

Participate
online



Join directly

or at



NFDI4
BIOIMAGE

Introduction to Bioimaging Research Data Management, NFDI4BIOIMAGE and I3D:bio

Workshop on Research Data Management for Microscopy and Image Analysis
GBM Compact Symposium 2024

Christian Schmidt, DKFZ Heidelberg
Project Coordinator, NFDI4BIOIMAGE and I3D:bio

Slides available at Zenodo, DOI: 10.5281/zenodo.13861026



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.

Data Management?

- Store
- Organize
- Annotate
- Process
- Share
- Publish
- Reuse



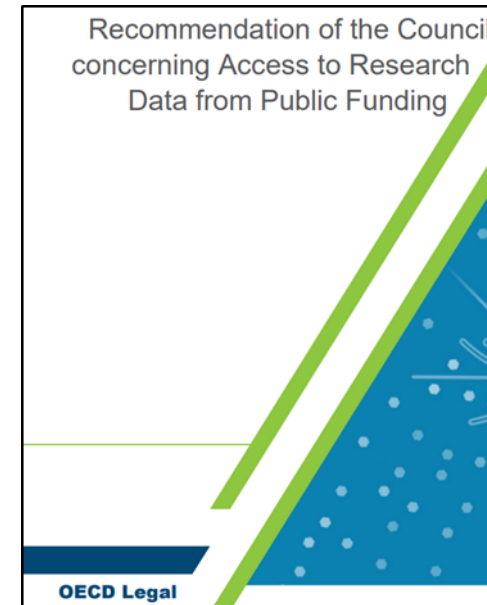
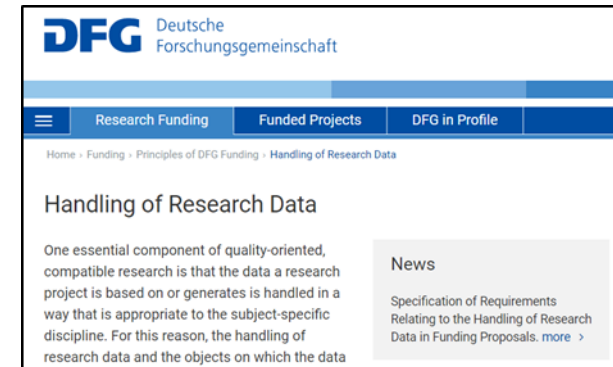
Why invest into Research Data Management?

Wordcloud Poll



Significance of research data management

- **Intrinsic motivation**
- Impact and efficiency of my research
- **Extrinsic factors**
- Funding agency demands
- Good Scientific Practice
- **Ethical aspect**
- Make science more sustainable
- Promote *Open Science* and data sharing

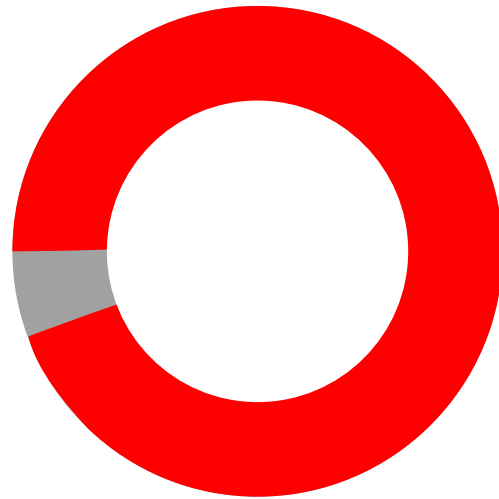


- https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html
- https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf
- https://twitter.com/BMBF_Bund/status/1571801906074337280?s=20&t=krDcwOPMuPwjs-VisYBgVg
- <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347>

Research Data Management Challenges

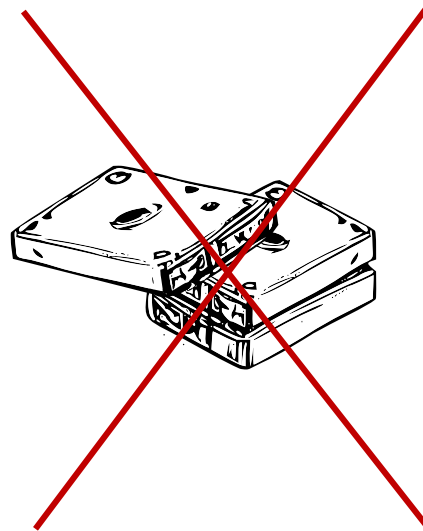
File storage and organization: The data deluge

A *lot* of data can pile up...

