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., Mallm, 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

Notes on the published version as of August 7th, 2024

This document contains the NFDI4BIOIMAGE grant application as submitted in the framework of the National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, NFDI) in 2021, with only minor modifications. The application was approved in November 2022, and work began in March 2023. However, the measures and milestones presented here reflect the work plan at the time of submission and may differ from current activities. As with all NFDI consortia, the approved budget was reduced by approximately 14% across all measures. Consequently, we do not expect all measures to be fulfilled during the five-year project phase as originally planned. Some plans have been adapted to account for novel developments and technical changes. In other cases, strategic decisions at the NFDI-wide level have influenced individual measures. The current work plan can be found on the project webpage at https://nfdi4bioimage.de.
Individuals named in the proposal denote those with specific roles. However, a much wider group of people were involved in the community-based process to establish the consortium. All contributors to the original text, regardless of whether they were explicitly mentioned, have been listed as co-authors.

Authors

Joshua Moore[1], Susanne Kunis[2], Björn Grüning[3], Markus Blank-Burian[4], Jan-Philipp Mallm[5], Torsten Stöter[6], Werner Zuschratter[6], Marc Thilo Figge[7], Anna Kreshuk[8], Christian Tischer[8], Robert Haase[9,§], Thomas Zobel[4], Pavol Bauer[6,$], Carl-Magnus Svensson[7], Ruman Gerst[7], Janina Hanne[1], Christian Schmidt[10,#] , Markus M. Becker[11], Thomas Bocklitz[12], Jan Bumberger[13], Claire Chalopin[14,&], Jianxu Chen[15], Paul Czodrowski[16,%], Timo Dickscheid[17], Carsten Fortmann-Grote[18], Jan Huisken[19], Jan Lohmann[20], Astrid Schauss[21], Martin Baumann[20], Carlo Beretta[20,*], Jean-Marie Burel[22], Vincent Heuveline[20], Rohini Kuner[20], Thomas Kuner[20], Matthias Landwehr[10], Andrea Leibfried[20], Roland Nitschke[3,23], Deepti Mittal[20,~], Dirk von Suchodoletz[3], Monica Valencia-Schneider[21], Peter Zentis[21], Dominik Brilhaus[24,25], Matthew Hartley[26], Bastian Hülsmann[16], Susanne Dunker[13], Antje Keppler[27] Aastha Mathur[27], Christian Meesters[16] Wiebke Möbius[28], Sven Nahnsen[29], Claudia Pfander[27], Stephanie Rehwald[30], Beatriz Serrano-Solano[27], Laura Vilardell Scholten[11], Raimund Vogl[4], Lutz Becks[10], *Elisa Ferrando-May*[5,10], *Stefanie Weidtkamp-Peters*[24]

shared senior authorship
corresponding authors

The list of authors is ordered as follows: Task Area Leaders (Responsible PIs at Co-Applicant Institutions) and their associates, Project management, Responsible PIs at Participant Institutions, members at Co-Applicant and Participant Institutions, Community Use Case partners, contributors and supporters, Deputy Spokesperson, Spokesperson.

Correspondence should be addressed to: Josh Moore (joshua.moore@gerbi-gmb.de), Elisa Ferrando-May (elisa.may@dkfz-heidelberg.de), or Stefanie Weidtkamp-Peters (stefanie.weidtkamp-peters@hhu.de)

Author Affiliations:

1 German BioImaging – Society for Microscopy and Image Analysis e.V. (GerBI-GMB), Konstanz
2 University of Osnabrück (UOS), Osnabrück
3 University of Freiburg (ALU-FR), Freiburg
4 University of Münster, Münster
5 German Cancer Research Center (DKFZ), Heidelberg
6 Leibniz Institute for Neurobiology (LIN), Magdeburg
7 Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute (HKI), Jena
8 European Molecular Biology Laboratory (EMBL), Heidelberg
9 Dresden University of Technology (TUD), Dresden
10 University of Konstanz (UKON), Konstanz
11 Leibniz Institute for Plasma Science and Technology (INP), Greifswald
12 Leibniz Institute of Photonic Technology (IPHT), Jena
13 Helmholtz Centre for Environmental Research (UFZ), Leipzig
14 University of Leipzig Medical Faculty, Innovation Center Computer Assisted Surgery, Leipzig
15 Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V., Dortmund
16 Johannes Gutenberg University of Mainz (JGU), Mainz
17 Forschungszentrum Jülich (FZJ), Jülich
18 Max Planck Institute for Evolutionary Biology (MPIEB), Plön
19 University of Göttingen (UGOE), Göttingen
20 Heidelberg University (UHD), Heidelberg
21 University of Cologne (UoC), Cologne
22 University of Dundee, Dundee, UK
23 Life Imaging Center, CIBSS and BIOSS – Centres for Biological Signalling Studies, University of Freiburg, Freiburg
24 Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf (HHU), Düsseldorf
25 Cluster of Excellence on Plant Sciences (CEPLAS), Düsseldorf
26 European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK
27 Euro-BioImaging ERIC Bio-Hub, at EMBL Heidelberg
28 Max-Planck-Institute for Multidisciplinary Sciences (MPINAT), Göttingen
29 Quantitative Biology Center (QBiC), University of Tübingen, Tübingen
30 University of Duisburg-Essen (UDUE), Duisburg and Essen

§ present address: University of Leipzig, Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig
$ present address: T-Systems International GmbH, Lübecker Str. 2, 39124 Magdeburg
# present address: German Cancer Research Center (DKFZ), Heidelberg
& present address: University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen (HAWK), Göttingen
% address at the time of submission: Dortmund Technical University (TUDo), Dortmund
* present address: Resolve Biosciences GmbH, Creative Campus Monheim, Monheim am Rhein
~ present address: Global Computational Biology & Digital Sciences, Boehringer Ingelheim Pharma GmbH & Co. KG

Table of Content

Bibliography and references:
In this proposal, bibliography entries are marked in brackets [#] (see 6.2 Bibliography), other references are marked with footnotes in uppercase format.

Note (2024): Where applicable, references published as preprints at the time of proposal submission were substituted with the corresponding peer-reviewed articles.

1 General Information

- **Name of the consortium in English and German**

  NFDI4BIOIMAGE – National research data infrastructure for microscopy and bioimage analysis

  NFDI4BIOIMAGE – Nationale Forschungsdateninfrastruktur für Mikroskopie und Bildanalyse

- **Summary of the proposal in English and German**

Bioimaging refers to a collection of methods to visualize the internal structures and mechanisms of living organisms. The fundamental tool, the microscope, has enabled seminal discoveries like that of the cell as the smallest unit of life, and continues to expand our understanding of biological processes. Today, we can follow the interaction of single molecules within nanoseconds in a living cell, and the development of complete small organisms like fish and flies over several days starting from the fertilized egg. Each image pixel encodes multiple spatiotemporal and spectral dimensions, compounding the massive volume and complexity of bioimage data. Proper handling of this data is indispensable for analysis and its lack has become a growing hindrance for the many disciplines of the life and biomedical sciences relying on bioimaging. No single domain has the expertise to tackle this bottleneck alone.

As a method-specific consortium, NFDI4BIOMAGE seeks to address these issues, enabling bioimaging data to be shared and re-used like they are acquired, i.e., independently of disciplinary boundaries. We will provide solutions for exploiting the full information content of bioimage data and enable new discoveries through sharing and re-analysis. Our RDM strategy is based on a robust needs analysis that derives not only from a community survey but also from over a decade of experience in German BioImaging, the German Society for Microscopy and Image Analysis. It considers the entire lifecycle of bioimaging data, from acquisition to archiving, including analysis and enabling re-use. A foundational element of this strategy is the definition of a common, cloud-compatible, and interoperable digital object that bundles binary images with their descriptive and provenance metadata. With members from plant biology to neuroscience, NFDI4BIOIMAGE will champion the standardization of bioimage data to create a framework that answers discipline-specific needs while ensuring communication and interoperability with data types and RDM systems across domains. Integration of bioimage data with, e.g., omics data as the basis for spatial omics, holds great promise for fields such as cancer medicine. Unlocking the full potential of bioimage data will rely on the development and broad availability of exceptional analysis tools and training sets. NFDI4BIOIMAGE will make these accessible and usable including cutting-edge AI-based methods in scalable cloud environments. NFDI4BIOIMAGE intersects with multiple NFDI consortia, most prominently with GHGA for linking image and genomics data and with DataPLANT on the definition of FAIR data objects. Last but not least, NFDI4BIOIMAGE is

internationally well connected and represents the opportunity for German scientists to keep path with and have a voice in several international initiatives focusing on the FAIRification of bioimage data as one of the main challenges for the advancement of knowledge in the life and biomedical sciences.

Bioimaging, oder auch biologische Bildgebung, bezeichnet die Gesamtheit der Verfahren, die der Visualisierung der Strukturen und Funktionsmechanismen lebender Organismen dienen. Dem grundlegenden Werkzeug des Bioimaging, dem Mikroskop, verdanken wir wegweisende Entdeckungen, wie die der Zelle als dem kleinsten Baustein des Lebens. Heute können wir verfolgen, wie einzelne Moleküle auf einer Zeitskala von Nanosekunden in einer lebenden Zelle miteinander wechselwirken, oder wie ganze Modellorganismen wie Fische oder Fliegen sich über mehrere Tage entwickeln. Dabei ist in jedem Pixel Information über mehrere Raum-, Zeit- und spektrale Dimensionen enthalten. Angesichts des massiven Volumens und der Komplexität, ist ein adäquates Datenmanagement für jede Art der Analyse von Bioimaging-Daten unerlässlich, allerdings bisher kaum realisiert. Für viele Disziplinen der Lebens- und Medizinwissenschaften, die auf Bioimaging angewiesen sind, ist dies ein substantielles Hindernis. Keine dieser Disziplinen alleine verfügt über die Expertise, diese Lücke zu schließen.

NFDI4BIOIMAGE setzt als methodenzentriertes Konsortium an dieser Stelle an, und will Lösungen erarbeiten, damit Bioimaging-Daten so geteilt und wiederverwendet, wie sie akquiriert werden, über disziplinäre Grenzen hinaus. Der volle Informationsgehalt dieser Daten soll ausgeschöpft, und aus der Re-Analyse neue Erkenntnisse gewonnen werden. Unsere FDM-Strategie basiert auf einer robusten Bedarfsanalyse, die nicht nur auf eine Community-weite Umfrage, sondern auch auf über zehn Jahren Erfahrung aus dem Netzwerk German BioImaging, der heutigen Deutschen Gesellschaft für Mikroskopie und Bildanalyse fußt. Ein Grundbaustein dieser Strategie ist die Definition eines gemeinsamen, Cloud-kompatiblen und interoperablen digitalen Objektes als Dateneinheit aus binären Bilddaten und zugehörigen Metadaten. NFDI4BIOIMAGE will eine Infrastruktur bereitstellen, die einerseits disziplinspezifischen Anforderungen gerecht wird, andererseits an andere Datentypen und FDM-Systeme verschiedener Domänen der Wissenschaft anschließt. Die Integration von Bild- und Omics-Daten ist hierfür ein vielversprechendes Beispiel mit wachsender Bedeutung für die Krebsmedizin. Damit solche Ansätze ihr Potential voll entfalten, brauchen wir leistungsfähige Bildanalyse Werkzeuge sowie Trainingsdatensätze zu deren Optimierung. NFDI4BIOIMAGE wird beides in einer skalierbaren Cloud-Umgebung zur Verfügung stellen, inklusive aktuellster KI-basierter Methoden. NFDI4BIOIMAGE hat neben anderen insbesondere Berührungspunkte mit mit GHGA zur Integration von Genom- und Bilddaten und mit DataPLANT zur Definition von *FAIR Data Objects*. Die ebenfalls bestehende starke internationale Vernetzung von NFDI4BIOIMAGE stellt eine Chance für deutsche WissenschaftlerInnen dar, an internationalen Initiativen mitzuwirken, die die FAIRifizierung von Bioimaging-Daten als eine Hauptherausforderung der Medizin- und Lebenswissenschaften erkannt haben.

