Towards Preservation of Life Science Data
with NFDI4BIOIMAGE

Robert Haase

ScaDS.AI, Leipzig University

NFDI4BIOIMAGE Core mission

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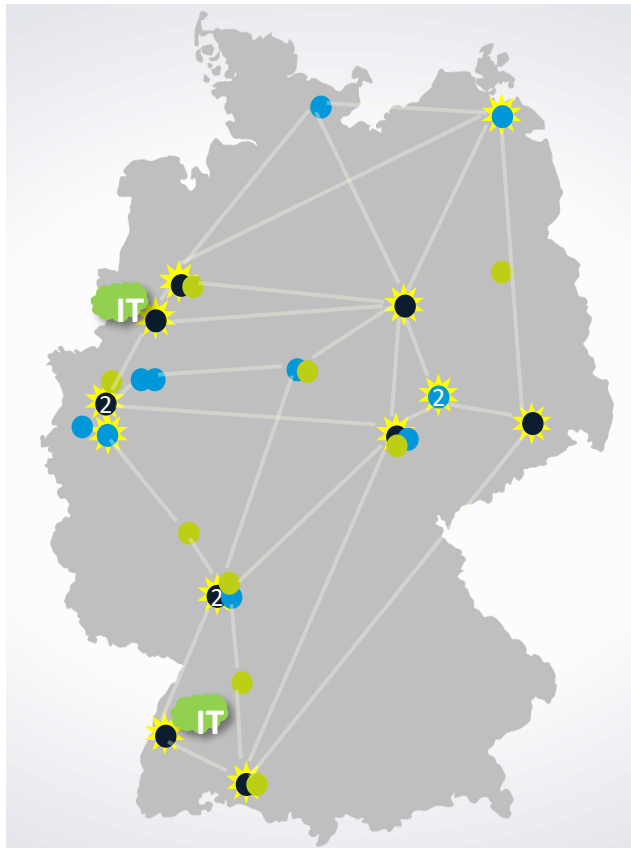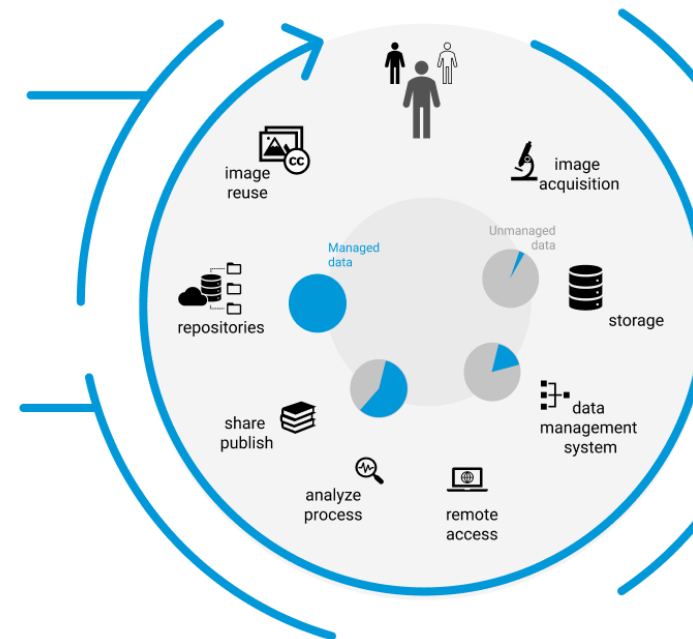

Objective 4
Capacitate researchers for FAIR image data management

Objective 3
Maximize the reach of **reproducible** image analysis workflows in the community
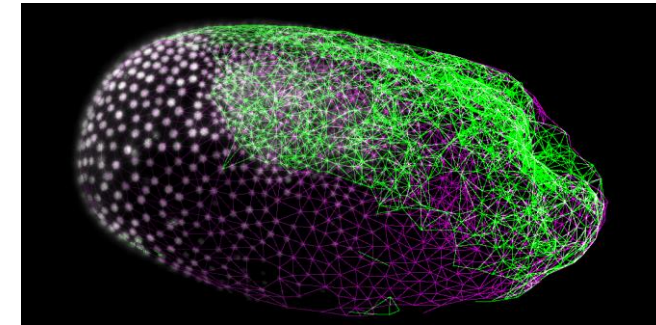
Objective 1
Champion the **standardization** of the „bioimage data" type