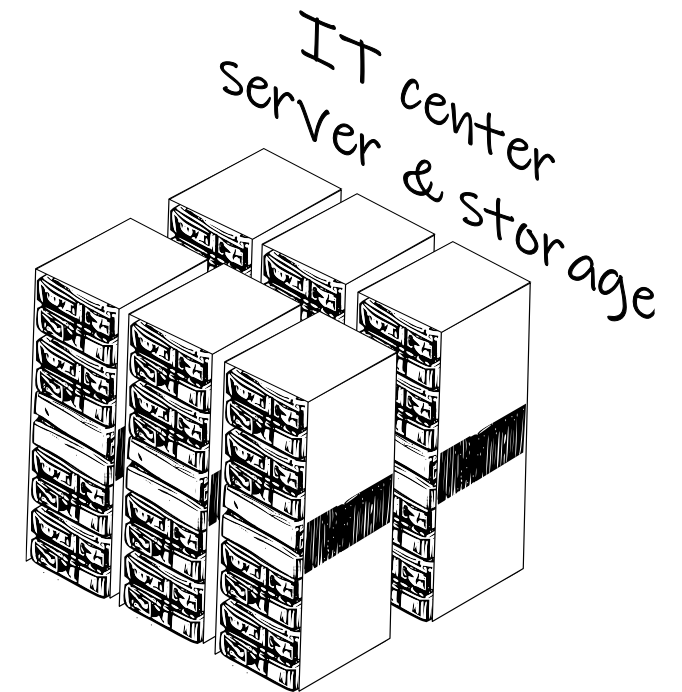


file system X:
945 GB (used) / 51 GB (free)

... external hard drives?
Could get lost or damaged!



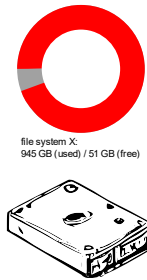
Institutional, backed-up storage
(often demanded by IT policy)



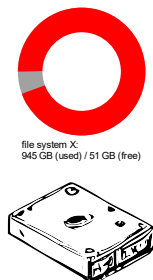
Research Data Management Challenges

The risk of data fragmentation

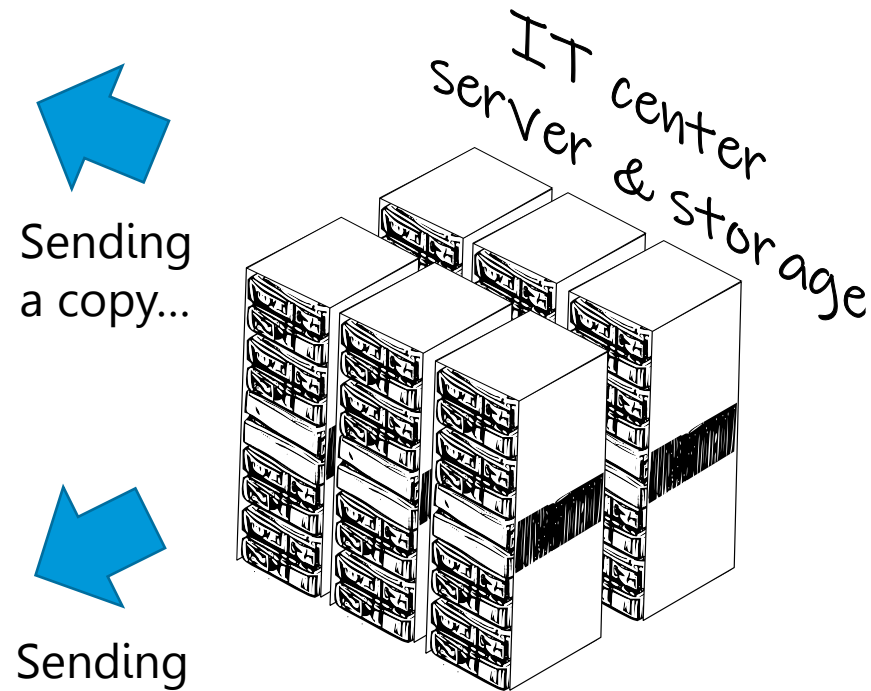
An external collaborator
doing image analysis



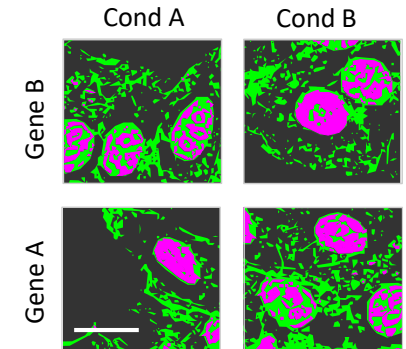
A lab colleague
needs the data on a
special workstation



Institutional, backed-up storage



Interim results to
show in Lab meeting /
seminars, etc.



Derived
data saved
elsewhere

Figure with
JPEG exports
stored locally

Research Data Management Challenges

Is the data understandable and reproducible?

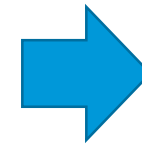
Paper submitted!