Applicant institution

Applicant institution	Location
Heinrich Heine University (HHU) Düsseldorf *Head:* Prof. Dr. Anja Steinbeck	Universitätsstraße 1 40225 Düsseldorf

- **Spokesperson**

Spokesperson	Institution, location
Prof. Dr. Stefanie Weidtkamp-Peters	Heinrich Heine University, Düsseldorf

- **Co-applicant institutions**

Co-applicant institutions	Location
German BioImaging – Society for Microscopy and Image Analysis (GerBI-GMB) *Head*: Prof. Dr. Stefanie Weidtkamp-Peters	c/o University of Konstanz 78547 Konstanz
University of Konstanz (UKON) *Head:* Prof. Dr. Katharina Holzinger	Universitätsstraße 10 78457 Konstanz
German Cancer Research Center (DKFZ) *Head:* Prof. Dr. Michael Baumann	Im Neuenheimer Feld 280 69120 Heidelberg
European Molecular Biology Laboratory (EMBL) *Head*: Prof. Edith Heard, Ph.D.	Meyerhofstr. 1 69117 Heidelberg
Osnabrück University (UOS) *Head*: Prof. Dr. Susanne Menzel-Riedl	Neuer Graben 29 / Schloss 49074 Osnabrück
Leibniz Institute for Neurobiology (LIN) *Head*: Prof. Dr. Stefan Remy	Brenneckerstr. 6 39118 Magdeburg
Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute (HKI) *Head*: Prof. Dr. Axel A. Brakhage	Beutenbergstr. 11a 07745 Jena
Leipzig University (UL) *Head: Prof. Dr. Eva Inés Obergfell*	*Ritterstr. 26* *04109 Leipzig*
University of Münster *Head*: Prof. Dr. Johannes Wessels	Schlossplatz 2 48149 Münster
University of Freiburg (ALU-FR) *Head*: Prof. Dr. Kerstin Krieglstein	Friedrichstraße 39 79098 Freiburg

- **Co-spokespersons**

Co-spokespersons	Institution, location	Task area(s)
Prof. Dr. Stefanie Weidtkamp-Peters	German BioImaging – Society for Microscopy and Image Analysis e. V. (GerBI-GMB)	**TA1**, **TA6**
Prof. Dr. Elisa May	University of Konstanz (UKON), Konstanz & German Cancer Research Center (DKFZ), Heidelberg	**TA6**, TA5

Dr. Jan-Philipp Mallm	German Cancer Research Center (DKFZ), Heidelberg	**TA3**
Dr. Anna Kreshuk	European Molecular Biology Laboratory (EMBL), Heidelberg	**TA4**
Dr. Susanne Kunis	Osnabrück University (UOS), Osnabrück	**TA1**
Dr. Werner Zuschratter	Leibniz Institute for Neurobiology (LIN), Magdeburg	**TA3**
Prof. Dr. Marc Thilo Figge	Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute (HKI), Jena	**TA4**
Dr. Robert Haase	Leipzig University (UL), *(before: Dresden University of Technology*)	**TA5**, TA4
Dr. Thomas Zobel	University of Münster	**TA5**
Dr. Markus Blank-Burian	University of Münster	**TA2**
Dr. Björn Grüning	Albert-Ludwigs-University of Freiburg (ALU-FR)	**TA2**

Note: As of January 2024, co-spokesperson Robert Haase moved from Dresden University of Technology (TUD) to Leipzig University (UL). Accordingly, UL joined the consortium as a co-applicant institution. TUD changed its status to a Participating Institution.

- **Participants**

Participating institutions	Location
Heidelberg University (UHD)	Heidelberg
University of Cologne (UoC)	Köln
Johannes Gutenberg University of Mainz (JGU)	*Mainz*
Georg-August-University of Göttingen (UGOE)	Göttingen
Leibniz Institute for Plasma Science and Technology (INP)	Greifswald
Helmholtz-Center for Environmental Research (UFZ)	Leipzig
Leibniz Institute of Photonic Technology (IPHT)	Jena
Leibniz Institute for Analytical Sciences (ISAS)	Dortmund
Dresden University of Technology (TUD)	Dresden
University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen (HAWK)	*Göttingen*
Forschungszentrum Jülich (FZJ)	Jülich
Max Planck Institute for Evolutionary Biology (MPIEB)	Plön

Note: The persons responsible for the project parts at the Technical University of Dortmund (TUDo, P. Czodrowski) and at the University of Leipzig Medical Faculty (UL-MED, C. Chalopin) each accepted a call to another institution before the start of the funding period. To enable optimal integration and collaboration on the goals of NFDI4BIOIMAGE, both TUDo and UL-MED kindly handed over the role of a Participating Institution to the respective new affiliated institutions of the responsible scientists. Accordingly, the JGU (P. Czodrowski) and the HAWK (C. Chalopin) joined the project as Participating Institutions as of March 2023.

Contribution of participating institutions:

At **Heidelberg University (UHD)**, a large bioimaging community exists across various disciplines (cell biology, neurosciences, stem cell biology, virology and more), including dedicated bioimaging research groups. The “Molecular foundations of life, health, and disease” constitutes one of four “Fields of Focus” as part of UHD’s Excellence Strategy (Research Council FoF1: J. Lohmann), representing the Heidelberg life sciences across faculties and institutions, in which bioimaging plays a key role. Three light microscopy core facilities on the Heidelberg campus offer user support and training. UHD has established the Competence Center for Research Data as a joint support infrastructure of the University Library (UB) and the University Computing Center (director of URZ: V. Heuveline, CIO at UHD). URZ provides various high-quality RDM tools and services such as a fully backed-up data storage (SDS@hd), a high-performance compute cluster (bwForCluster MLS&WISO) and cloud computing services (heiCLOUD). As a project partner in bw2FDM (with UKON), UHD regularly organizes the E-Science-Tage, a well-established German exchange platform on RDM-related topics. Collaboratively, URZ, UB, and CRC 1158 are developing a coordinated approach to bioimaging RDM to enable the use of the bioimage RDM platform OMERO (OME Remote Objects), computational services like Galaxy[1], and integration of these systems with electronic lab notebooks (elabFTW, a central service of URZ), topics in the focus of the Task Area (TA) 2 of NFDI4BIOIMAGE. The Nikon Imaging Center and the Electron Microscopy Core Facility will act as community connectors to make these services available for the benefit of >200 users p.a and beyond.

University of Cologne (UoC) is the second largest German university with a broad research profile covering all areas of science. The Imaging Facility of the Cluster of Excellence (EXC) on Cellular Stress Response in Aging-Associated Diseases, CECAD, with its director A. Schauss and bioimage analyst P. Zentis, has strong expertise in light and electron microscopy and image data management and will contribute to the task area on multimodal data integration (TA3). It collaborates with the Regional Computing Centre of UoC (M. Valencia-Schneider). Moreover, UoC via the Cluster of Excellence in Plant Sciences, CEPLAS (a joint initiative of UoC, HHU, FZJ and the MPI for Plant Breeding Research), and the NFDI consortium DataPLANT, will contribute to the definition of metadata formats and annotation tools (TA1). The involved CEPLAS data steward is D. Brilhaus (HHU). With three collaborative research centers (CRC1218, CRC1403, CRC1451) using bioimaging technologies, UoC is a strong partner for disseminating the results of our consortium (TA5).

At the **Johannes Gutenberg University of Mainz (JGU)**, the group of P. Czodrowski (formerly: Technical University of Dortmund) will collaborate with NFDI4BIOIMAGE on the exploitation of cell-painting assays for drug design. Here, AI-based methods are used on benchmark datasets to predict cell phenotypes for novel compounds of known structure, or to model novel structures based on a desired phenotype (TA4). The required metadata standards and annotation workflows to generate public FAIR HCS cell painting datasets will be elaborated in collaboration with TA1 and the planned DeBioData consortium.

The **University of Göttingen (UGOE)** is part of a research nucleus in Lower Saxony created by the partnership of the University (incl. the medical faculty) and non-university research institutes forming the Göttingen Campus. Among the shared activities have been two bioimaging-centered Clusters of Excellence[2] and the Göttingen eResearch Alliance. J. Huisken is a newly appointed

---

[1] https://usegalaxy.eu/

[2] Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB) & Multiscale Bioimaging: from Molecular Machines to Networks of Excitable Cells (MBEXC)

Alexander-von-Humboldt-Professor for Multiscale Biology at UGOE. From 2016 to 2021 he was Director of Medical Engineering at the Morgridge Institute for Research at Madison, Wisconsin, USA. J. Huisken is a pioneer of light-sheet microscopy technology and has recently established the Flamingo project[3] of shareable microscopes as an innovative approach to provide a large community of researchers with access to high-end advanced imaging [1]. With J. Huisken's return to Germany, UGOE will be home to the Flamingo project in Germany and the wider European area, providing the unique chance to accompany standardization for clearing protocols, metadata annotations and (remote) image data processing in close alignment with the consortium's objectives. The goal is to facilitate highly reliable image data generation with portable, modular, and customizable microscopes and potentially remote image acquisition. The Huisken lab develops tools to benchmark local infrastructures for this purpose, and collaborates on cloud-native data standards and data sharing solutions (TA1). The Flamingo project furthermore acts as a test platform to implement location-independent storage and sharing solutions and contributes to the dissemination of the consortium's standards and guidelines (TA5, TA6).

The **Leibniz Institute for Plasma Science and Technology (INP)** is the largest non-university institute in the field of cold plasmas in Europe and has a leading role in defining RDM standards in plasma research, an interdisciplinary field making use of a multitude of methods, including imaging of biological and non-living material. The BMBF-funded project QPTDat (Quality Assurance and Linking of Research Data in Plasma Technology) coordinated by M. Becker at INP, has developed important milestones of plasma science RDM, e.g., the definition of a standardized metadata schema (Plasma-MDS [2]) or the implementation of a plasma technology knowledge graph (that shall be further developed in NFDI4Phys, where INP is co-applicant). INP developed, hosts and maintains the Interdisciplinary Plasma Technology Data Platform (INPTDAT)[4]. INP is part of the Leibniz Research Alliance Health Technologies together with HKI, IPHT, ISAS, and LIN. In NFDI4BIOIMAGE, it will collaborate on the interoperability of image and plasma metadata schemes and ontologies (TA1, TA3), and will collect RDM needs of the plasma science community (TA5).

The **Helmholtz-Center for Environmental Research (UFZ)** has extensive experience in the use of bioimaging methods for integrative environmental research, and contributes a central use case for managing data from high throughput / high content screening (HCS) and imaging flow cytometry (S. Dunker). For large-scale image data annotation, curation and analysis, automation steps are vital. UFZ will exploit the automation tools developed in NFDI4BIOIMAGE (TA3 and TA4) which in turn take advantage of the hardware specifications for analysis on HPC clusters developed in collaboration with TA2. UFZ shares experience on the aforementioned data types and with the implementation of novel workflows. Recommendations for metadata specifications for HCS workflows are contributed to TA1. The RDM Expert Board and Service Team (Head:

---

[3] https://involv3d.org

[4] Becker M. et al. (2019): INPTDAT – a new data platform for plasma technology, doi: 10.5281/zenodo.3500282.

J. Bumberger) will share expertise with the consortium, in particular based on common interests in use cases related to high-content imaging at JGU and at UKON. Moreover, UFZ is a participant in four funded NFDI consortia. Hence, it is in an ideal position to contribute HCS-related data stewardship support to NFDI4BIOIMAGE and collaborating consortia (TA5).

The **Leibniz Institute of Photonic Technology (IPHT)** develops biophotonics methods for application in the biological and medical sciences. IPHT is part of the Leibniz Research Alliance Health Technologies. The Leibniz Center for Photonics in Infection Research (LPI)[5] is an infrastructure coordinated by IPHT and funded through the National Roadmap for Research Infrastructures that enables the targeted cross-disciplinary development of novel optical and photonics approaches for the diagnosis and therapy of infectious diseases. T. Bocklitz, head of the Photonic Data Science group at IPHT, leads the section “AI for diagnostics and therapy” that is responsible for establishing LPI’s data infrastructure. IPHT is well-embedded in the EXC “Balance of the Microverse” in Jena, which includes the microscopy core facility Microverse Imaging Center. T. Bocklitz coordinates interdisciplinary bioimaging RDM efforts in the Jena area. As a participant in NFDI4Chem, he develops a spectroscopy database. In NFDI4BIOIMAGE, IPHT contributes by building an overarching metadata database for data linking (TA1, TA3). Furthermore, the linkage and interoperability of microscopy and photonics data types, as well as the topic of electronic lab notebook (ELN)-cross-connectivity (e.g., bridging bioimage data management platforms to the NFDI4Chem-specific ELN “Chemotion”) are interests of IPHT (TA3).

The **Leibniz Institute for Analytical Sciences (ISAS)** combines expertise from biology, chemistry, physics and computer science for developing analytical technologies. ISAS is part of the Leibniz Research Alliance Health Technologies and has recently invested into opening a new project group on Analysis of Microscopic Biomedical Images (AMBIOM), in particular for large scale 3D image datasets as those generated, e.g., by light-sheet fluorescence microscopy. The group leader, J. Chen, is a former member of the Allen Institute for Cell Science and has extensive experience in this field [3, 4]. ISAS contributes to NFDI4BIOIMAGE with expertise on AI-based image analysis and segmentation tools, and will collaborate on software interoperability and the collection of benchmarking datasets (TA4). This resource will be available via the NFDI to foster AI-tool development while promoting FAIR data standards.

At **HAWK University of Applied Sciences and Arts Hildesheim/Holzminden/Göttingen** the research group of C. Chalopin (formerly Leipzig University Medical Faculty) exploits non-invasive imaging methods for guiding surgery and is interested in the further development of hyperspectral imaging (HSI) for the characterization of tissue structures and anomalies. In this emerging field of non-standard imaging, file and metadata standards are still largely undefined. In NFDI4BIOIMAGE, HAWK will collaborate predominantly with TA1, but also with TA4 to co-develop and test next-generation file formats (NGFF) as suitable multimodal vector-based image

---

[5] https://lpi-jena.de/en

file type for HSI or further imaging methods providing 3D and 4D vector-based image data. Appropriate metadata vocabularies and ontologies will be developed and linked to NGFF. HAWK will explore bridges to the medical image standards DICOM and collaborate with NFDI4Health with regard to patient data confidentiality and security (TA6).

The **Forschungszentrum Jülich (FZJ)** has a focus in basic research inspired by uses in the areas of energy research, bioeconomy and information science. Regarding the latter, the Institute of Neuroscience and Medicine (INM1) studies how information is processed by the human brain, for which bioimaging is among the key technologies used. The imaging core facility holds international reputation in relevant techniques like high-throughput microscopic brain imaging and histological processing of whole human brain samples. FZJ leads the H2020 flagship "Human Brain Project", which establishes the EBRAINS European infrastructure for brain research. T. Dickscheid is head of the research group "Big Data Analytics", which focuses on bioimaging RDM and bioimage analysis in the area of large-scale whole-slide imaging of human brain sections. His lab facilitates distributed workflow management in collaboration with the high-performance computing facility "Jülich Supercomuting Center (JSC)". In NFDI4BIOIMAGE, T. Dickscheid shares his expertise on the Brain Imaging Data Structure as a contributor to TA3. He maintains a close cooperation with the developers of datalad software[6] as well as with P. Ritter (Charité Berlin; spokesperson of the NFDI-Neuro initiative) on brain atlases in reproducible workflows, and as such provides a strong link to the NFDI-Neuro consortium. Moreover, FZJ is involved in a collaborative effort within CEPLAS (together with UoC, HHU) for data integration using the bioimage management platform OMERO.

