



NFDI 4
BIOIMAGE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE
FOR MICROSCOPY AND BIOIMAGE ANALYSIS

ScaDS.AI
DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE



TECHNISCHE
UNIVERSITÄT
DRESDEN

UNIVERSITÄT
LEIPZIG

Towards Preservation of Life Science Data with NFDI4BIOIMAGE

Robert Haase

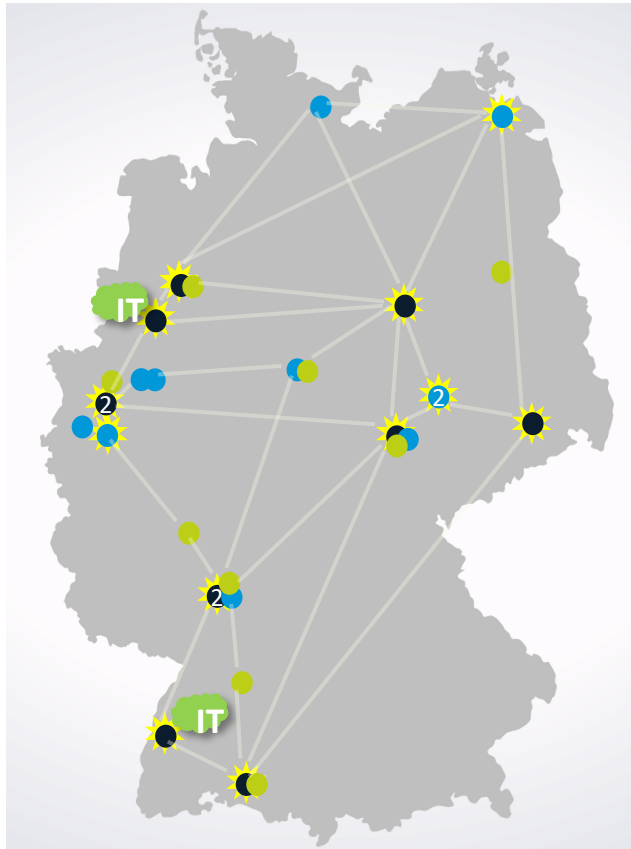
ScaDS.AI, Leipzig University



With the exception of logos and cited third-party content, the content of these slides is shared under the terms of the [Creative Commons Attribution License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/) unless the content is marked otherwise.

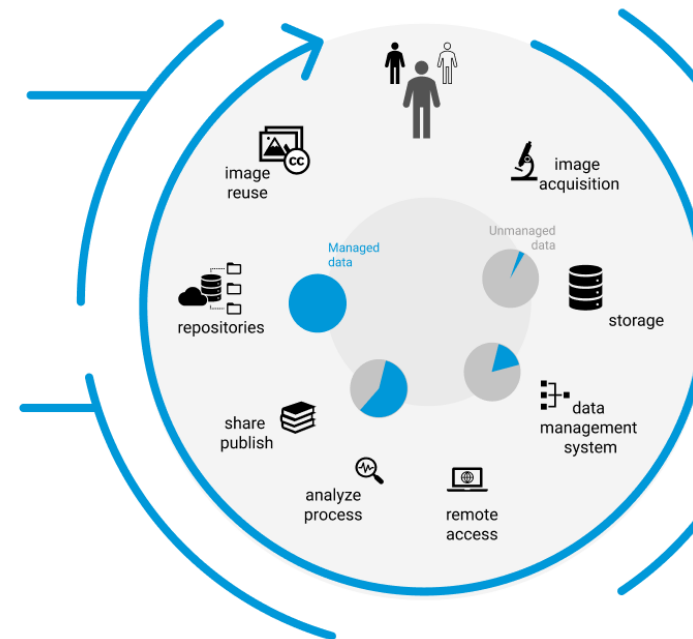
NFDI4BIOIMAGE Core mission

- Enable FAIR bioimage data management for German researchers, across disciplines and embedded in the international framework.