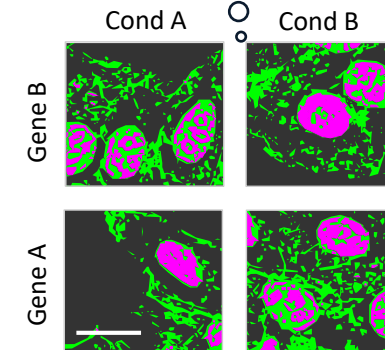
Reviewer:
„Please show
reproducibility of
your analysis
pipeline.“

Editor:
„Please provide
the underlying
original data.“

Repository Provider:
„Please enrich your
data with a complete
set of metadata.“



How exactly was this
analyzed?



What metadata?

Which were the
original file(s) here?

Research Data Management Challenges

Is data trustworthy and reusable for others?

Paper published. What else could happen?



Other researcher:

„We have developed a new tool and would like to reuse your data for our research“



Reusable research data might increase your scientific impact! *



Science Sleuth / Integrity check algorithms / AI:

„There is an indication that the figure was illegitimately manipulated“

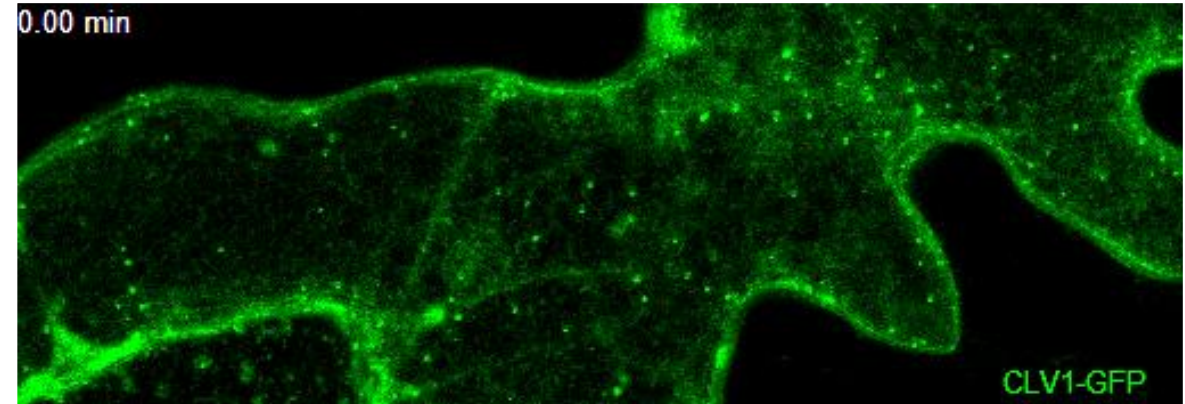


RDM and documentation are measures of scientific integrity and protect from false accusations

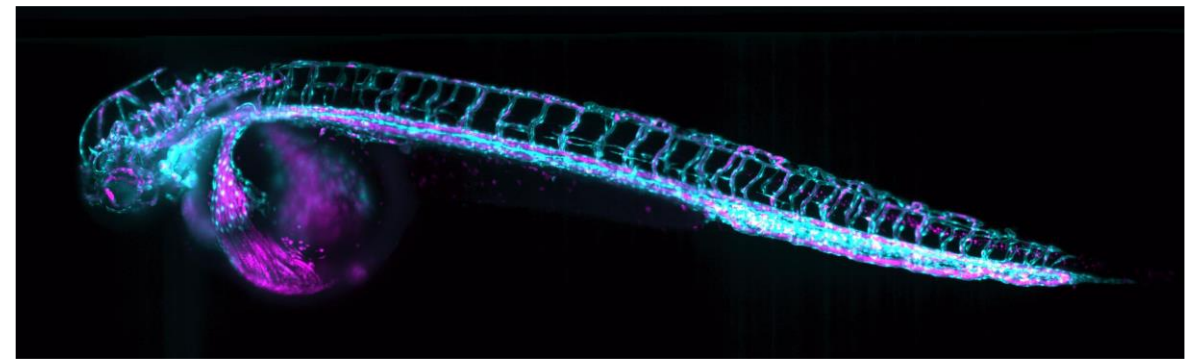
The bioimage data type

Microscopy data is (often):

- high-dimensional (X, Y, Z, Channel, Time, ...)
- saved in proprietary file formats
- of large file size (often GB- or even TB-range)
- produced with complex experimental setups
- used for quantitative analysis → derived data
- ... i.e., **cumbersome to manage?**



Courtesy of: Y. Stahl, S. Weidtkamp-Peters, HHU Düsseldorf



Courtesy of: Jan Huiskens, University of Göttingen

What defines „good“ data management?

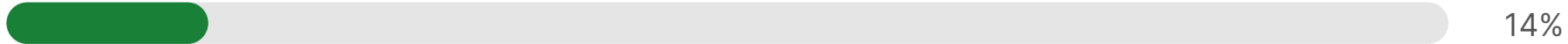




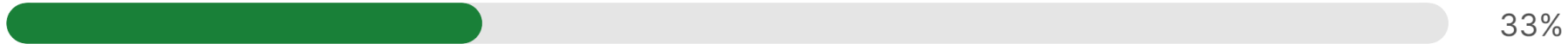
The term "FAIR Principles"...