The **Max Planck Institute for Evolutionary Biology (MPIEB)** has a three-tiered research focus on Evolutionary Genetics, Evolutionary Theory and Microbial Population Biology, for which the combination of various technologies and methods including microscopy plays an important role. The Scientific Computing Unit (Head: C. Fortmann-Grote) not only supplies IT-support for local research projects but develops software solutions for general purposes in an open, integrative manner. Delivered tools are explained in the unit's own SciComp Blog[7] and released for re-use and adaptation via GitHub[8]. Running an OMERO instance at MPIEB, the SciComp Unit has contributed to Bulk Map Annotations for Image Data and will contribute to NFDI4BIOIMAGE expertise in metadata ontologies, annotations both at the domain-specific and at the generic level (TA1), collected also in collaboration with the wider Max-Planck Bioimaging Network. MPIEB will work on the integration of these standards with the data management platform SEEK[9] and the ELN & LIMS instance openBIS[10] (TA3). For knowledge extraction from diverse data sources,

---

[6] https://www.datalad.org/

[7] https://mpievolbio-scicomp.pages.gwdg.de/blog/

[8] https://github.com/mpievolbio-scicomp/obat/

[9] https://seek4science.org

[10] https://openbis.ch

MPIEB is establishing a prototype knowledge base for the model organism *P. fluorescens SBW25* using semantic web technology including imaging data sources (TA3).

Participating individuals	Institution, location
Dr. Jean-Marie Burel, Open Microscopy Environment Consortium	University of Dundee, UK

Contribution of J.M. Burel (University of Dundee)

The Open Microscopy Environment (OME) is a consortium of open-source contributors building research infrastructure for bioimaging data. A number of open-source projects which are developed and maintained by OME are critical to this application. Bio-Formats is the premier translation library in the community for accessing proprietary file formats. OMERO is a leading RDM platform used by many participants. As a member of the project management team for OME, J.M. Burel is aware of the NFDI4BIOIMAGE roadmap and will help to coordinate across the growing number of development groups so that new software products and meta(data) specifications work together optimally.

- **Names and numbers of the DFG review boards (DFG-Fachkollegien) that reflect the subject orientation of the proposed consortium**
  - 201 Basic Research in Biology and Medicine
  - 202 Plant Sciences
  - 203 Zoology
  - 204 Microbiology, Virology & Immunology
  - 206 Neurosciences
  - 308 Optics, Quantum Optics and Physics of Atoms, Molecules and Plasmas
  - 409 Computer Science

2 Scope and Objectives

2.1 Research domains or research methods addressed by the consortium, specific aim(s)

Scientific techniques that enable visualizing the internal structures and dynamics of biological specimens have had an enormous impact on the progress of biology and medicine, as epitomized by the discovery of the cell in the mid-seventeenth century, made possible by the invention of the microscope. Today, bioimaging is a pervasive and indispensable methodological approach in the life and biomedical sciences, across all disciplines – from botany to biochemistry, from virology to material science. Recent technological advancements have endowed us with the ability to image biological samples at the level of the single molecules up to that of the complete organism (Figure 1). This breadth of range applies also to the temporal dimension and the visualized molecular features, making modern bioimaging a highly complex multi-dimensional and quantitative methodology.

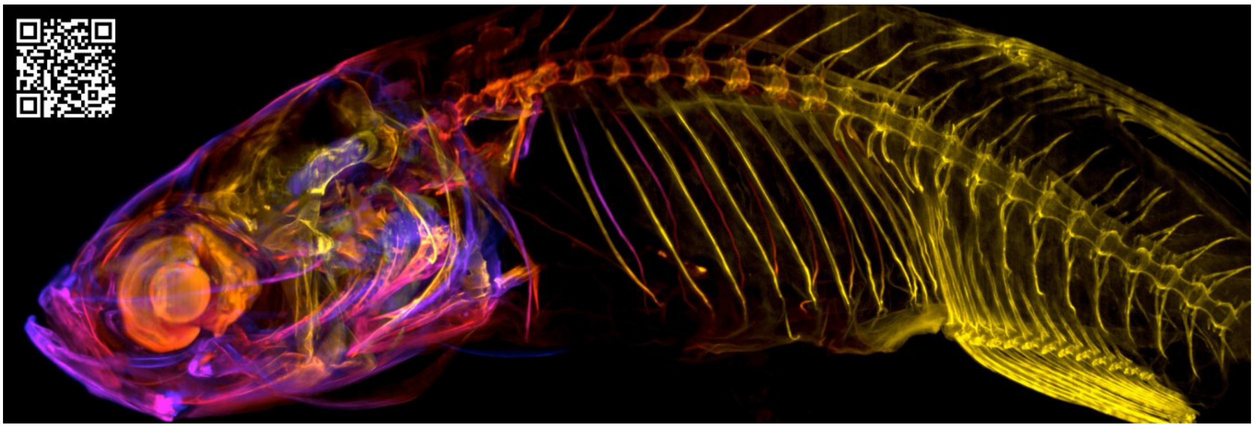

Figure 1: Example image of a biological specimen. An optically cleared, one month old zebrafish was imaged in a light-sheet microscope by acquisition in three tiles generating a full image of 8905x3000 pixels in 75 slices (at 20 µm axial res.), shown as maximum intensity projection in depth color coding. While this single projection of a fixed sample reaches almost 4 GB in size, a time series of a living specimen, e.g., for the study of developmental processes, generates TB-sized data (Courtesy of J. Huisken and K. Weiss).
The QR-code links to the image in an OMERO database.

Bioimaging experiments, e.g., using a light-sheet microscope, easily deliver several TB of data per examined specimen. Thus, bioimaging is a truly, even if not yet fully appreciated, big data discipline, and faces the same, if not harder, challenges than the classical big data producing omics technologies of the life sciences. Bioimaging confronts the scientific community with a substantial and not in the least resolved bottleneck that restrains this method from exerting its full potential for scientific discoveries. The feat will consist in finding ways to properly deal with the huge amount of complex and unstructured data, so that researchers, for whom access to high-end imaging equipment has become relatively easy, are not overwhelmed and discard data. The aim is to give them viable solutions at hand for storing, processing, analyzing, and, first and foremost, sharing bioimaging data. This aim includes educating the research community to adopt common standards as the absolute prerequisite for productive data re-use. Our consortium seeks to join the NFDI to tackle these unresolved issues. We are encouraged by the feedback from existing and planned NFDI consortia in the life and health sciences, who have highlighted (bio)imaging data as very relevant for their communities, but lack the capacity and/or the expertise to address this data type in a comprehensive manner.

Bioimaging data encompasses all data generated by the diverse modalities of electron and light microscopy, including spectroscopic and multi-parametric, correlative approaches and applied to biological or biomimetic specimens. NFDI4BIOIMAGE does not address medical or clinical imaging, where data protection issues come prominently into play and need to be dealt with specifically by the communities of interest as represented, e.g., in NFDI4Health and/or NFDI4Patho. NFDI4BIOIMAGE will strive to exploit and expand existing bridges between the bioimaging and the medical imaging domain with the vision of facilitating the use of biological and pre-clinical imaging data for diagnostic and therapeutic purposes. Taken together, the scope of our consortium is centered around one data type, bioimaging data, as defined above, and encompasses the whole life-cycle of this data, from acquisition to archiving, including data analysis. Thus, the scope of NFDI4BIOIMAGE has a disciplinary aspect too, localized in the domain of bioimage informatics, in particular in the investigation of novel AI-based methods for bioimage analysis.

Realizing the aims of NFDI4BIOIMAGE requires overcoming several barriers. First, the stigma associated with bioimaging data as cumbersome to handle, share, and archive, must be addressed by delivering practicable data management solutions. Secondly, we must avoid bioimaging data being stuck in silos and data graveyards because the local environment where they are acquired lacks the proper technical expertise and infrastructure. Here, best practices as well as raising awareness of funders about the relevance and requirements of bioimaging data will help. Thirdly, we must minimize waste of resources by parallel developments and isolated solutions, a trend that will be effectively counteracted by the networking and harmonization efforts of NFDI4BIOIMAGE. Regarding the latter, we have seen several activities materializing in the last few years trying to tackle the bioimage data challenge, the most frequent being the deployments of local instances of the OMERO platform[11]. The aim is to bring together these relatively disconnected efforts, and distribute the expertise systematically and comprehensively while driving further the development of bioimage data management solutions that go beyond any single platform and are ready for future demands. Closing the gap between local, on-premise and cloud-based services is certainly one such demand. This will include the development of new data formats. Importantly, this harmonization and distribution endeavour must take into account the specific needs of the individual disciplines where solutions are to be adopted. Ideally, NFDI4BIOIMAGE will become an expert hub for adapting bioimage RDM solutions to the requirements of the majority of life science disciplines and by doing so will have the leverage to push standardization of formats and procedures. The vision is to increase the bioimage RDM competency in the community to a level that enables a shift in perspective: instead of being an annoying and cumbersome task that scientists attend to when projects are already at an advanced stage, RDM is embraced at the very beginning of the bioimage data lifecycle, i.e., as part of the experimental planning before image acquisition. The growing body of available FAIRified bioimage data will lead researchers to re-evaluate and potentially re-use existing data to answer new questions shortening the path to new discoveries. Once initiated by the bioimaging community, the trend to FAIR bioimage data management is likely to become an integral part of everyday good scientific practice in any discipline. On an even larger scale, this process will pave the way for the wider adoption of data mining approaches (consider, e.g., the ubiquity of BLAST searches!) making bioimage data accessible to the next level of knowledge extraction. From the recognition of the true value of FAIR bioimage data, new research approaches will arise, similarly to the field of molecular dynamics that was sparked and enabled by the introduction of the Protein Data Bank [5]. We predict that the application of AI-based methods on huge integrated datasets that include bioimage data will have a transformative impact on the life and medical sciences in the decades to come.

2.2 Objectives and measuring success

From the description above, one can directly derive the following four objectives as the most important ones on the road to FAIR bioimage data management.

1. **Champion the standardization of the “bioimage data” type:** A common unified standard in bioimaging, capturing large multidimensional bioimaging measurements and its related metadata, is needed to overcome the stigma of cumbersomeness and enable interoperability between commercial vendors, publishers, and the research community. This process, that has been initiated by our bioimaging community already, will lead to the development of

---

[11] https://openmicroscopy.org/omero

interoperable vocabularies for the description of the wide variety of bioimaging modalities and applications. A direct measure of success of this objective will be the number of German datasets deposited to international archives and repositories. Task Area 1 will lead this objective.

2. **Provide scalable infrastructure for FAIR image data**: Rather than hosting a single national repository for bioimaging data, we aim to develop a roadmap for federating and sharing distributed storage resources including linking to international repositories. A digital testing environment, a “sandbox”, for management tools emerging from the consortium will be provided by Task Area 2. Here, the documentation of access to this environment is an obvious measure for progress made in this objective. The value of bioimaging data will increase with their interoperability, e.g., if bioimaging data can be efficiently linked to data types like omics data. This work towards data integration, led by Task Area 3, depends on discipline-specific input. The total amount of linked data allows us to measure success in this objective.
3. **Maximize the reach of reproducible image analysis workflows in the community**: Improvements in bioimage data management will make it much easier to bring (AI-based) methods for bioimage analysis to the hand of the end-user and promote the reproducibility and sustainability of image analysis workflows. Task Area 4 will focus on this objective with technical support from Task Area 2 and support from Task Area 5 to reach out to the community. Counting the number of image analysis jobs submitted to central tools that will be provided by us will show the success of this objective.
4. **Capacitate researchers for FAIR image data management**: In preparation of this proposal, we have conducted a survey[12] and invited national and international researchers of all levels to share with us their experience on image data management. This (and other surveys) confirm that knowledge about and access to image data management and analysis tools is currently scarce in the bioimaging community. Empowering the community to grow its RDM literacy by active contribution is the main goal of Task Area 5. We will monitor the success of our measures by repeating the survey annually. We will adjust our efforts according to the results.

The composition of our consortium, a clear governance structure with a dedicated Task Area for coordination and networking (TA6) and a focused work programme will help us to achieve these objectives.