Objective 4
Capacitate researchers
for FAIR image data
management

Objective 3
Maximize the reach
of **reproducible** image
analysis workflows
in the community



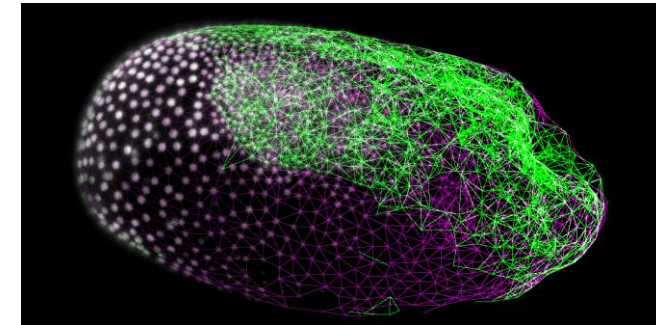
Objective 1
Champion the
standardization
of the „bioimage
data“ type

Objective 2
Provide scalable
infrastructure
for FAIR image data

What types of research data do we want to keep?

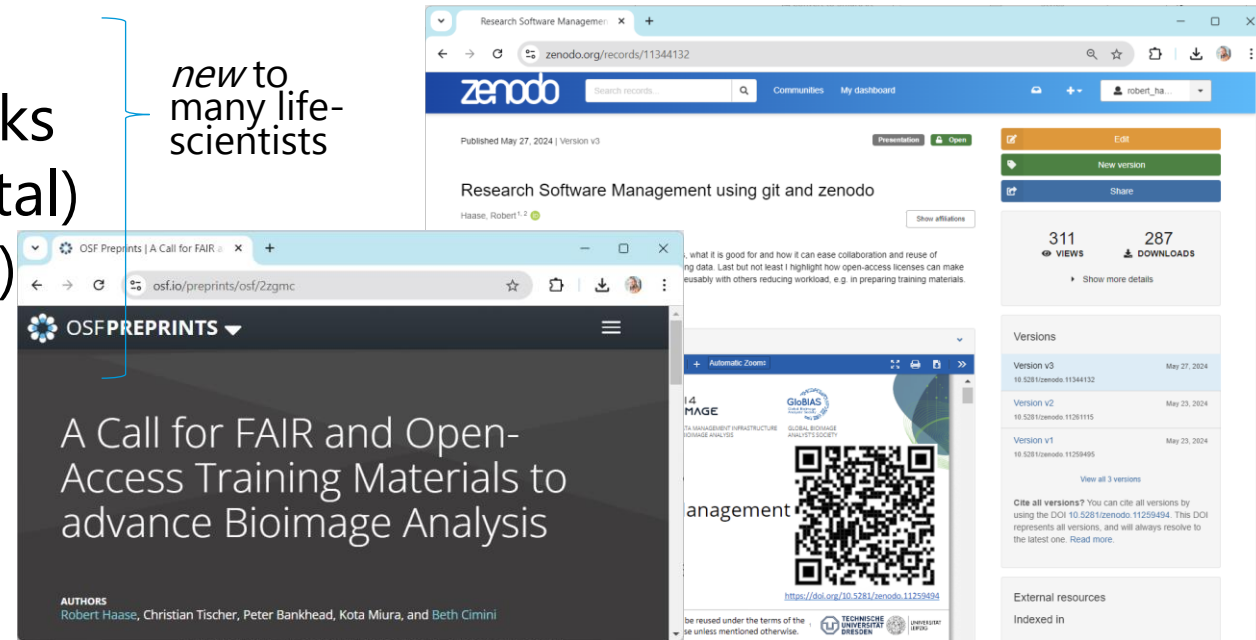
- Huge variety of imaging data [types]
 - 2D RGB images, from thousands of experiments
 - 3D electron microscopy data
 - 3D timelapse data
 - Uncommon data types (point-based, hyperspectral, ...)
- Code
 - Analysis scripts, Jupyter notebooks
 - Workflows (nextflow, galaxy, fractal)
 - Environments (docker, conda, Fiji)
 - Modular, re-usable software
- Training materials
 - Research data management
 - Image analysis basics
 - Machine / Deep Learning

} Terabyte range



Tribolium castaneum, lightsheet microscopy

*new to
many life-
scientists*



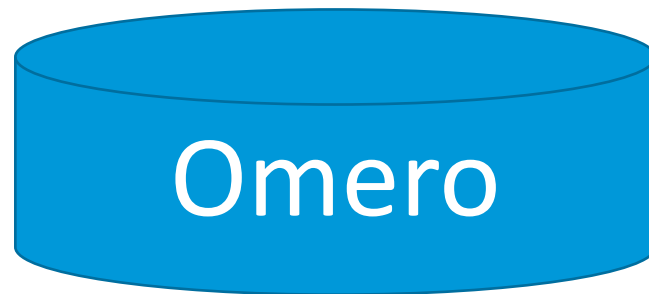
<https://osf.io/preprints/osf/2zgmcc>

<https://zenodo.org/records/11344132>

Which data is preserved where?