Multiple Choice Poll

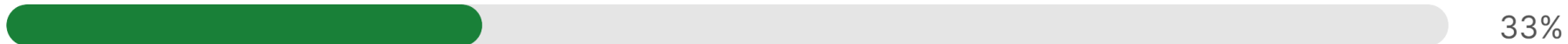
is unknown to me



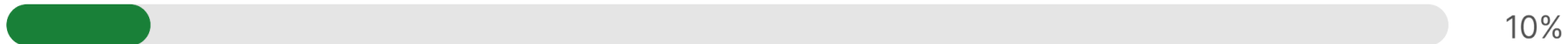
sounds familiar but I cannot explain it's meaning



is familiar, I can explain the meaning to some extent



is clear to me, I can explain it in detail.



is clear, and I already publish (most of) my data according to the FAIR principles.



The FAIR principles for data and data stewardship

Findable

Accessible

Interoperable

Reusable

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

Guiding principles

(not a dictation of technical solutions)

Goal is machine-readability

(not only human ability to reuse)

- Wilkinson et al., 2016, Scientific Data, DOI: 10.1038/sdata.2016.18, CC-BY 4.0 (<http://creativecommons.org/licenses/by/4.0>)
- Jacobsen et al., 2020, FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence, DOI: 10.1162/dint_r_00024

The FAIR principles for data and data stewardship

Findable

Accessible

Interoperable

Reusable

Box 2 | The FAIR Guiding Principles

To be Findable

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described using rich metadata
- F3. metadata are available for search and retrieval
- F4. (meta)data are associated with detailed provenance

To be Accessible

- A1. (meta)data are available to all
- A1.1 the data are available to all
- A1.2 the metadata are available to all
- A2. metadata are available to all

To be Interoperable

- I1. (meta)data are available to all
- I2. (meta)data are available to all
- I3. (meta)data are available to all

To be Reusable

- R1. meta(data) are released with a clear and accessible data usage license
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

**„AI’s Dirty Little Secret:
Without FAIR Data, It’s Just Fancy Math“**

– Josh Moore, Susanne Kunis, ELMI 2024
(<https://zenodo.org/doi/10.5281/zenodo.11235512>)

- Wilkinson et al., 2016, Scientific Data, DOI: 10.1038/sdata.2016.18, CC-BY 4.0 (<http://creativecommons.org/licenses/by/4.0>)
- Jacobsen et al., 2020, FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence, DOI: 10.1162/dint_r_00024

The FAIR principles for data and data stewardship

Findable

Accessible

Interoperable

Reusable

Box 2 | The FAIR Guiding Principles

To be Findable

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described in a rich, structured metadata
- F3. metadata are registered in a searchable, open registry
- F4. (meta)data are associated with detailed provenance

To be Accessible

- A1. (meta)data are available to all
- A1.1 the data are available to all
- A1.2 the data are available to all
- A2. metadata are available to all

To be Interoperable

- I1. (meta)data are available to all
- I2. (meta)data are available to all
- I3. (meta)data are available to all

To be Reusable

- R1. (meta)data are released with a clear and accessible data usage license
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

„Most of the world’s bioimaging data lacks a clear path to being shared.“

– Josh Moore, Susanne Kunis, ELMI 2024

(<https://zenodo.org/doi/10.5281/zenodo.11235512>)