---

[12] https://nfdi4bioimage.de/en/survey/

3 Consortium

These members of NFDI4BIOIMAGE are also participating in other consortia *or consortia initiatives:*

Participating institution (or a person from it)	**is also a participating in**
HHU	DataPLANT, NFDI4Microbiota
DKFZ	GHGA, NFDI4Health, NFDI4Immuno
EMBL	GHGA, NFDI4Microbiota, DAPHNE4NFDI
LIN	NFDI-Neuro
TUD	GHGA, NFDI4Chem, NFDI4DS, NFDI4Earth, NFDI4Ing, PUNCH4NFDI, Text+, NFDIxCS, FAIRmat, DAPHNE4NFDI
WWU	PUNCH4NFDI, MaRDI, NFDI4Earth, NFDI4Memory, NFDI4Objects, TheoRes
HKI	NFDI4Microbiota
ALU-FR	DataPLANT, NFDI-MatWerk, PUNCH4NFDI, Text+, NFDI-Neuro, NFDI4Memory, DAPHNE4NFDI
UKON	DataPLANT, METHODS
FZJ	DAPHNE4NFDI, NFDI4Earth, NFDI4Ing, NFDI4Microbiota, NFDI-MatWerk, PUNCH4NFDI, Text+, NFDI-Neuro, DataPLANT, FAIRmat, FAIRagro, NFDI4Phys, NFDIxCS
UoC	GHGA, NFDI4Culture, NFDI4Health, NFDI4Chem, NFDI4Culture, NFDI4Earth, NFDI4Microbiota, PUNCH4NFDI, Text+, DAPHNE4NFDI, BERD@NFDI, Text+, METHODS, NFDI4Memory, NFDI4Mobility, NFDI4Objects, NFDI4Phys
UFZ	NFDI4Microbiota, NFDI4Earth, NFDI4Biodiversity, NFDI4Chem, DeBioData, FAIRagro, InnoMatSafety
IPHT	NFDI4Chem
INP	NFDI4Phys
ISAS	FAIRmat
UHD	GHGA, NFDI4Culture, PUNCH4NFDI, NFDI-Neuro, NFDI4Memory, NFDI4Objects, TheoRes
UGOE	DAPHNE4NFDI, NFDI4BioDiversität, NFDI4Chem, NFDI4Culture, NFDI4Earth, NFDI4Health, NFDI4Microbiota, PUNCH4NFDI, Text+, FAIRmat, NFDI4Objects, TheoRes

3.1 Composition of the consortium and its embedding in the community of interest

The composition of NFDI4BIOMAGE as well as the roles of its members are the result of a community-driven proposal preparation process. Information about the NFDI was distributed via multiple channels, in particular the German BioImaging society and the RDM4mic[13] group, but also personal contacts, since the first NFDI conference in 2019. Updates about the process were given at large international microscopy conferences and a participation by the bioimaging community was discussed extensively and officially approved at the GerBI-GMB General Assembly (ELMI[14] meeting 2019). A formal call to get involved was made at a networking meeting in December 2020. Task Areas and co-spokespersons were then formalized through April 2021. Community members were further invited to contribute either as **participants** (**PTPs**, with financial support, listed in section 1) or as **community use cases** (**CUC**s, with project support through data stewards, where appropriate, listed below and in TA5) until mid-July 2021. On July 15th 2021, the final consortium composition was presented at a public community (online)

---

[13] Research Data Management for Microscopy, https://german-bioimaging.github.io/RDM4mic.github.io/

[14] European Light Microscopy Initiative, https://elmi.embl.org/

meeting[15]. As the NFDI4BIOIMAGE must evolve dynamically, an important early-phase task will be to implement the possibility for new members to join and to adjust the consortium composition according to the strategic goals and the mission of the infrastructure (see Governance, sec 3.4).

Our consortium at project start includes all stakeholders of the bioimaging RDM ecosystem: experts in image (meta)data formats and annotation tools (TA1), providers of technical infrastructure (TA2), specialists in multimodal data integration (TA3), developers of bioimage analysis tools (TA4), well-established bioimage community builders and coordinators (TA5 and TA6) as well as a consistent group of interested end-users providing exemplary use cases. Importantly, several of our members are leaders of imaging or image analysis core facilities, service-oriented scientists who are committed to the principles of shared infrastructure. The participation of core facilities thus warrants the direct and efficient involvement of the community of interest. The bioimaging community is particularly large and diverse as it involves all researchers developing and applying microscopy methods in their projects as well as scientists relying on bioimage data for knowledge generation or tool development. In the following we describe the group of co-applicants, their field of expertise and their community involvement.

At the **Heinrich Heine University Düsseldorf** (HHU), the applicant institution, bioimaging is widely applied in a large number of fundamental research projects in biology, chemistry and physics as well as in biomedical and pre-clinical studies. The University's imaging core, the Center for Advanced Imaging (CAi), thus plays a key role in enabling research projects at HHU. The head of CAi is **S. Weidtkamp-Peters,** spokesperson of the consortium. Beyond supporting individual research groups, the CAi participates in the University's life science research consortia, i.e., the Cluster of Excellence CEPLAS, and the CRC1208. S. Weidtkamp-Peters leads the central (Z-) project for microscopy support and co-leads the information infrastructure (INF)-project on RDM of this CRC. In 2015 she pioneered the idea of including bioimage RDM into the service portfolio of the imaging facility and obtained funding through the DFG core facility program for hosting an OMERO instance for bioimage data management. OME Remote Objects (OMERO) is a leading open-source software platform, provided by the OME consortium for the management of bioimaging data. CAi thus became one of the very first imaging cores in Germany offering bioimage RDM services. This project was not only declared a pilot for the institutional RDM practices by the central IT-services, it was also instrumental for integrating bioimage RDM into undergraduate student classes. The respective teaching materials, including all educational resources produced at the CAi for workshops and trainings on bioimaging techniques, are shared with the community[16]. Presently, S. Weidtkamp-Peters is a member of the HHU Task Force RDM, which advises the rectorate on the RDM needs of the various research areas and aims at

---

[15] https://nfdi4bioimage.de/get-involved/community-meeting-2021/

[16] see, e.g.: FLIM Symposium 2021 – Talks & Training: https://www.youtube.com/playlist?list=PLbi8IyTCMI1lL5RhpOLxRRy8RBtwgUTI1

providing tailored support to all scientists of the University. In 2021, together with three other members of the NFDI4BIOIMAGE consortium (E. May, UKON/DKFZ; S. Kunis UOS; R. Nitschke, ALU-FR), she was granted funding in the frame of the DFG program “Information Infrastructure for Research Data” for developing a distributed open bioimage data infrastructure exclusively focused on OMERO. This project, termed “Information Infrastructure for BioImage Data” (I3D:bio), in which S. Weidtkamp-Peters is responsible for elaborating best practices for the implementation of OMERO databases for local IT-departments, fits well into the much broader scope of NFDI4BIOIMAGE and will generate significant synergies. Finally, HHU recently established a central RDM team, which coordinates INF-project activities, and is involved in several overarching RDM initiatives incl. fdm.nrw[17] and DINI-nestor[18].

In 2021, S. Weidtkamp-Peters also became Chair of Board of the co-applicant organization **German BioImaging - Gesellschaft für Mikroskopie und Bildanalyse e. V.** (**GerBI-GMB,** managing director J. Hanne). GerBI-GMB is the scientific society representing the interests of imaging scientists in Germany. It is the successor organization of the DFG-funded German BioImaging network (2012-2017), and as of Oct. 2021 counts 164 members including 62 imaging core facilities (54 German and 8 international) and 12 corporate members. In over a decade of community work, GerBI-GMB has gained extensive experience in fostering communication and transfer of know-how, organizing trainings, workshops, and promoting topics of relevance for imaging core facilities. Image data has been at the center of the activities of GerBI-GMB since the very beginning with a working group dedicated to image analysis. In 2019, this working group aligned with an interest group on bioimage data management, the RDM4mic group that was initiated by S. Weidtkamp-Peters, S. Kunis (UOS) and A. Schauss (UoC) to connect German research groups and core facilities that use the OMERO platform. These activities have been continuously expanding, and GerBI-GMB has now become a strategic partner for OME, resulting in the award of a grant from the Chan-Zuckerberg Initiative (CZI) in the framework “Essential Open Software for Science” to J. Moore, a collaborator at GerBI-GMB, who designed the OMERO database and is also on the OME project management team. GerBI-GMB with the RDM4mic group has contributed to the recently published recommendations on metadata for biological imaging REMBI [6]. Furthermore, S. Kunis and S. Weidtkamp-Peters together with co-authors from GerBI-GMB and University of Dundee published work on a metadata annotation tool, MDEmic, and its application in a specific research context [7].

The **University of Osnabrück (UOS)** hosts the interdisciplinary research Centre for Cellular Nanoanalytics (CellNanOs) in which complex questions concerning the influence of protein microenvironments on organelle function and ultimately on the whole cell are addressed by working groups from physics, chemistry, biology and mathematics. This requires a

---

[17] „Landesinitiative Forschungsdatenmanagement“ in Northrhine Westphalia, fdm.nrw

[18] https://dini.de/ag/dininestor-ag-forschungsdaten/

comprehensive arsenal of cell biological and biophysical methods, in which high-resolution microscopy plays a key role. The scientific work has been supported since 2016 by a suitable data management environment with a focus on the annotation of metadata. **S. Kunis** from UOS, who is responsible for the on-site data management at the integrated bioimaging facility iBIOS at CellNanOs, has the lead of TA1. In the aforementioned I3D:bio project, S. Kunis is responsible for the technical infrastructure work package. She is a trained computer scientist with special expertise in bioimage metadata formats and annotation and developed the metadata annotation tool MDEmic. This tool is integrated in OMERO as OMERO.mde and allows ontology-based facilitated metadata curation [7]. As GerBI-GMB's working group leader on "image analysis and data management" and as co-leader of the working group "metadata" in the Quality Assessment and Reproducibility of Images in Light Microscopy (QUAREP-LiMi) consortium [8, 9], she facilitates global metadata format and schema harmonization for bioimaging data.

The **University of Konstanz (UKON)** is one of eleven universities or university consortia that receive funding as Universities of Excellence in Germany. One of the flagship projects of the Excellence Strategy is the E-science strategy which ensures outstanding data and information structures as well as optimized workflows across all levels. It also encompasses enhanced research data management and open science services. These measures build on a solid foundation of research data management activity. With a long history of a leading role in the open access movement in Germany, the Team Open Science has been established since 2019. The team offers individual support along the whole research (data) lifecycle and manages the distribution of relevant library and IT-services to the researchers across all scientific disciplines. In Baden-Württemberg, the UKON leads the statewide RDM project bw2FDM, project partner of the Science Data Center BioDATEN, and is running the leading national information platform https://forschungsdaten.info. The UKON and its members are highly active in state-wide and nation-wide initiatives like DINI nestor, GO FAIR as well as in YERUN[19] and ERUA[20] in the European context. UKON will bring expertise on metadata, policies, information infrastructures, and community building to the consortium with TA5. With forschungsdaten.info, an established distribution channel to the community can be used. Co-spokesperson for UKON is **E. May**, former Head of the University's Bioimaging Center and founding Chair of GerBI-GMB. For several years, E. May acted as the national contact person for the European infrastructure project for biological and biomedical imaging Euro-BioImaging (EuBI). In the abovementioned DFG project I3D:bio, E. May leads the work package on Communication, Training and Networking. Since 09/2021, E. May has a dual appointment at the UKON and at the German Cancer Research Center DKFZ in Heidelberg, where she took over the role of Chief Enabling Technology Officer and is

---

[19] Young European Research Universities, https://www.yerun.eu/

[20] European Reform University Alliance, https://erua-eui.eu/

responsible for the research infrastructures at the Center. In NFDI4BIOIMAGE, E. May has the co-lead of TA6 on coordination and networking.

The **German Cancer Research Center (DKFZ)** is Germany's largest biomedical research institute and is internationally renowned for its achievements in cancer biology and cancer medicine. Through its outstanding expertise in genomics the center is leading in the development of novel cancer therapies and personalized oncology. It is supported by a wide portfolio of scientific infrastructures, among them one of Europe's most prominent next-generation sequencing facilities, a large microscopy core and a dedicated Omics-IT and Data Management facility. As the coordinating institution of the German Human Genome-Phenome Archive (GHGA), funded by the NFDI since 2020, the DKFZ is building the infrastructure for the secure management of human omics data. A heavily pursued research direction at the institute is currently spatially-resolved omics, i.e., the combination of different omics technologies, in particular sequencing data with spatial information from microscopic imaging. Spatial transcriptomics[21] at DKFZ is performed in the newly created single-cell Open Lab facility, headed by **J.-P. Mallm**, who is the co-lead of TA3 on multimodal data integration. Previously, he has worked within larger consortia on multiomics data generation and analysis and was head of the chromatin and RNA methods lab at the Heidelberg Center for Personalized Oncology (HIPO). The single-cell Open Lab provides infrastructure and training for its users and works closely with the Omics-IT and Data Management Core Facility to ensure efficient data storage and analysis. New recruitments at the DKFZ as the ones of M. Gerstung (AI in Oncology, former group leader at the European Bioinformatics Institute, EMBL-EBI) and F. Hartmann (Systems Immunology and Single Cell Biology) will significantly advance spatial omics-based research at the center in the coming years in the areas of in-situ sequencing and imaging mass spectrometry, respectively.

The intergovernmental organization **European Molecular Biology Laboratory (EMBL)** with six sites across Europe, headquartered in Heidelberg, is a key player in European molecular biology research. Among its core missions is offering scientific services to researchers in member states and the world, where EMBL provides expertise in bioimaging and bioimage analysis**. A. Kreshuk**, group leader at the Cell Biology and Biophysics unit, has renowned expertise in 3D large-scale bioimage analysis, and has developed the *ilastik* software tool enabling researchers in the community to use advanced AI-based image analysis without prior programming experience [10]. In NFDI4BIOIMAGE, she will co-lead the TA4 on bioimage informatics. EMBL provides bioimage analysis training and support, aided by the Centre for Bioimage Analysis (CBA), headed by C. Tischer. Furthermore, EMBL hosts, operates and maintains access to the dedicated bioimage data repository, the BioImage Archive (BIA, [11]) as an Open and FAIR service at EMBL-EBI for which M. Hartley was recently appointed head. EMBL's Advanced Light Microscopy Facility

---

[21] Nature Methods „Method of the Year 2020“, https://doi.org/10.1038/s41592-020-01042-x

(ALMF) is a founding node in the European Life Science Research Infrastructure (EuBI ERIC) and has been a core seed of ELMI, the European Light Microscopy Initiative[22]. EMBL plays a crucial role in both EuBI, as the host of the Bio-Hub headed by A. Keppler (TA6), as well as the Global BioImaging (GBI) Network. Furthermore, the EuBI Bio-Hub Team at EMBL hosts the Image Data Services coordinated by A. Mathur, and the Industry Board coordinated by C Pfander. EuBI Bio-Hub also provides links to the European Open Science Cloud (EOSC) through its participation in EOSC-Life, EOSC-Future, and the recently awarded BY-COVID project. The new EMBL Imaging Centre which allows open access to high-end imaging technologies has already opened in pilot regime, with full operation expected to start in 2022.[23]

With the two co-applicants **DKFZ** and **EMBL**, and with **UHD** as participant institution, NFDI4BIOIMAGE has a strong pillar in the Heidelberg area, and will be able to leverage high-profile inter-institutional collaborations that have been launched recently. In January 2021, EMBL and DKFZ have signed a Memorandum of Understanding to join forces in the advancement of life science data processing and analysis, scientific imaging and small molecule screening[24] All three institutions are member in the Heidelberg-Mannheim Health and Life Science Alliance[25], a recently launched liaison of medical faculties, university hospitals, life science research institutions, and industrial partners in the Heidelberg-Mannheim area with the aim of creating a world-class science hub for the life sciences, health management and medical engineering. NFDI4BIOIMAGE, with its focus on data sharing and knowledge exchange perfectly fits into this landscape and will provide a contact point for the Alliance in all matters regarding bioimage data management.