- Imaging data
 - Bioimage Archive
 - [Image Data Resource]
 - Zenodo

Preferred: institutional Omero server

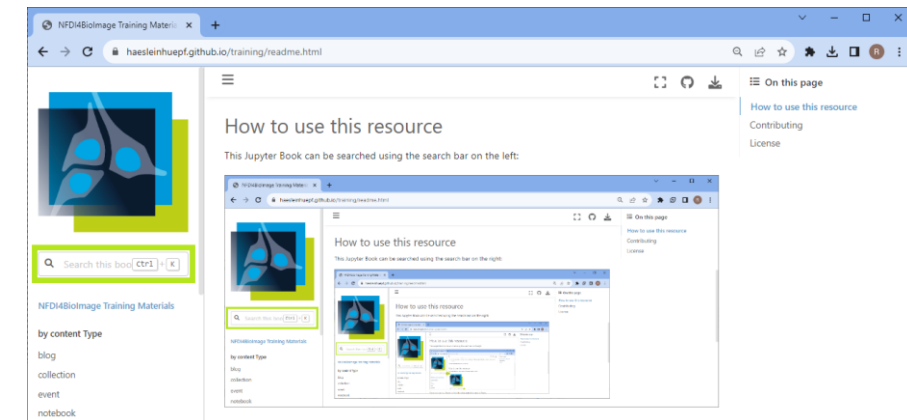


- Code
 - Github
 - Zenodo
 - Protocols.io

If nothing above works: institutional git server



- Training materials
 - Zenodo
 - F1000
 - Github
- No matter where, we collect links in an index:



<https://www.openmicroscopy.org/omero/>
<https://www.ebi.ac.uk/bioimage-archive/>

<https://nfdi4bioimage.github.io/training>

How do we make sure the data is safe as well as re-usable?

- Task-Area 1: Image (meta)data formats and standardization

- M1.1 Define FAIR Image Objects (FAIR-IO)
- M1.2 Increase usability of Linked Metadata
- M1.3 Enabled cloud-native image formats
- M1.4 Formalize FAIR-IO RFC process
- M1.5 Implement community requirements & use cases

- Task-Area 2: Technical infrastructure and cloud resources

- M2.1 Infrastructure as a Service
- M2.2 Community Software as a Service
- M2.3 Development and extension of community services
- M2.4 Desktop as a Service

- Task-Area 5: Training and community integration

- M5.1 NFDI4BIOIMAGE training portfolio
- M5.2 Implement a cross-site search index for bioimaging RDM training resources
- M5.3 Community-driven process for curation and long-term viability of training materials
- M5.4 Data Stewardship and Help Desk activities

Who is in charge at the different stages in data lifecycle?

- Responsible: The PI
- In charge: The scientist, overseen by PI, potentially delegated to senior scientist (defined in: Data Management Plan)
- Data Stewards don't take care of project specific data



