Current Opinion in Systems Biology* 4:105–113. DOI: https://doi.org/10.1016/j.coisb.2017.07.011.

Sarkans, Ugis, Wah Chiu, Lucy Collinson, Michele C. Darrow, Jan Ellenberg, David Grunwald, Jean-Karim Hériché, et al. 2021. "REMBI: Recommended Metadata for Biological Images – enabling reuse of microscopy data in biology". *Nature Methods* 18 (12): 1418–1422. DOI: https://doi.org/10.1038/s41592-021-01166-8.

Schmidt, Christian, Janina Hanne, Josh Moore, Christian Meesters, Elisa Ferrando-May, and Stefanie Weidtkamp-Peters and. 2022. "Research data management for bioimaging: the 2021 NFDI4BIOIMAGE community survey". *F1000Research* 11:638. DOI: https://doi.org/10.12688/f1000research.121714.2.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersbergand, Gabrielle Appleton, Myles Axtonand, Arie Baakand, Niklas Blombergand, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific data* 3 (1): 1–9. DOI: https://doi.org/10.1038/sdata.2016.18.

Zobel, Thomas, Sarah Weischer, and Jens Wendt. 2022. *OMERO for microscopy research data management - A use case example from the Münster Imaging Network.* Technical report. Wiley Analytical Science. https://analyticalscience.wiley.com/do/10.1002/was.0004000267.