Objective 2
Provide scalable **infrastructure** for FAIR image data

image reuse

image acquisition

Managed data

Unmanaged data

repositories

storage

share publish

data management system

analyze process

remote access

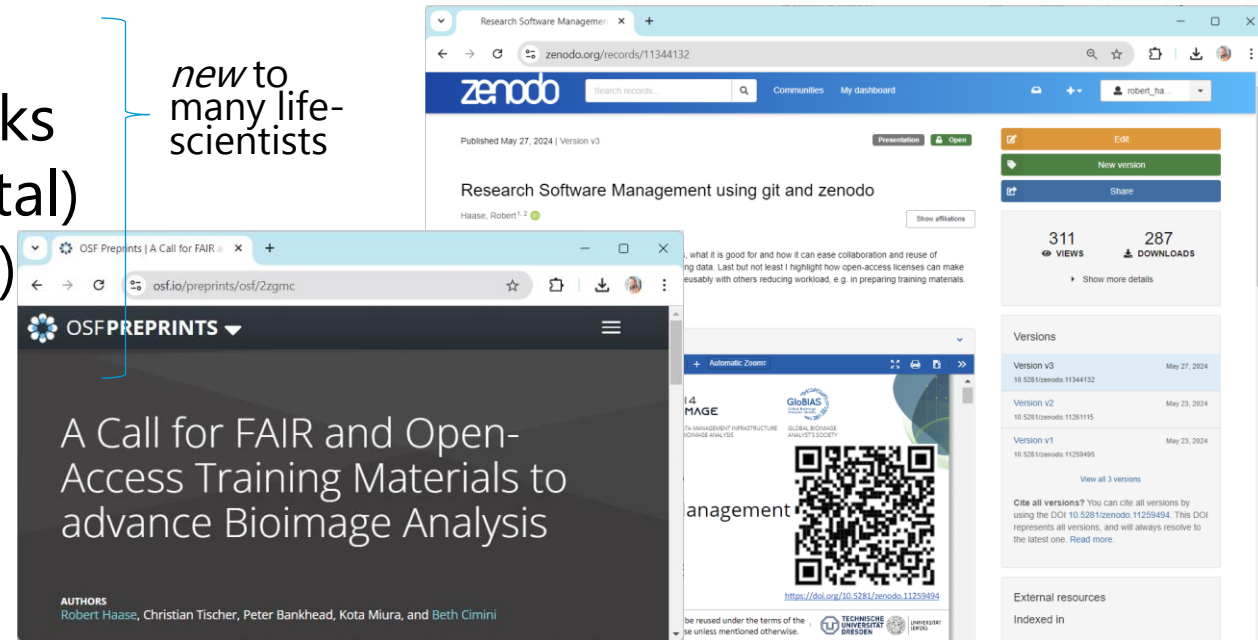# What types of research data do we want to keep?

- Huge variety of imaging data [types]
  - 2D RGB images, from thousands of experiments
  - 3D electron microscopy data } Terabyte range
  - 3D timelapse data
  - Uncommon data types (point-based, hyperspectral, …)

Tribolium castaneum, lightsheet microscopy

- Code
  - Analysis scripts, Jupyter notebooks
  - Workflows (nextflow, galaxy, fractal)
  - Environments (docker, conda, Fiji)
  - Modular, re-usable software

new to many life-scientists

- Training materials
  - Research data management
  - Image analysis basics
  - Machine / Deep Learning

https://osf.io/preprints/osf/2zgmc   https://zenodo.org/records/11344132

Which data is preserved where?