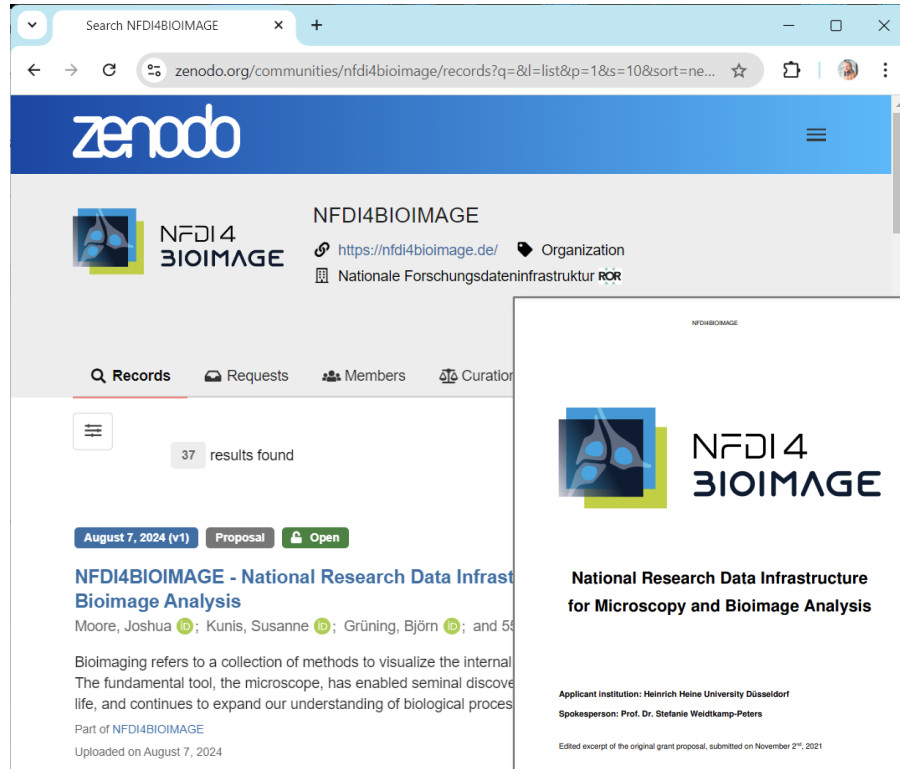
Created with Dall-E 3

Challenges

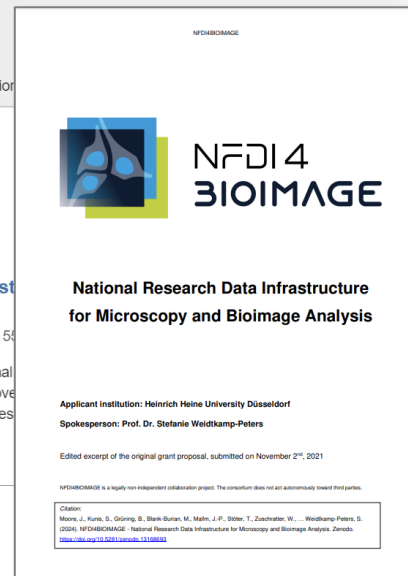
- Lack of incentives for sharing data
 - Young PIs care most about re-usability of data
- Different philosophies in the community
 - Open source software developers are “all in” when it comes to sharing
 - Experimentalists with “hot” research data, they are afraid of sharing (“Available on request”)
- Everyone has heard about the FAIR principles, but scientists don’t know:
 - Where to publish data
 - What license to use
 - Who *owns* the data
 - Who to talk to, to figure things out

Communication is key

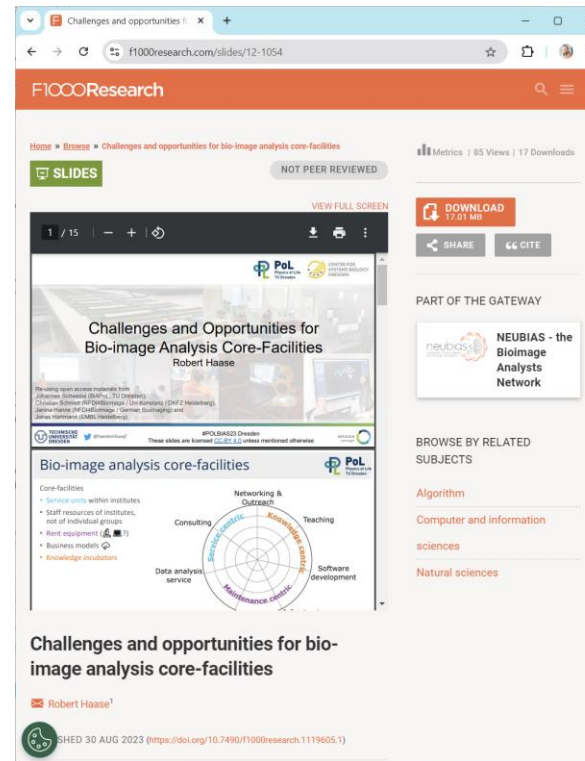
- Spreading the word about best-practices, on various platforms