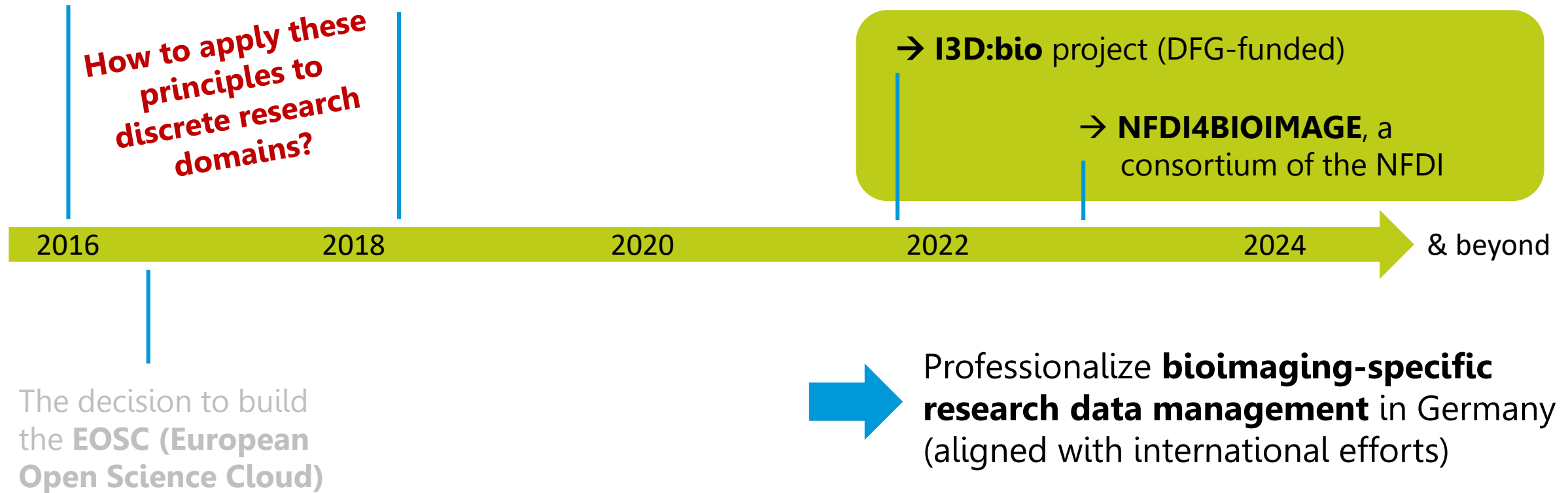
- Wilkinson et al., 2016, Scientific Data, DOI: 10.1038/sdata.2016.18, CC-BY 4.0 (<http://creativecommons.org/licenses/by/4.0>)
- Jacobsen et al., 2020, FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence, DOI: 10.1162/dint_r_00024

What is needed to make data reusable?

The „FAIR
principles“
published

In Germany:

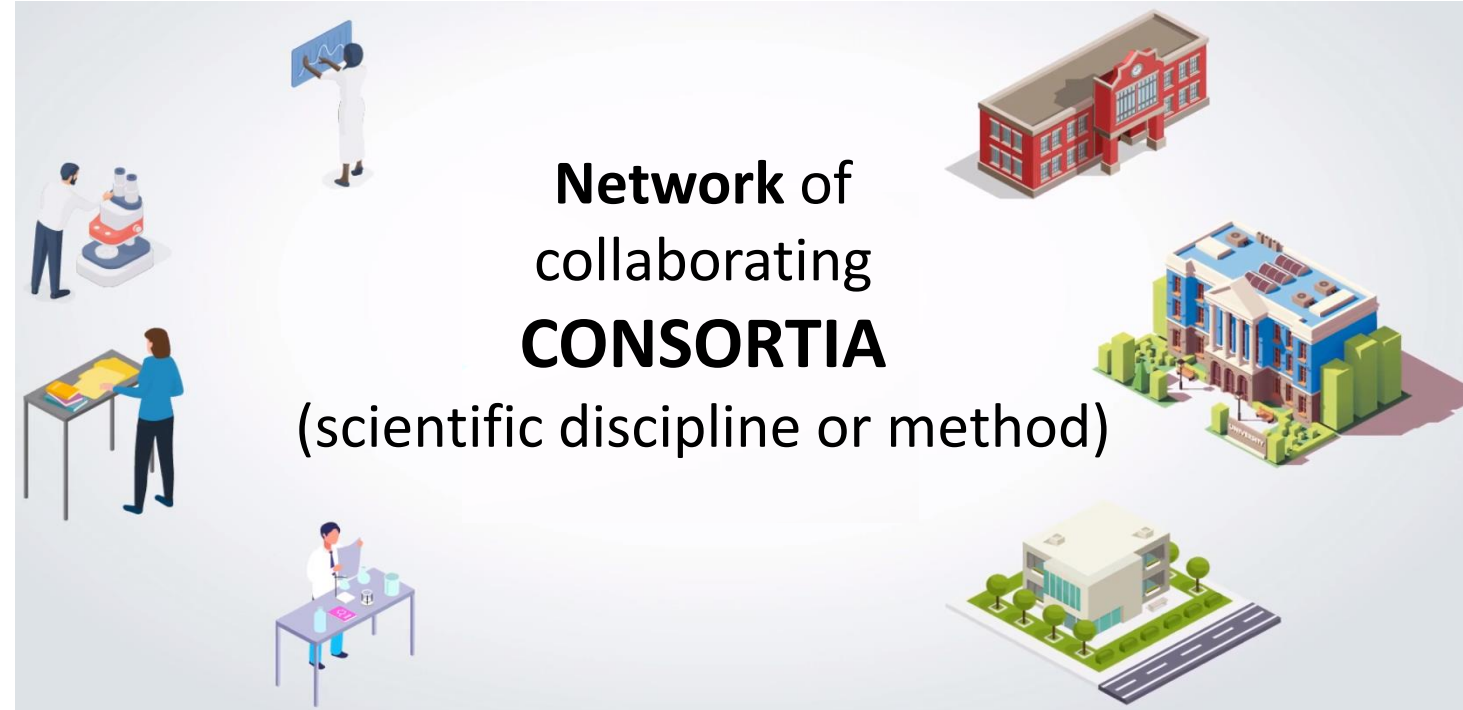
- Small-scale funding lines for data infrastructures
- Building the **Nationale Forschungsdateninfrastruktur (NFDI)**