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

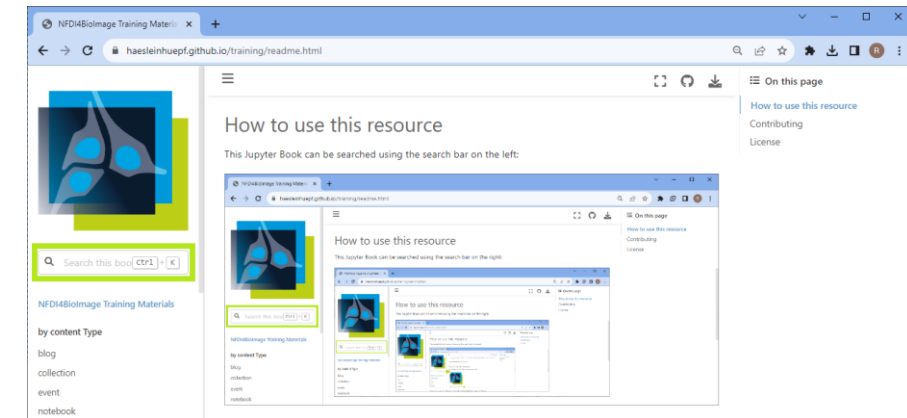Preferred: institutional Omero server

- 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://nfdi4bioimage.github.io/training

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

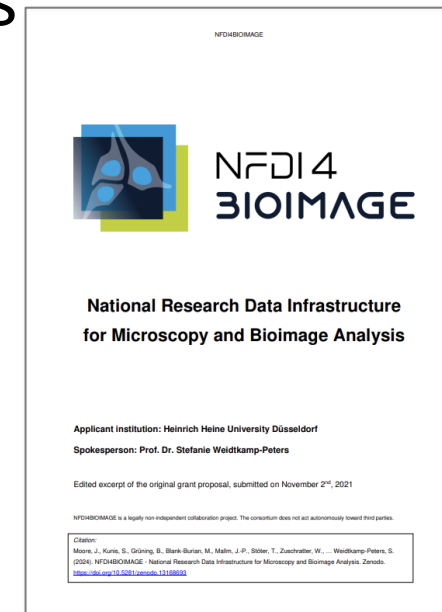- ## Task-Area 1: Image (meta)data formats and standardization

  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

NFDI4BIOIMAGE

National Research Data Infrastructure for Microscopy and Bioimage Analysis

Applicant institution: Heinrich Heine University Düsseldorf

Spokesperson: Prof. Dr. Stefanie Weidtkamp-Peters

Edited excerpt of the original grant proposal, submitted on November 2nd, 2021

NFDI4BIOIMAGE is a legally non independent collaboration project. The consortium does not act autonomously toward third parties.

Citation:
Moore, J., Kunis, S., Grüning, B., Blank-Burian, M., Malm, J.-P., Stöter, T., Zuschratter, W., ... Weidtkamp-Peters, S. (2024). NFDI4BIOIMAGE - National Research Data Infrastructure for Microscopy and Bioimage Analysis. Zenodo. https://doi.org/10.5281/zenodo.13168693

https://zenodo.org/records/13168693

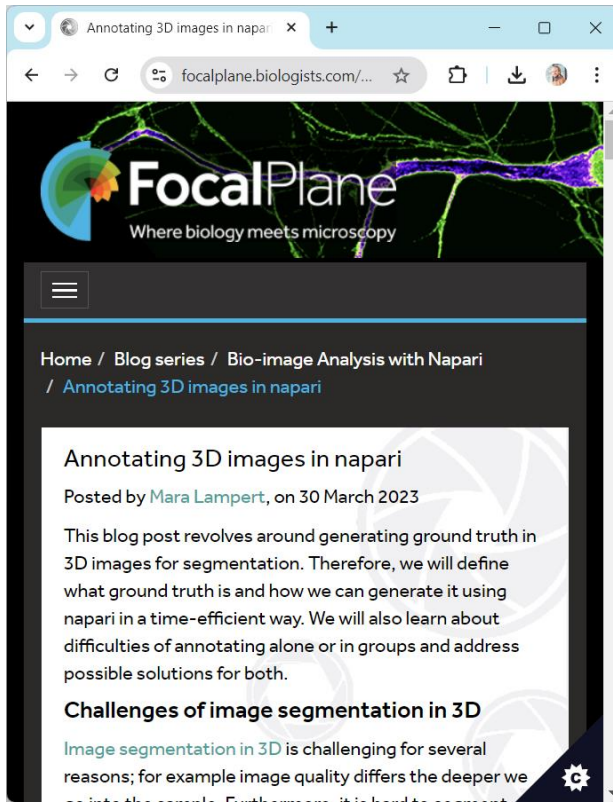# Who is in charge at the different stages in data lifecycle?