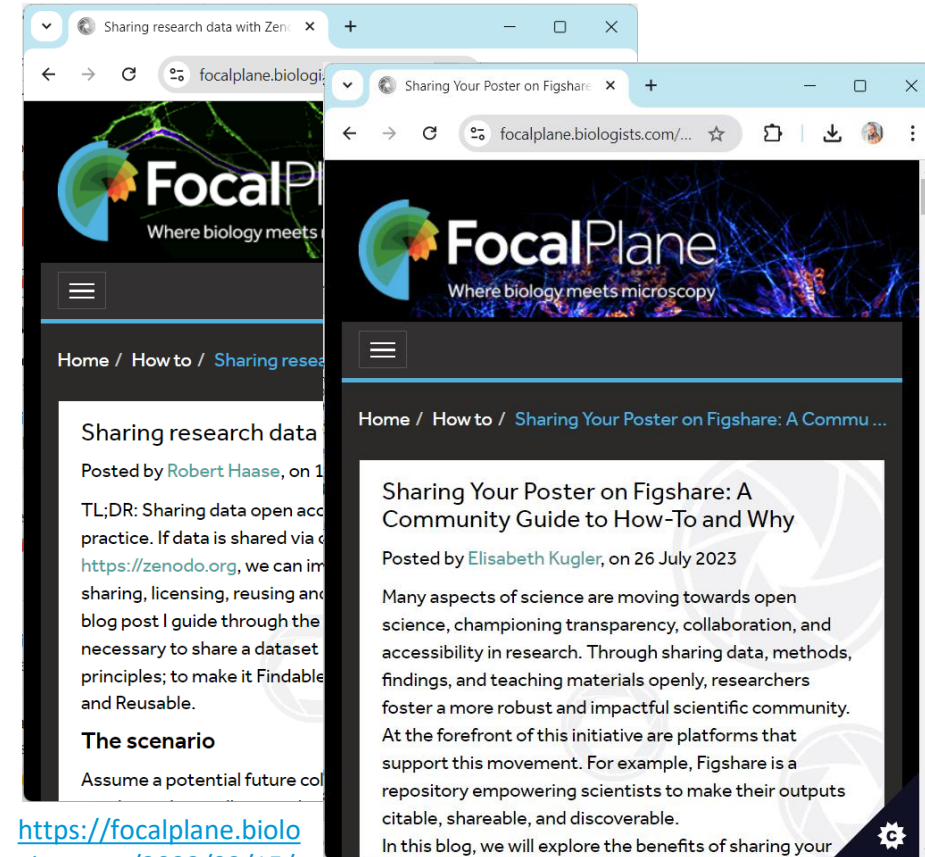
<https://zenodo.org/communities/nfdi4bioimage>