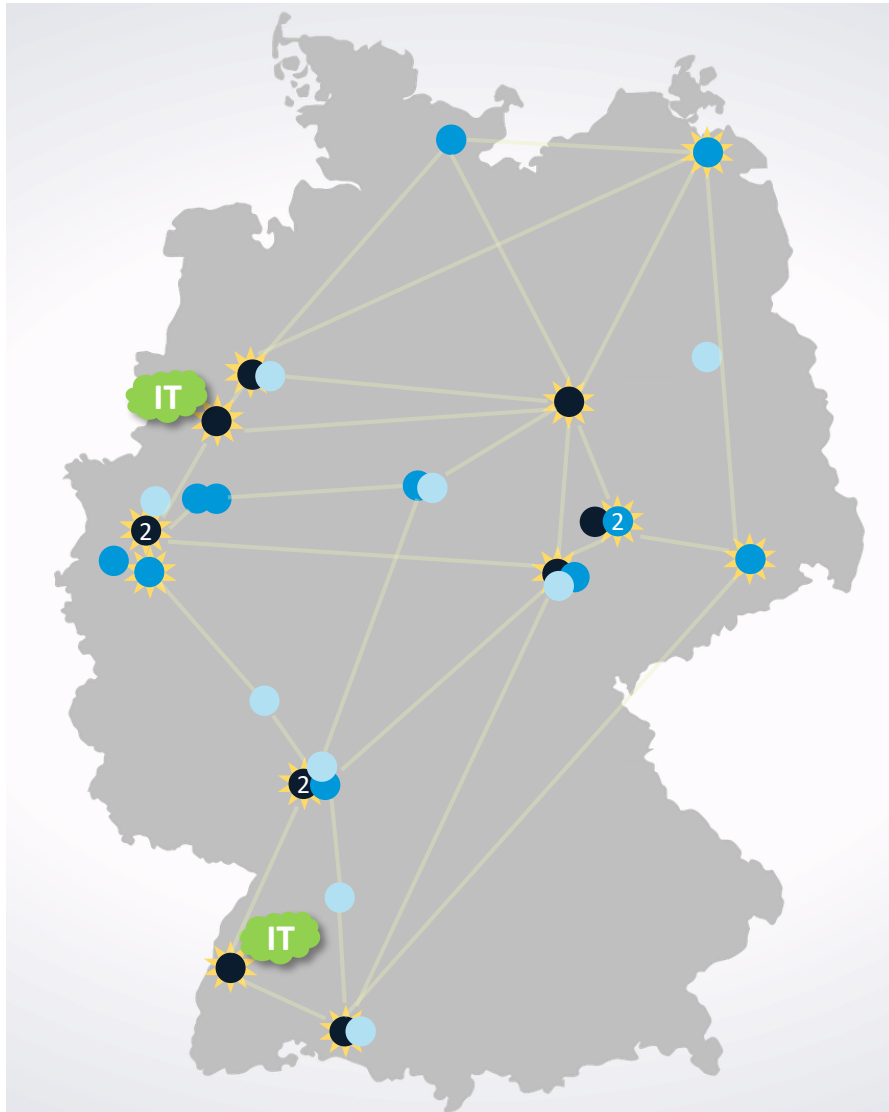
“The aim (...) is to systematically manage (...) research data, provide long-term data (...) accessibility, and network the data both nationally and internationally.”

<https://nfdi.de>



Adapted from: <https://www.dfg.de/foerderung/programme/nfdi/informationsmaterialien/index.html>, copyright: Deutsche Forschungsgemeinschaft

The consortium NFDI4BIOIMAGE



- Involved institutions all over Germany

- Lead: Heinrich Heine University Düsseldorf
Speaker: Prof. Dr. Stefanie Weidtkamp-Peters

 IT resources for the consortium

 **Data stewards (DaSts) & Research Software Engineers (RSE)**

 Community Use Cases

● Responsible Co-Applicant Institutions
● Participating Institutions

We offer a Help Desk to support you!

Research Data Management for Microscopy and BioImage Analysis

Startseite > Help Desk

The NFDI4BIOIMAGE Community Help Desk

Receive support from our Data Stewardship Team

The NFDI4BIOIMAGE Help Desk is the point of contact for researchers and research-associated staff working with bioimaging data to request for support with bioimaging data management. Our team of **Data Stewards** will review your request and contact you within a couple of days. We will discuss with you and our network which could be the optimal way to support you.

Help Desk support can mean supporting you in several ways:

- Pointing to relevant resources and solutions you might not have been aware of yet
- Helping to moderate help requests within the broader community (e.g., image.sc)
- Discussing and supporting a local solution for your case
- Integrating you as a **Community Use Case** partner within NFDI4BIOIMAGE's work.