The **Leibniz Institute for Neurobiology (LIN)** in Magdeburg is a research institute conducting fundamental research on learning and memory and their underlying neuronal mechanisms. Studying learning and memory on different spatial and temporal scales requires multimodal imaging techniques at various levels and generates huge amounts of high-dimensional data, reaching levels of complexity that can only be unraveled by the use of state-of-the-art information technology. This not only includes machine learning AI and other advanced methods for feature extraction, it also crucially requires effective rules and standards for data management, data safety, and data dissemination. LIN is a member of the Leibniz Research Alliance Health Technologies and will bring expertise to the consortium within TA3 that focuses on the integration of cross-disciplinary data types with bioimaging data, as well as their connection to electronic lab notebooks (ELNs) and laboratory information management systems (LIMS). TA3 is co-led by LIN with **W. Zuschratter**. He is head of the microscopy subdivision of the Combinatorial

---

[22] https://elmi.embl.org/

[23] https://www.embl.org/about/info/imaging-centre/

[24] https://www.embl.org/news/lab-matters/mou-dkfz-embl/

[25] https://www.heidelberg-mannheim-alliance.org/

Neuroimaging Core Facility (CNI). W. Zuschratter is member of the DIN-standards committee “Optics and Precision Mechanics“, subcommittee “Microscopes“ NA 027-01-04 AA and co-lead of the central (Z-)project “multimodal imaging platform” of the CRC 854. Moreover, he heads a project within the recently funded biophotonic research network “Time resolved Raman- and metabolic Imaging spectroscopy unit (TIRAMISU); BMBF FKZ 13N15790. Besides providing imaging infrastructure, CNI also offers dedicated data management, image analysis & software engineering expertise (T. Stöter) for the departments and research groups of LIN. T. Stöter is an initial contributor to the COST action Network of European BioImage Analysts (NEUBIAS)[26] and later became a member of its management committee. In this context, it is significant that LIN is co-applicant in NFDI-Neuro and thus plays an integrative role between both consortia by building a bridge through strong collaboration with the imaging-intensive research field of neurobiology.

Research at the **Leibniz Institute for Natural Product Research and Infection Biology - Hans Knöll Institute (Leibniz-HKI)** in Jena focuses on the investigation of infection processes and microbial natural products using the latest scientific methods and technologies. The aim is to translate the findings into new diagnostic methods and therapies. The research group Applied Systems Biology (ASB) at Leibniz-HKI, headed by **M. T. Figge,** co-lead of TA4, is specialized in image-based applied systems biology. ASB has more than ten years of experience in the analysis of multimodal biomedical data (e.g., microscopy data and micro-CT data), the modeling of biological processes and data management. Furthermore, ASB has expertise in traditional statistical methods for data analysis, image analysis, and artificial intelligence (AI). Within AI, Leibniz-HKI focuses in particular on the field of computer vision, which is used to identify different cell and tissue types from microscopy data. The development of novel image analysis algorithms is made with a user-friendly focus, as for example in MISA++ and JIPipe. In addition, image-based modeling approaches are applied, for example, to model whole blood infections and the immune response to fungal infections in lung alveoli. ASB also has experience and resources for data management in general and for the management of an OMERO instance at the Leibniz-HKI in particular. ASB is part of the Jena Cluster of Excellence “Balance of the Microverse”, where M. T. Figge leads the “Data Synopsis” research area and is a member of the Microverse Imaging Center management board. He is principal investigator in projects on image-based systems biology in two funded CRCs – FungiNet 124 and PolyTarget 1278.

At the **University of Leipzig (UL)**, Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), **R. Haase**, formerly group leader at PoL Dresden, is a recognized bio-image analysis expert, who has contributed novel tools like GPU-acceleration for Fiji [12]. R. Haase is a strong community advocate for open-source software development [13], collaboration with A. Kreshuk)

---

[26] European Cooperation in Science and Technology (COST), https://eubias.org/NEUBIAS/

and produces openly accessible community training materials in image analysis[27] which have been accessed more than 80000 times in about 1.5 years, and are among the six most often highlighted resources in the 2020 Survey of the NIH Center for Open Bioimage Analysis [14]. Fiji was also developed at the Dresden biomedical science campus [15][28]. R. Haase was active in NEUBIAS, lastly as scientific organizer of NEUBIAS training schools [16]. He is an experienced lecturer of Applied Bio-image Analysis. In NFDI4BIOIMAGE, R. Haase contributes to TA4 and co-leads TA5 on Communication and Training leveraging his qualification and passion for community exchange and education. The **University of Technology Dresden (TUD)** has a strong focus on science and engineering in the biomedical context. For example, the recently founded Cluster of Excellence “Physics of Life” (PoL) bridges the disciplines Physics, Biology and Computer Science to unravel the very basic principles of how molecules form cells and how cells form tissues. TUD changed its status to Participating Institution as of January 2024. It has a well-established bioimaging infrastructure. The Biopolis Dresden Imaging Platform (BioDIP) connects 11 facilities from five institutions incl. R. Haase’s group. BioDIP puts TUD in an excellent position to contribute to and disburse bioimaging RDM training. BioDIP is furthermore participating in I3D:bio as a naïve site for testing OMERO implementation guidelines.

With more than 45,000 students, the University of Münster (formerly Westfälische Wilhelms-Universität Münster is one of the largest universities in Germany. “Cell Dynamics and Imaging” is one of its core research focus areas, which is underlined by nine different CRCs and CRUs. The Cells in Motion Interfaculty Centre (CIM) is a central, cross-faculty research institution, bringing together and supporting researchers from a wide variety of disciplines and emerged from the Cluster of Excellence “Cells in Motion” (2012-2019). **T. Zobel**, co-lead of TA5, is a trained developmental biologist and imaging specialist. He is the head of the light microscopy unit in the Münster Imaging Network, which is part of the CIM, providing access to high-end microscopy, and supports research projects in CRC1348, CRC1009, CRC1450 and CRU342. Besides the access to high-end microscopy, the Imaging Network also helps its 350+ users to analyze and store their data and provides support in RDM related questions with respect to bioimaging data. The new Multiscale Imaging Center, which will open in spring 2022, will greatly enhance WWUs capacity for transdisciplinary imaging and further boost the existing expertise. T. Zobel has six years of experience in teaching RDM tools and is a co-speaker of the RDM4mic group. As an integral part of research infrastructures, the IT center of the University of Münster (head and CIO R. Vogl) offers both hardware and software solutions, and is closely collaborating with the Imaging Network. The collaborative competence center eScience, bridging IT and the university and federal state library (ULB), provides research infrastructure and data management services. **M. Blank-Burian**, leading the Research Infrastructures section, will co-lead the TA2 on technical

---

[27] For example, see R. Haase’s public lectures on BioImage Analysis: https://www.youtube.com/watch?v=e-2DbkUwKk4&list=PL5ESQNfM5lc7SAMstEu082ivW4BDMvd0U

[28] https://analyticalscience.wiley.com/do/10.1002/was.0004000112

infrastructure provision in NFDI4BIOIMAGE. He contributes extensive experience in Infrastructure as a Service (IaaS) and Software as a Service (SaaS) provision. He manages the University of Münster Cloud, based on OpenStack and Kubernetes, bringing cloud-computing expertise and infrastructure to the consortium.

Founded in 1457, the **University of Freiburg (ALU-FR)** is one of the oldest German universities and is now one of the nation's leading research and teaching institutions, evidenced inter alia by its membership in the League of European Research Universities (LERU). It actively fosters interdisciplinary research, and it is one of the few universities offering world-class research environments in the classical as well as in the modern disciplines. ALU-FR has a strong focus in bioinformatics and contributes to TA2 with technical infrastructure, access to a wide range of research data management services and compute resources like HPC systems, the de.NBI-cloud and bwCloud. Co-lead of TA2 is **B. Grüning**, a trained bioinformatician and head of the Galaxy Europe working group. He is the principal investigator of the de.NBI-epi project, responsible for the de.NBI-cloud (Freiburg), co-founder and co-chair of the ELIXIR Galaxy Community, technical coordinator of ELIXIR-DE, and the ELIXIR Tools platform ExCo. In EOSC-life he is Task Lead in the Tools and Workflows work package. With the support of the eScience department (led by D. v. Suchodoletz) of the Freiburg computer center, he provides various base and high-level services as well as expertise for research data management. D. v. Suchodoletz is an infrastructure specialist, and his group activities center around large-scale research storage and compute infrastructures as well as research data management. He is co-leading the Research Data Management group of the University and is highly connected within the RDM working group of Baden-Württemberg. He is the speaker of the NFDI DataPLANT consortium and directly involved in its infrastructure development.

Our consortium is well-balanced with respect to highly qualified young investigators in responsible roles (Kunis, Zobel, Haase, Grüning, Mallm, Blank-Burian) and a significant portion of female TA leads (Weidtkamp-Peters, May, Kunis, Kreshuk; 36%). Taking into account both applicants and participants (for the latter, see sec 1) the composition reflects the integration of all relevant RDM stakeholders in the bioimaging community and has the following specific strengths:

- The inclusion of directors of imaging core facilities (HHU, WWU, UKON/DKFZ, LIN, UOS, UoC) who are also involved in active research ensures not only service-orientation and user proximity, but also an excellent position to stay up-to-date about new developments in a rapidly evolving area of technology like bioimaging and its related RDM.
- With the co-applicant GerBI-GMB, we take advantage of the leading communication and interaction platform for bioimaging scientists in Germany and its wide-reaching national and international network. Moreover, GerBI-GMB has implemented an organizational structure

and decision-making processes[29] that are well-accepted in the community and can be a starting point for the consortium.

- IT-professionals bring in competence on technical infrastructure for interdisciplinary and bioimaging-tailored RDM (ALU-FR, University of Münster, DKFZ, EMBL, LIN, UHD, FZJ).
- We include software developers with a track record in the area of RDM tools for bioimaging (GerBI-GMB, UOS, ALU-FR, TUD, MPIEB, IPHT, HAWK, EMBL, OME, UHD, INP).
- Library and information service specialists bridge generic and discipline-specific RDM expertise (HHU, ALU-FR, UHD, UKON, UoC).
- We include image analysis experts for "point-and-click"-, automated, and AI-based image analysis (EMBL, HKI, TUD, ISAS, JGU, HAWK, FZJ).
- Members of existing relevant community initiatives in bioimaging / bioimage analysis RDM are included (OME consortium, RDM4mic, NEUBIAS, QUAREP-LiMi, EuBI, GBI, EOSC, ELIXIR, Galaxy, de.NBI).
- As CUCs we have already enrolled a number of sites who have volunteered to test the tools and services that will be elaborated by the consortium (S. Rehwald, Universität Duisburg-Essen, UDUE; W. Moebius, Max Planck-Institute for Experimental Medicine, MPIEM; J. Steinkötter, Max-Delbrück Center Berlin, MDC; C. Meesters and B. Huelsmann, Johannes Gutenberg Universität Mainz, JGU; S. Nahnsen, Quantitative Biology Center Tübingen, QBiC; and several other use cases associated with co-applicant and participant institutions (see TA5, M5.4).

3.2 The consortium within the NFDI

NFDI4BIOIMAGE is conceived as a data-type-centered consortium that addresses the need of the scientific community for the management of bioimaging data. It thus differs from the majority of NFDI consortia established so far which are mainly discipline-driven. We believe that bioimaging data, because of its ever-increasing wealth and complexity, requires a dedicated RDM initiative targeting all fields of life and biomedical sciences. This notion is underpinned by the feedback received from existing or planned NFDI consortia (e.g., DataPLANT, NFDI4Biodiversity, NFDI-Neuro, NFDI4Immuno). They highly welcome a professional infrastructure for bioimage data management that provides tools and best practices, and with which they can collaborate on the development and interoperability of discipline-specific formats and standards. In particular, for the development of novel, AI-based approaches for bioimage analysis, which is a circumscribed and strong disciplinary component of our consortium, well-annotated image data is of prime importance across research domains. This integrating role is one of the key assets of NFDI4BIOIMAGE for the NFDI (Figure 2).