<https://zenodo.org/records/13168693>



<https://f1000research.com/slides/12-1054>

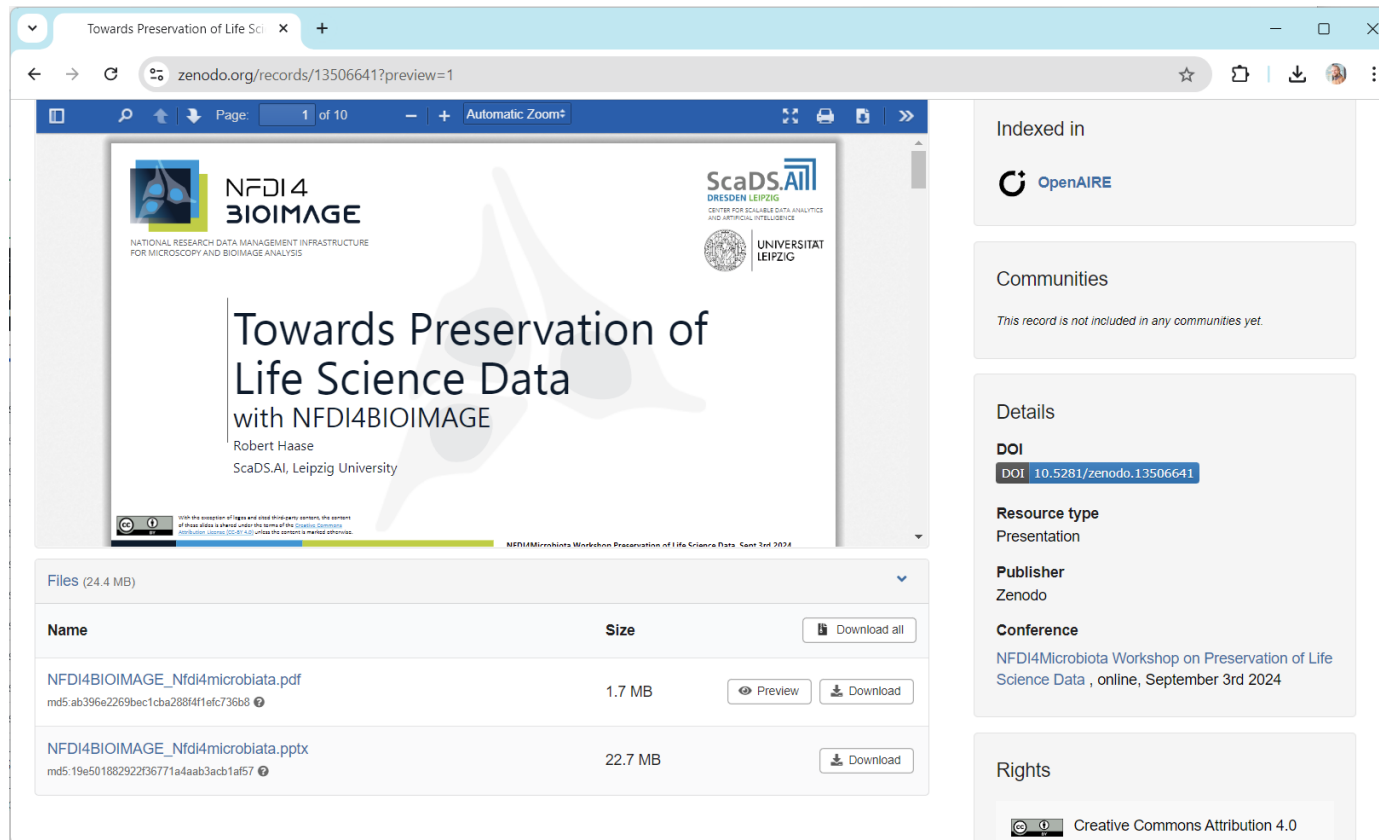


<https://focalplane.biologists.com/2023/02/15/sharing-research-data-with-zenodo/>

<https://focalplane.biologists.com/2023/07/26/sharing-your-poster-on-figshare/>

Communication is key

- Living a good example:
If us RDM-experts don't share things, why should our community?



The screenshot displays a Zenodo record page for a presentation titled "Towards Preservation of Life Science Data with NFDI4BIOIMAGE" by Robert Haase from ScaDS.AI, Leipzig University. The page includes a cover image, a list of files for download (a PDF and a PPTX), and various metadata fields such as the DOI (10.5281/zenodo.13506641), OpenAIRE index, and Creative Commons Attribution 4.0 license.



<https://doi.org/10.5281/zenodo.13506641>

Acknowledgements



NFDI4BIOIMAGE

- task area leads
- Participants
- coordinators
- data stewards
- research software engineers
- (international) collaborators

All-hands Meeting NFDI4BIOIMAGE, Oct. 23, Düsseldorf

Funded by the Deutsche
Forschungsgemeinschaft (DFG, German
Research Foundation) under the
National Research Data Infrastructure –
NFDI 46/1 – 501864659

DFG
Deutsche
Forschungsgemeinschaft

In cooperation with

nfdi

ScaDS.AI
DRESDEN LEIPZIG



UNIVERSITÄT
LEIPZIG

We acknowledge the financial support by the
Federal Ministry of Education and Research of
Germany and by Sächsische Staatsministerium
für Wissenschaft, Kultur und Tourismus in the
programme Center of Excellence for AI-research
„Center for Scalable Data Analytics and Artificial
Intelligence Dresden/Leipzig“, project
identification number: ScaDS.AI