Here, you can read more about the Help Desk Concept.
Use the form below to contact us! We are looking forward to your message.

NFDI4BIOIMAGE Help Desk

HELP REQUEST FORM

Your name*

Your email address

Kontakt

The NFDI4BIOIMAGE Data Stewardship Team is headed by:

Dr. Vanessa Fuchs
Heinrich-Heine University Düsseldorf

Dr. Maximilian Müller
University of Konstanz

For support requests, please use the Help Desk entry mask.
If you wish to establish a non-support request contact with the team, please reach out via email.

E-Mail schreiben

Goals and Task Areas +

Consortium members +

Project Office

Data Stewardship Team -

Mohsen Ahmadi

Vanessa Fuchs

Ksenia Krooß

Riccardo Massei

Maximilian Müller

Jens Wendt

Cornelia Wetzker

Partners

Section work

The NFDI4BIOIMAGE Data Stewardship Team

Get to know our task force for community support!

Our consortium offers support for bioimaging data management within the scientific community. No matter what your field of research is, we're here to support handling data acquired with microscopes. To do so, NFDI4BIOIMAGE has formed a Data Stewardship Team addressing your needs - from the community for the community.

To contact our Data Stewards Team for support in bioimaging data management, please use the Help Desk.

Who are the Data Stewards (DaSts)?

All NFDI4BIOIMAGE Data Stewards are scientists from different research fields with a focus on bioimaging. Within NFDI4BIOIMAGE, they work in different Task Areas and bundle their knowledge and expertise as the DaSt team. Thus, the DaSt team can offer community support for a wide range of cases. If required, our DaSt team will help finding additional contact partners (e.g., in other NFDI consortia). Here, we introduce the team:

THE DATA STEWARDS TEAM OF NFDI4BIOIMAGE

Mohsen Ahmadi
Background: Biochemistry & Microscopy
Affiliation: INP Greifswald

Vanessa Fuchs
Background: Plant Sciences
Affiliation: Heinrich Heine University Düsseldorf

Riccardo Massei
Background: Environmental Sciences and Toxicology
Affiliation: Helmholtz Center f. Env. Res. (UFZ), Leipzig

Maximilian Müller
Background: Ecotoxicology
Affiliation: University of Konstanz

Jens Wendt
Background: Electrical Eng./Information Tech. & Biomedical Eng.
Affiliation: University of Münster

Cornelia Wetzker
Background: Molecular Biology, Immunology, Zoology
Affiliation: Dresden Technical University

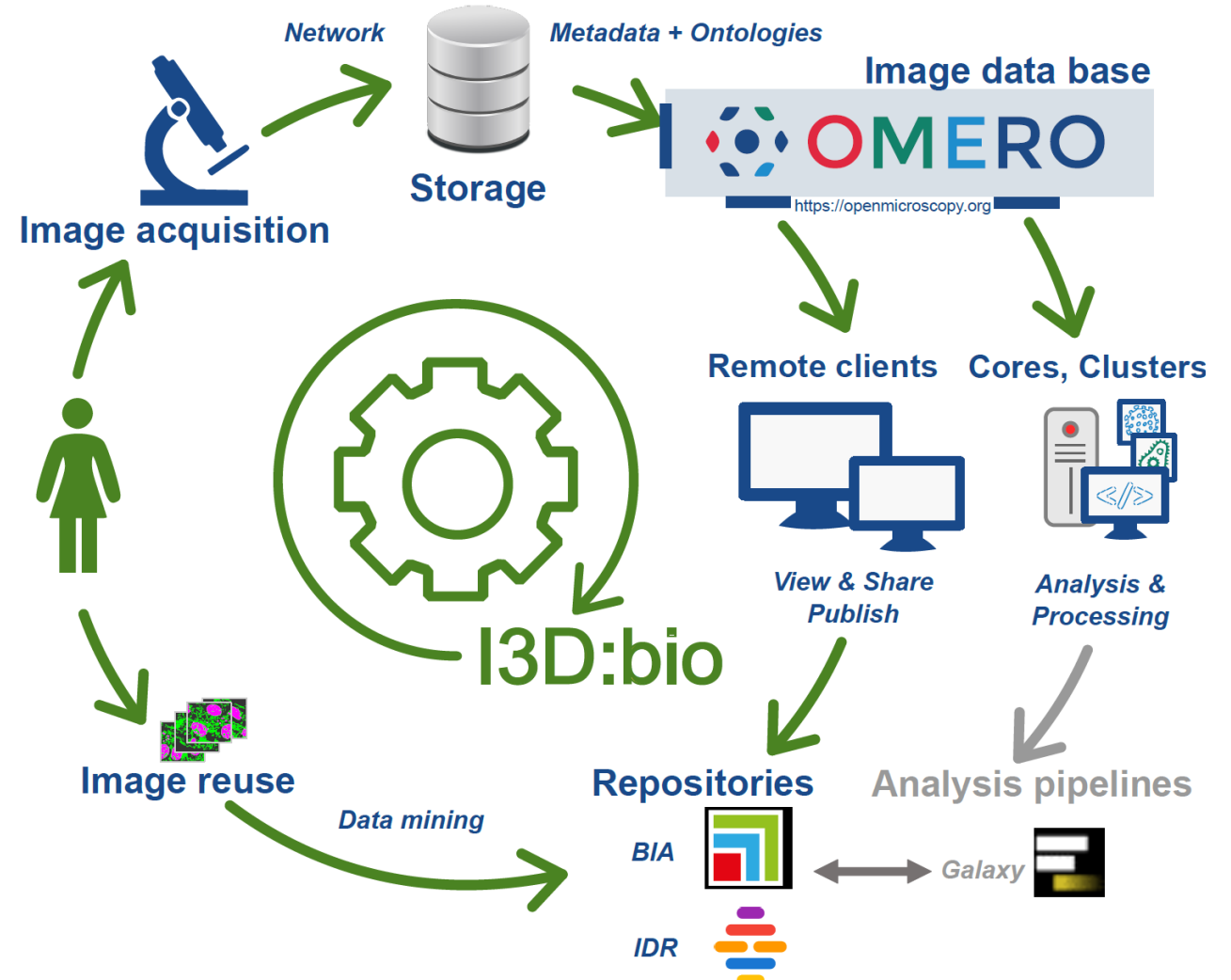
Ksenia Krooß
Background: Plant Sciences
Affiliation: Heinrich Heine University Düsseldorf