In the following, we outline existing and planned interactions with other NFDI consortia:

---

[29] https://gerbi-gmb.de/society-structure

DataPLANT (1st Call): According to an internal survey, 50% of the DataPLANT community uses bioimaging methods, but RDM standards for bioimaging and bioimage analysis are insufficiently addressed. Collaboration between NFDI4BIOIMAGE and DataPLANT aims at defining standards for compatibility with the Annotated Research Context (ARC), a community-specific implementation of “FAIR Digital Objects” (FDO) by DataPLANT. This collaboration has already been initiated between members of RDM4mic at HHU and UoC in advance of our proposal, which has allowed us to define discrete collaborative work packages for our work programme. We will leverage synergies by building a connection between ARC as a generic plant science standard and bioimaging formats and platforms (e.g., OMERO). We will extend this collaboration further within the NFDI by building a bridge to standard file organization structures for microscopy based on the BIDS standard used by NFDI-Neuro (TA1 and TA3). The DataPLANT consortium has implemented a Data Stewards commissioning model similar to our Help Desk and Data Stewardship concept (sec 3.5). We will also facilitate collaboration between our consortia’s data steward groups and with data stewards employed at Excellence Clusters and CRCs, where available (e.g., D. Brilhaus, HHU). Via our co-applicant ALU-FR (applicant in DataPLANT) we leverage synergies to develop an open DaaS-infrastructure, which we plan to collaboratively explore and contribute as cross-cutting topic (D. v. Suchodoletz, B. Grüning).

GHGA (1st Call): This consortium is related in structure to the NFDI4BIOIMAGE as being data-type centered. GHGA is establishing an infrastructure specialized for human omics data, including genome sequencing data. Close interactions between GHGA and NFDI4BIOIMAGE will be highly beneficial and are supported by the structure of our consortium. The speaker of GHGA (O. Stegle) and – in NFDI4BIOIMAGE – the co-lead of TA6 (E. May) and the co-lead of TA3 responsible for spatial multiomics data integration (J.P. Mallm) are all affiliated with DKFZ. This constellation maximizes the potential for seminal contributions for the linkage of complex imaging data to genomics data, while being fully compatible with the data privacy and associated ethico-legal questions.

NFDI4Patho (3rd Call): The focus of this consortium is on Digital Pathology with the aim to enable access and sharing of whole slide images from histopathology to promote the development of AI-based analysis methods and bringing these into clinical application. Major hurdles are data privacy protection and medical product licensing regulations that are unique for medical imaging in contrast to the open source and open data practices favored by the bioimaging community. In three joint meetings in 2020 and 2021, we identified potential areas of collaborations, in particular metadata and file format interoperability, and data sharing across management systems. A shared long-term vision for data accessibility and re-use would be making imaging data from basic research studies easily available alongside histopathology results in the context of multidisciplinary tumor boards.

DeBioData (3rd Call): With a focus on chemical biology, the DeBioData consortium aims to harmonize RDM standards for basic research in molecular processes of human diseases and associated pre-clinical drug development. To this end, software packages for data standards, decentralized storage, entry-querying and data access will be developed. Among the diverse data types generated in this research area, imaging plays an important role for high-throughput assays and cellular process modelling. Previous collaborations between members of DeBioData and members of NFDI4BIOIMAGE have showcased the importance for developing shared RDM standards (e.g., within EOSC-Life demonstrators[30]). Shared goals will be to facilitate rich metadata curations for HCS-data with multiple feature annotations, e.g., by Cell Painting assays. For this purpose, we will work together with a shared PTP case (JGU, P. Czodrowski). NFDI4BIOIMAGE will contribute schemes and workflows to facilitate deposition of large datasets to BIA or to the Image Data Resource (IDR)[31], a repository for highly annotated reference datasets. DeBioData will ensure interoperability with their software packages and data linkage across relevant repositories for connected data types (e.g. chemical compound libraries).

NFDI-Neuro (3rd Call): Neuroscience research acquires data from sub-cellular scale of individual neurons up to the systems scale of whole brains. The planned NFDI initiative for neuroscience (NFDI-Neuro) is concerned with multi-modal neuro-scientific data. It has a strong focus on medical imaging standards, but bioimaging as covered by NFDI4BIOIMAGE is also a widely applied modality. Therefore, we see a clear potential for synergies between the two NFDI initiatives. In TA3 (LIN, W. Zuschratter) we build upon the BIDS directory structure developed within the neuroscience community and used by NFDI-Neuro as a microscopy-tailored directory structure extension. At the same time, we want to contribute to BIDS in the BIDS Extension Proposal for Microscopy with FZJ (T. Dickscheid). Employing BIDS as a data structure for bioimaging data enables easier data transfer to neuroscience repositories, such as EBRAINS, or the public GIN[32] repository developed and hosted by UHD (CRC 1158 members R. Kuner, T. Kuner, C. Beretta, D. Mittal are individual participants of the NFDI-Neuro initiative).

NFDI4Immuno (3rd Call): This consortium addresses the development and implementation of a comprehensive data model for all data types that are relevant in the field of immunology. Imaging data have been identified by the consortium as increasingly important for immunological research as they allow, e.g., for the identification and detailed localization of immune cells in complex histological structures. In a workshop on August 16th, 2021, we identified the following collaborative topics: establish and share databases for reagents that are widely used both in microscopy and flow cytometry, i.e., chemical fluorophores and fluorescent proteins as well as antibodies, and develop common ontologies for these reagents and molecules. We will further collaborate in a tangible use case scenario provided by NFDI4Immuno on implementing a workflow for the submission process of imaging data as part of a representative dataset to an

---

[30] https://www.eosc-life.eu/d6/

[31] https://idr.openmicroscopy.org/

[32] https://gin.g-node.org

appropriate repository and on the development of RDF-compatible datasets (JSON-LD, knowledge graphs) and shared authentication and authorization infrastructure (AAI).

NFDI4Phys (3rd Call): The NFDI4Phys consortium addresses RDM needs in various areas of disciplinary and transdisciplinary physics. Especially in the fields of biological physics and application-oriented research in low-temperature plasma physics, bioimaging represents an important investigation method and the RDM solutions and standards developed within the framework of NFDI4BIOIMAGE are of particular interest. We will collaborate with NFDI4Phys via our PTP M. Becker (INP), who is also a participant in NFDI4Phys. On the one hand, domain-specific requirements regarding bioimaging data will be recorded and taken into account in the development of targeted RDM tools. On the other hand, concrete solutions for the integration and harmonization of disciplinary RDM developed in the context of NFDI4Phys will be designed together with the INP (TA1 and TA3). To support the establishment of standardized RDM solutions for handling bioimaging data in physics, we also plan a regular exchange with other relevant use cases in NFDI4Phys (TA5 and TA6).

NFDI4Health (2nd Call): Patient-derived data in epidemiology, public health and clinical-trial-based medicine are the focus of NFDI4Health, which includes a broad spectrum of data types from a multitude of data sources from providers with different governance structures and processes. Record linkage for patients and complying with highest standards for data privacy protection are key challenges. One subsection is the medical imaging field (radiology, cardiology, pathology, etc.) primarily relying on the DICOM standard, which plays a minor role in the bioimaging community. While data conversion to, e.g., OME-TIFF is possible, often different image data management platforms are used (e.g., PACS and XNAT in (pre-)clinical imaging vs. OMERO in bioimaging). As the developments of AI-based bioimaging methods follow common principles in both bioimaging and medical imaging, better interoperability of formats and software would produce important synergies. Moreover, developments in bioimaging become increasingly promising for novel clinical applications. One example is the case of C. Chalopin (HAWK) focusing on hyperspectral imaging and automated analysis for in-surgery application. As addressed in a workshop of the NFDI4Health consortium (Feb. 4th, 2021), we aim to identify use cases for bioimaging - medical imaging interoperability.

In addition, we discussed common topics of interest with **InnoMatSafety** (shared training materials, library of modular standard operating procedures for ELNs, data quality control); and **NFDI4Chem** (integration of imaging data in the Chem ELNs). We see potential for collaboration with consortia **NFDI4Microbiota** and **NFDI4Biodiversity**, whose research domains make wide use of bioimaging.

Beyond synergizing with single NFDI consortia, NFDI4BIOIMAGE is highly interested in contributing to **cross-cutting topic sections** within the NFDI and also sees potential for co-developing basic services. These plans are outlined in detail in the measure M6.7 of the work programme including a dedicated “Data Formats Ambassador”. We believe that the solid embedding of the NFDI4BIOIMAGE in the community puts our consortium in a unique position to interpret day-to-day user needs on the one hand and disseminate the practices and benefits of (bioimage) RDM on the other hand. Since bioimaging projects are intrinsically interdisciplinary combining aspects from biology, physics, chemistry and also medicine, we expect NFDI4BIOIMAGE to be attractive for scientists working at the boundary between different disciplines who do not find their RDM needs covered by strictly disciplinary consortia. Finally, we foresee an impact of NFDI4BIOIMAGE on the wider public, as microscopy images are prime vehicles for communicating scientific content to lay audiences, which will only grow as the accessibility and usability of data increases.

3.3 International networking

The NFDI4BIOIMAGE leverages the manifold international cooperations of its members and will make them coalesce into a coherent networking strategy (TA6). It also builds upon more than a decade of community building and networking by GerBI-GMB. A close and long-lasting collaboration exists with OME (see 3.1) through which our consortium can connect to 100s of unique OMERO instances and thus interact with bioimage data management worldwide. The TA6 leads were invited speakers at a recent OME Community meeting with more than 240 attendees. The international initiative Quality and Reproducibility for Instruments and Images in Light Microscopy (QUAREP-LiMi)[33] was spearheaded by the Life Imaging Center at our co-applicant institution ALU-FR and meanwhile counts 300 members from academia and industry across the globe [9]. It has installed a working group dedicated to microscopy metadata and is an important communication channel to microscopy equipment manufacturers but also to national standard-setting bodies like ISO[34] and NIST[35]. Our co-applicant EMBL will be crucial for connecting the NFDI4BIOIMAGE to the European BioImage Archive (BIA) hosted at EMBL-EBI. The BIA is a public archive for (published) biological image data. It is an explicit goal of the consortium to promote access to and usage of BIA among German imaging scientists, for which EMBL has committed to provide capacity. EMBL is also the Image Data Service provider for EuBI ERIC supporting EuBI Nodes with data services and is involved in various projects of the EOSC. Thus, EMBL will be instrumental for connecting the NFDI4BIOIMAGE to EU-level activities and develop an infrastructure that aligns with the open and FAIR scientific data space that is in the making in the European Research Area. The co-lead of TA6 at UKON/DKFZ has been national contact person for EuBI for several years and as such is familiar with the European landscape of bioimaging community organizations, like France BioImaging, BioImagingUK, Czech-Bioimaging, Swedish Bioimaging and more. A further strong connection to European initiatives is provided by

---

[33] https://quarep.org/

[34] International Organization for Standardization, https://www.iso.org/home.html

[35] National Institute of Standards and Technology, https://www.nist.gov/

our co-applicant ALU-FR who hosts the European Galaxy Server and bridges to ELIXIR. A recent demonstrator project[36] funded through EOSC-Life for the integration of IDR and Cellprofiler into Galaxy and carried out at EMBL and ALU-FR showcases the potential for collaborations of international impact and visibility in our consortium. A further important international partner of NFDI4BIOIMAGE is Global BioImaging (GBI), an international network of advanced bioimaging facilities and communities coordinated by EMBL. GBI has gained a high profile as a promoter of knowledge exchange and training for imaging scientists and features a working group on image data management. Both leads of TA6 have been invited speakers at GBI Exchange of Experience workshops. On the topic of image analysis, there are links between GerBI-GMB and the Network of European Image Analysts NEUBIAS. The network was created in the frame of a COST action, during which the speaker of the image analysis working group of GerBI-GMB coordinated the German chapter of NEUBIAS. NEUBIAS and GerBI-GMB have jointly promoted the novel "Rent-the-trainer" concept as a training format for image analysis that will be taken up by TA5 in the NFDI4BIOIMAGE. The working group Benchmarking and Datasets of NEUBIAS elaborates standards for the interoperability of image analysis software, a topic of interest for TA4.

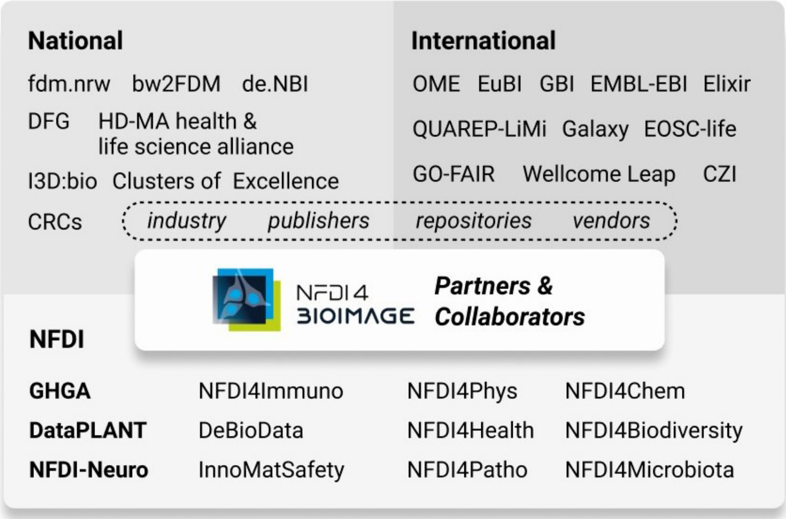

Figure 2: National and International embedding of the NFDI4BIOIMAGE consortium

Outside of Europe, close contacts exist with BioImaging North America (BINA), an organization bringing together bioimaging communities in Canada, the United States and Mexico. A collaboration on metadata annotation tools for OMERO as part of a package of complementary tools which facilitate metadata annotation of light microscopy image data (Micro-Meta-App [17], MethodsJ2 [18], recently resulted in a common publication [7]. The Australian Characterisation Commons At Scale[37] project represents another national initiative which has the aim to overcome the data challenges of natural and life sciences including all modalities of imaging data.