- 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

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 close to the community

https://focalplane.biologists.com/2023/03/30/annotating-3d-images-in-napari/

https://focalplane.biologists.com/2021/09/04/collaborative-bio-image-analysis-script-editing-with-git/

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

https://focalplane.biologists.com/2023/07/26/sharing-your-poster-on-figshare/
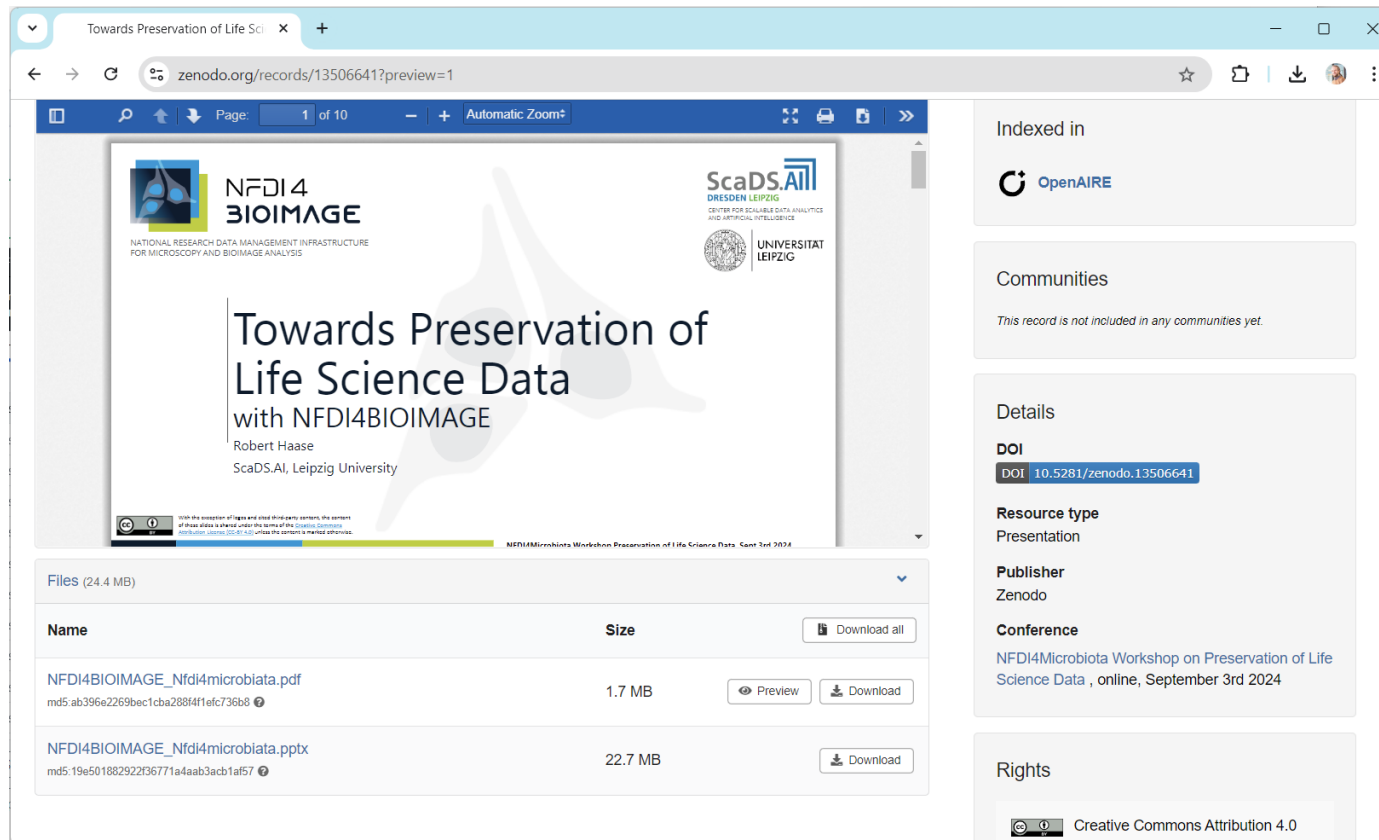
Communication is key

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

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

Acknowledgements

All-hands Meeting NFDI4BIOIMAGE, Oct. 23, Düsseldorf

NFDI4BIOIMAGE

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

DFG
Deutsche Forschungsgemeinschaft

In cooperation with

nfdi