<https://nfdi4bioimage.de>

I3D:bio – OMERO-centered RDM project

Information Infrastructure for BioImage Data (I3D:bio)

- **DFG-funded:** 2022-2025
- **Partners:** HHU Düsseldorf, Uni Osnabrück, Uni Freiburg, DKFZ Heidelberg
- **Goal:**
 - Implement new OMERO instances in Germany
 - Provide training for OMERO users



The I3D:bio knowledge hub



I3D:bio – Information Infrastructure for BioImage Data

A Microscopy Research Data Management Resource

ABOUT I3D:bio ▾

About Bioimaging Data ▾

Teaching Material & Events ▾

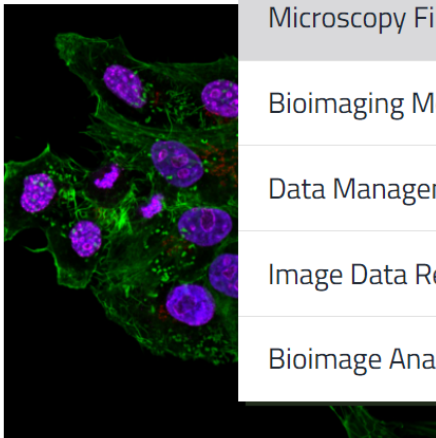
Guides & Resources ▾

Help Desk & Contact

Welcome to I3D:bio

Data Life Cycle

management resource for bioimaging with a focus on light microscopy



Microscopy File Formats

Bioimaging Metadata

Data Management Platforms

Image Data Repositories

Bioimage Analysis

resource to aid with...

and teaching about microscopy research data management?

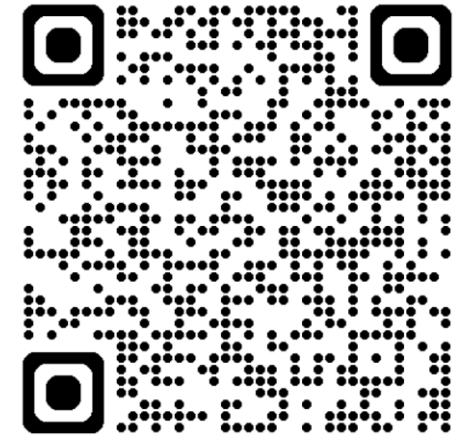
ing your microscopy research data?

sharing and archiving your images?

our image data with experimental protocols and analysis results?

up microscopy research data management at your institute or facility?

se I3D:bio project pages and contact us for questions and support.



<https://www.i3dbio.de>

What's next?

- OMERO as a tool for bioimaging data management

Tom Boissonnet



- Reproducible image analysis workflows with OMERO software APIs

Michele Bortolomeazzi



- Publishing datasets in public archives for bioimage data

Ksenia Krooß



- Meet us at the Research Data Management Booth throughout the day (+ *Lena Krämer*)



Acknowledgments

We thank the Gesellschaft für Biochemie und Molekularbiologie (GBM) for inviting us to this Symposium.

We thank all members, contributors, and supporters of NFDI4BIOIMAGE and I3D:bio

Polls for participation from the audience were created with slido.com



Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 462231789

Contact

Dr. Christian Schmidt
German Cancer Research Center (DKFZ), Heidelberg
Christian01.schmidt (AT) dkfz-heidelberg.de