---

[36] https://www.eosc-life.eu/wp-content/uploads/2021/02/D6-Wrap-up-Seminar.pptx.pdf

[37] https://ardc.edu.au/project/the-australian-characterisation-commons-at-scale-accs/

We will exploit these existing contacts (Figure 2) and build new ones for bringing NFDI4BIOIMAGE as a new player on the international stage of bioimage data management. As an established NFDI consortium, we will connect to overarching scientific RDM activities and contribute the perspective and represent the interest of the (German) bioimaging community. Links exist to OpenAIRE via UKON and UGOE, and with FAIRsharing.org[38], where the EuBI BioHub curates a resources collection. Further relevant RDM initiatives are the Research Data Alliance (RDA)[39], GO FAIR[40], the FDO Forum[41], and FAIRsFAIR[42]. In particular, our concept for an FDO description for bioimage data will be a topic of common interest (TA1). We will consider hosting activities under the umbrella of these broad organizations to engage with the wider scientific RDM community. The Networking Coordinator (TA6) will be in charge of promoting the international visibility and integration of NFDI4BIOIMAGE, and further explore the landscape of interesting partners.

3.4 Organisational structure and viability

The organisational structure of NFDI4BIOIMAGE is conceived to facilitate the development and operation of the infrastructure with defined roles, decision making responsibilities and internal communication channels. It enables integrating the needs and perspectives of all consortium members and warrants communication with and input from the whole community of interest. The initial steering bodies (Figure 3) are composed of members from the proposal-submitting institutions. Importantly, the governance will enable new members that may emerge from the community in the course of the project to join the consortium. The structure also enables responding to future developments of the NFDI as a whole that may result in recommendations or guidelines for the organizational structure of consortia. NFDI4BIOIMAGE will be founded as a “consortium according to the Bund-Länder-Vereinbarung BLV“, as foreseen by DFG regulations. To work together and interact with other consortia, e.g. on cross-cutting topics, under the roof of the NFDI association, we aim to become a “consortium according to the NFDI statutes”, a process that will be facilitated by our governance (TA6, M6.1).

Board of Directors (BoD): This Board includes the leaders of the coordination and networking task area (TA6), of which one is the **spokesperson** of the consortium, and three members elected from the Steering Board (see below). The spokesperson represents the consortium towards the NFDI directorate, the NFDI consortia, the DFG and the general public. She has the oversight of the whole project and represents the applicant institution, which in turn oversees correct terms of

---

[38] https://beta.fairsharing.org
[39] https://rd-alliance.org
[40] https://www.go-fair.org/
[41] https://fairdo.org/
[42] https://www.fairsfair.eu/

use and regular reporting. The BoD is the main decision-making body of NFDI4BIOIMAGE. It has the financial authority over the budget and decides on funding allocations. It reports regularly to the Steering Board (every 8-10 weeks). For decisions of larger impact (changes in the composition of the consortium, funding allocations deviating from the proposal, etc.), the BoD elaborates a decision paper to be discussed in the Steering Board.

The **NFDI4BIOIMAGE Office** consists of the Managing Director of NFDI4BIOIMAGE (@HHU) and a Networking Coordinator (@UKON/DKFZ) working closely together to support the BoD. It will be their responsibility to ensure that the project measures and decisions implemented by the Board of Directors comply with the DFG Rules of Good Scientific Practice, the DFG funding guidelines, the general laws, and the rules of the budget management office of the applicant institution (HHU). The NFDI4BIOIMAGE Office executes the tasks of TA6 regarding overall operational implementation (mainly @HHU) and national and international networking (mainly @DKFZ). It will prepare documents for meetings, elections, and assemblies, organize internal and external communication channels, and execute mandates by the BoD or planned actions

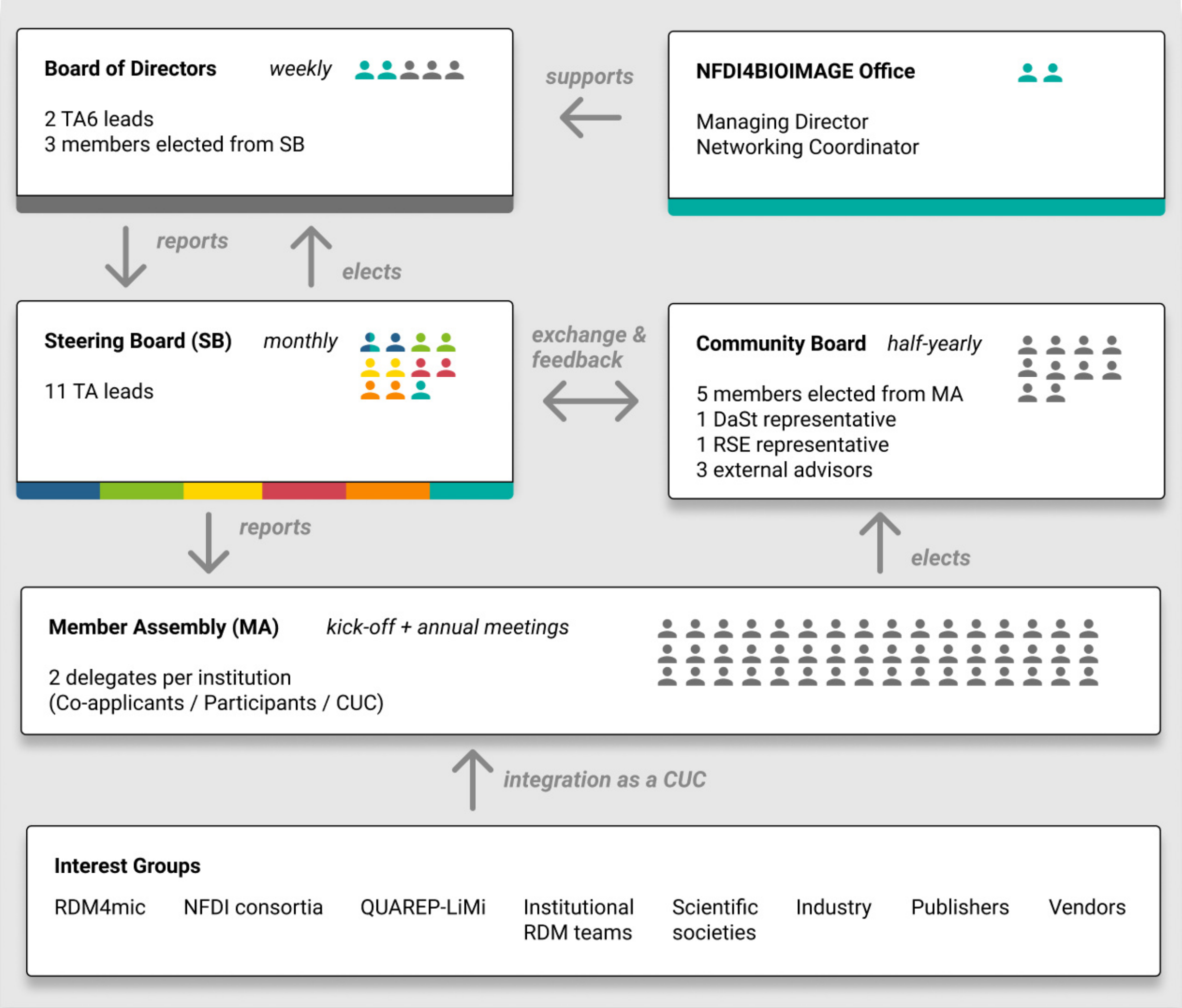

Figure 3: Organisational Structure of NFDI4BIOIMAGE

according to the proposal. Monitoring the output of the infrastructure and preparing reports and drafts for future proposals are also duties of the NFDI4BIOIMAGE Office.

Steering Board (SB): The SB is the main deliberation body of NFDI4BIOIMAGE, meeting once a month. It will provide strategic advice to the BoD regarding funds allocation, modifications of the work programme, future directions, organization, strategic partnerships etc. It consists *ex officio* of the co-spokespersons, which are the TA leads. The co-spokespersons are responsible for achieving the results described in the work programme of their respective TAs, coordinating the contributions of the PTPs and involving the CUCs. The spokesperson of NFDI4BIOIMAGE acts as the inaugural Chair of the SB. Three members of the SB are elected to the BoD. In the course of the project, SB members may take over dedicated roles, if necessary (e.g., industry coordination). The SB is responsible in particular for recommending the commissioning of Data Stewards to support RDM projects from the community, is advised by the Community Board and reports to the Member Assembly.

The **Community Board (CB)** consists of five members elected by the Member Assembly, one person representing the Data Stewards, and one person representing the Research Software Engineers (RSE, see sec 3.5). This board ensures that the interests and needs of all members of the consortium are represented and communicated to the SB. The CB elects one Chair who is in regular contact with the Chair of the SB and can attend meetings of the latter as a guest. In addition, the BoD, after consultation with the SB, will propose three external advisors to join the CB.

The **Member Assembly (MA)** collects all members of the consortium (applicant, co-applicants, PTPs and CUCs). The PTPs are integrated into one or more measures in the work programme, and receive funds accordingly. They work together with the co-applicants on a sustained basis to realize the goal of building the NFDI4BIOIMAGE infrastructure. CUCs are stakeholders who are interested in the deliverables and services of the infrastructure and are willing to serve as a test-site or want to connect their local activities in bioimage RDM to the NFDI4BIOIMAGE. The exchange with CUCs is meant to fine-tune and validate the outputs of the work program and is based on mutual interest. CUCs do not receive direct financial support but are prioritized for Data Stewardship support (sec 3.5). They are represented in the governance of the consortium via the MA. During the project phase, community members, e.g., from one of the interest groups, can propose to become a CUC via our contact channels (sec 3.5 and TA6, M6.4). Each member institution mandates two representatives to the MA that will meet twice in the first year and at least once a year in the following. The MA elects five members into the CB and is presented reports on the activities and achievements of the consortium by the SB.

Interest groups comprise all groups or organizations that are confronted with bioimaging data in their profession and stand for the broad anchorage of the consortium in the community. They comprise institutional RDM teams, specialized bioimaging data groups like RDM4mic and QUAREP-LiMi, but also members of industry, of scientific societies, and other NFDI consortia (see 3.2). Their voice is heard through the many members of the consortium who interact with or are active in these groups, some of which are fully open for participation. Moreover, interest groups will be invited to attend yearly public meetings, of which one has been held during the proposal preparation phase (July 15th, 2021). These groups not only ensure that NFDI4BIOIMAGE remains informed about the rapidly evolving bioimage RDM needs in the community, they also increase the consortium's impact and visibility (as well as the visibility of the NFDI as a whole) (see TA6, M6.5). In addition, the consortium's operational model (see sec 3.5) is designed to provide low-barrier contact points for the accessibility of services and support. A significant portion of the requested funding is planned for community support by Help Desk and Data Stewardship. Data Stewardship support requests from the community are reviewed and approved by the SB.

Viability: The consortium structure entails several features that promote its long-term viability. Firstly, NFDI4BIOIMAGE builds upon existing and well-established community structures that are self-funded. These include interest groups like RDM4mic and QUAREP-LiMi, but foremost the co-applicant GerBI-GMB. The society is a self-sustained, non-profit association that has successfully mastered a change at the highest management level when the Chair moved from UKON to HHU in March 2021. GerBI-GMB has gained a high reputation and visibility (average increase in members p.a. ~35% since 2017) and is supported by a network of dozens of relevant stakeholders in bioimaging from academia and industry. Over the last 2-3 years, bioimage RDM has emerged as a field of focus within the society, catalyzed by the NFDI, and has resulted in the new DFG project "I3D:bio" for building an OMERO-centered distributed infrastructure in Germany starting in January 2022. Very recently, GerBI-GMB could also attract funds from prestigious international organizations, the Chan-Zuckerberg Initiative and Wellcome Leap, which will not only synergize with the NFDI4BIOIMAGE but put GerBI-GMB center stage of global biomage RDM activities.

Secondly, the overall goals of the NFDI4BIOIMAGE and its planned measures are well-aligned with the strategic priorities and the RDM activities of the applicant and co-applicant institutions (see also sec 3.1). Not only institutional RDM teams, also RDM initiatives at the States level have been informed about and partly participated in developing concepts for the NFDI4BIOIMAGE. Embedding our consortium in the larger German RDM landscape is an important component of our sustainability strategy. The commitment of the co-applicant institutions is underpinned by their in-kind contributions, in particular concerning the provision of technical infrastructure for launching the NFDI4BIOIMAGE project, but also by recent large investments like the new EMBL Imaging Centre, the Multiscale Imaging Center under construction at University of Münster, or the federated storage system for RDM and long-term data publication of Baden-Württemberg, bwSFS.

We do not plan to raise usage fees during the project phase, but we are aware that these are an important element of sustainability for which we need to elaborate a cost model. A possible scenario might entail lump sums in the form of membership fees, with which we have gained positive experience within GerBI-GMB. The actual financial burden for individual users will very much depend on the evolution of the whole NFDI and possible future funding lines to support it.

Several institutions who are members in NFDI4BIOIMAGE have already established bi- and multilateral collaboration frameworks with the aim, amongst other, to harmonize RDM practices and facilitate data sharing. These networks represent an additional layer of institutional embedding of our consortium and will also promote sustainability. Examples are:

- Heidelberg-Mannheim Health and Life Science Alliance (EMBL, DKFZ, UHD)
- DKFZ and EMBL Memorandum of Understanding
- Helmholtz Imaging Platform HIP and Helmholtz Metadata Consortium HMC (DKFZ, FZJ, UFZ, MDC)
- Multi-site Clusters of Excellence (HHU & UoC; IPHT & HKI)
- RDM initiatives of the States (bw2FDM: ALU-FR, UKON, UHD; fdm.nrw: HHU, DUE, University of Münster)

The excellent international links of the NFDI4BIOIMAGE will also contribute to the viability of the infrastructure. In particular, through EMBL and EuBI the consortium is in an optimal position to bridge to the Bioimage Archive at EMBL-EBI (see sec 3.3), a long-term resource for the scientific community. TA6 will specifically work on strengthening these strategic international partnerships.

3.5 Operating model

NFDI4BIOIMAGE starts as a project coordinated by HHU to create an infrastructure for bioimage data services. The roadmap for achieving this goal is outlined in the work programme in which all co-applicants are responsible for implementing specific measures while collaborating with each other, with PTPs, and CUCs. The necessary funding is allocated by HHU according to the work programme. The administration of the budget lies with HHU. Our governance permits responding to new demands that may come up during the project time. We consider a certain degree of flexibility for allocating the budget essential in the dynamic and diverse field of bioimaging. The operational model will evolve as the infrastructure takes shape. We plan a single entry point to the NFDI4BIOIMAGE in the form of a webportal where users are presented the structure and the services of the infrastructure and how to actively engage. A central component of this resource is a bioimage RDM Help Desk that collects and manages requests for assistance from users. Behind the Help Desk is a distributed team of experts, the **NFDI4BIOMAGE Data Stewards** (**DaSts**). These are staff located at the applicant and at selected co-applicant institutions who work on the measures of the respective Task Areas, and, in addition to their project work, assist users across

Germany, remotely or on-site, in solving their bioimage RDM issues (see M5.4 and M6.3). The DaSt sites were chosen based on i) their specific expertise and its relevance for the community, ii) prior experience with data stewardship consulting, and iii) the geographic distribution. Thus, the DaSts form a network of experts with complementary skills that can be deployed across Germany. Their dual involvement in project work and service has several benefits: DaSts are up to date about the latest developments and deliver state-of-the-art tools and advice; they can directly feed insights from their "field work" back into the Task Areas ensuring that the latter always follow user needs; and they can flexibly adjust the time allocated to projects tasks and service depending on demand thereby increasing the overall efficiency.

DaSt engagement will vary depending on the raised issues. For basic questions, users will be directed to ready-to-use solutions and/or training material available online (image.sc, etc). Minor issues that require DaSt input are managed by the team autonomously. The team will meet weekly to discuss user requests. When a substantial use case comes up, DaSt can propose it as a new CUC. To this end, CUC candidates file a short proposal that is evaluated by the DaSts concerning feasibility and projected DaSt time and submitted to the BoD for decision. We have already pre-identified several CUCs that will serve as pilots for the DaSt operating model. All CUC data stewardship actions are documented including among others information on: Which problems were addressed? Which deliverables of NFDI4BIOIMAGE and other consortia could be used? What were the lessons learnt to improve the Data Stewardship guidelines and training materials?

Research software engineers (RSE) working in the NFDI4BIOMAGE will form an operational unit, the **RSE** team, which will build working groups where necessary. A substantial part of the tasks to be accomplished during the project phase entails writing code to create or modify software i.e. for expanding ontologies or modify software to implement the data standards and make them appropriately accessible, such as by implementing specific requirements for visualizing the various formats or appropriate application programming interfaces (APIs) for meta(data) access. The RSE team will come together to respond to a specific demand on a temporary basis mandated by the respective TA leads who are themselves experienced RSEs. To speed up development work, the RSE team will use agile project management tools and DevOps environments. Some of these elements are currently set up by GerBI-GMB in the frame of a CZI-funded project and will be provided to the NFDI4BIOIMAGE in-kind. Online communication channels (e.g. Matrix, Zulip) are set up and monitored by TA6, M6.4. All software products will be released in open access repositories with version control to encourage community participation. Documentation is crucial and is supervised by the overall project coordination in TA6 who also takes care of distributing the results from the RSE team in the consortium. The hardware resources needed by the RSE team for development and testing purposes are provided in-kind and free of charge by University of Münster, ALU-FR and de.NBI via TA2. Requests are communicated directly to TA2 staff. Major requests or issues are brought to the attention of the SB via the RSE representative in the CB, if necessary. The NFDI4BIOMAGE will not create a central storage for bioimaging data but aims at federating existing storage capacities in the consortium through the implementation of standards. Data access and searchability across storage sites is a declared goal. In the beginning, project partners rely on their in-house resources or, on a larger scale, existing frameworks for cloud storage, and work towards harmonizing configurations within the consortium. At the same time, we will facilitate access to European public archives by developing and divulging best practices for submission to the BIA hosted by EMBL-EBI.

4 Research Data Management Strategy

The strategy outlined below is based on more than ten years of experience in the GerBI-GMB network, expertise at co-applicant and participant institutions, and activities for systematic assessment of the current situation and challenges: participation and presentation of the NFDI4BIOIMAGE initiative in conferences and workshops (E-Science Days 2021, GerBI community 2020-2021, RDM4mic/OMERO 2019-2021, OME community 2020-2021, ELMI 2021, NFDI 2019-2021, NFDI4Health workshop 2021, RDMO-meeting on DMPs for the NFDI 2021, fdm.nrw jour fixe 2021), exchange with national and international partners, and especially the dedicated 2021 NFDI4BIOIMAGE community survey [19]. The ~25 min long questionnaire addressed knowledge and use of bioimaging methods, method-specific and general data handling, data sharing, and image analysis tools. It was open for participation between June and July 2021 and openly available via our website. It was announced at conferences, via mailing lists and social media. Here, we show excerpts while the full dataset is available online [19].

Regardless of where we discuss with bioimaging specialists, the complexity and heterogeneity of bioimaging data as outlined in section 2 are a consistent, unsolved block to researchers making their bioimaging data FAIR, despite the fact that open data sharing and publication are mandated by various funding agencies[43,44]. Certainly, significant advancements have been made over the past twenty years, but each stage of the standard bioimaging data lifecycle — from the *acquisition* of the data to its *management* in a server or the cloud through the process of *discovery,* and finally to *publication* — still struggles with the lack of a coherent RDM strategy (Figure 4).

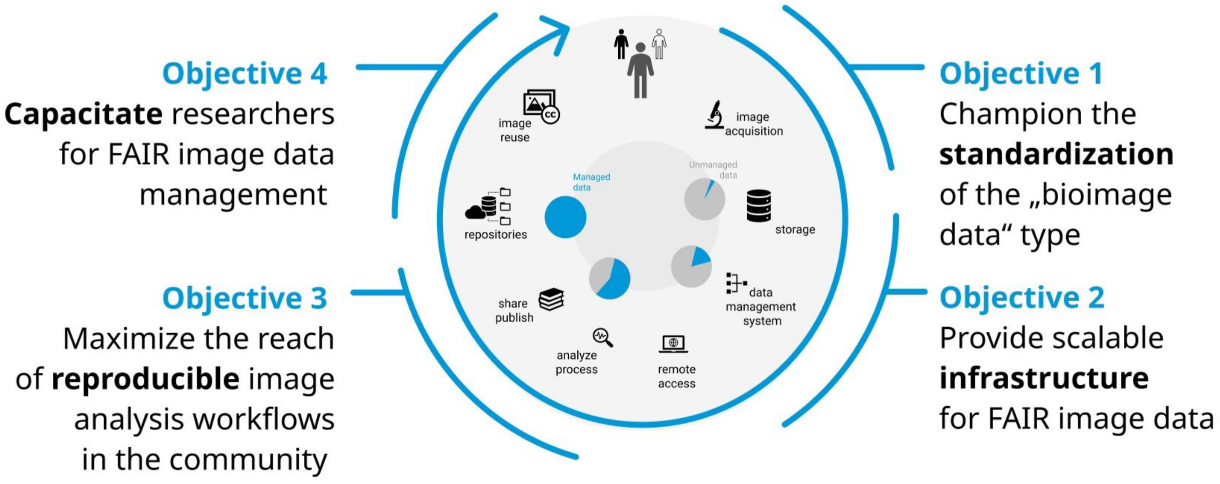

Figure 4: The envisioned bioimaging data lifecycle in which increased management of data leads to more re-use.

---

[43] https://wellcome.ac.uk/grant-funding/guidance/data-software-materials-management-and-sharing-policy

[44] http://data.europa.eu/eli/dir/2019/1024/oj

4.1 State of the art and needs analysis

Acquisition. Most bioimaging acquisition systems including light- and electron-microscopes produce data in formats specific to a particular vendor. Developing a custom format allows vendor' hardware to optimally write large and complex datasets. For downstream consumption, however, these proprietary file formats (PFFs) can only be read on the vendor-provided software. Admittedly, a few vendors have taken steps to make their file formats more accessible, e.g., with free libraries for accessing their formats. Nevertheless, the community largely relies on an open-source library, Bio-Formats [20], for accessing the detected binary data, hardware provenance, as well as metadata of the acquisition process itself. This translation step is a bottleneck to analysis, wasting time on every access. Use of open formats can reduce such overhead. However, existing formats, like OME-TIFF, experience other bottlenecks for higher-dimensional data or cloud-storage. Next-generation file formats (NGFF) seek to overcome these various issues with a cloud-native design. OME-NGFF, developed openly by the bioimaging community [21], offers an image layer on top of NGFF. Within NFDI4BIOIMAGE, TA1 would facilitate the conversion of all bioimaging data to such an open format. The long-term vision, however, would be for all imaging data to be provided directly in an open format with the vendor-specific metadata written for all to read.

Management. A relatively small percentage of bioimaging data today is under management. In the 2021 NFDI4BIOIMAGE community survey, it was found that only ~36% of users had or were planning on installing a data management platform (Figure 5). The reason lies in the scale of bioimaging data that makes centralized management, including storage, viewing, and sharing, appear too costly for the perceived value. Where central management is lacking, ad hoc and suboptimal solutions in the form of isolated instances are often the result. Local custom solutions are built but often fail to find the long-term support needed to keep data available. Several open-source data management servers have been developed and maintained independently of single labs. Projects like OMERO [22, 23], openBIS [24, 25], Cytomine [26] and others provide diverse data management frameworks that can be freely installed and managed locally. According to our

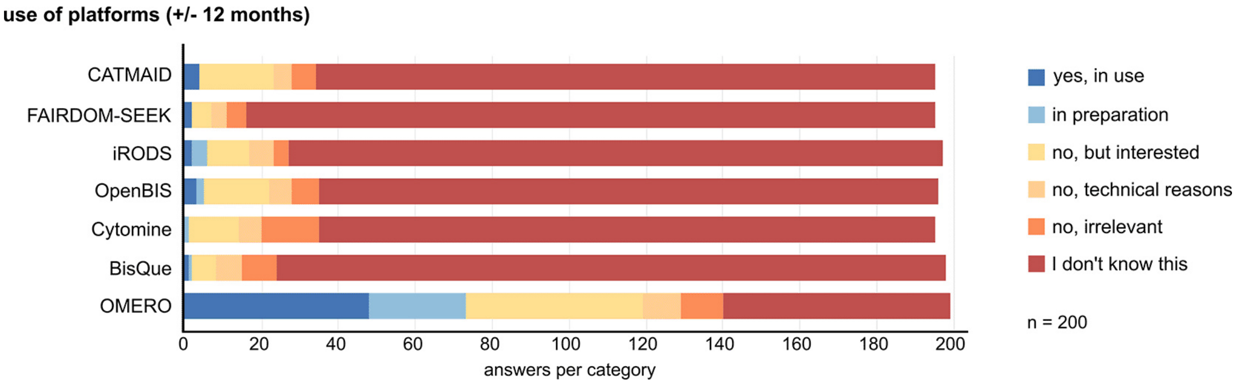

Figure 5: Survey results from 200 participants regarding their use or familiarity with the most prevalent bioimaging RDM systems (Original question: Did you (do you) use the following RDM platforms (+/- 12 months). Taken from [19]

survey and in line with our consortium member's experience, the most frequently used server in our community is OMERO. For this reason, and throughout this initiative, many tasks focus first on integration of OMERO but are open to further systems in time.

Uptake of such data management servers is a persistent issue. Though there are currently hundreds of OMERO installations world-wide, each with tens to hundreds of users, the overall cost of management and maintenance prevents many institutions and labs from using them for storing and sharing large datasets. NFDI4BIOIMAGE's Task Area 2 will strive to make the deployment of data management solutions easier for members of the German bioimaging community.

Discovery. The lack of sharing of bioimage data creates several problems for analysis and interpretation. First, there is a reproducibility crisis in the biomedical sciences which challenges the integrity of modern research [27, 28]. In the absence of public sources of original bioimaging data, funders, universities, the pharmaceutical industry and the wider scientific community cannot easily interrogate or validate the data. However, these data form the basis for peer-reviewed papers that leverage advanced bioimaging technologies. Second, there are limited opportunities for integration of independent datasets, so new modes of discovery through advanced data mining and machine learning are impaired. For a systematic view, however, it is vital that results and discoveries are combined across domains. TA3 will connect and enrich bioimaging and other domain-specific data despite fundamental formatting differences and thereby enable, critically, the searchability of complex research results. Further, as shown in the recent COBA[45] bioimage analysis survey [14] described in TA4, most image analysis is performed using "Open-source point-and- click" software, making it difficult if not impossible to record precisely how an analysis result is produced. Combined with the number of overlapping solutions and undefined output types, the current state of image analysis mirrors the confusion of the input formats. TA4 proposes to provide and standardize workflow tools for a more FAIR representation of bioimaging analysis.

Publication. Finally, combined with the lack of suitable infrastructure and standards (formats, services, workflows), the complexity and size of bioimaging data has slowed the development of centralized repositories, as is common in fields like proteomics and sequencing. A 2015 joint white paper by Euro-BioImaging and ELIXIR[46] laid out a strategy for linking biological image data to biomolecular resources. This initiative led to a call for archives [11] and subsequent funding from the BBSRC, UK to develop the first general-purpose image database, the Image Data Resource (IDR) [29]. Submission to and hosting data in the IDR is costless, but only "reference images" are accepted for publication. The EMBL-EBI is currently expanding this concept to build the BIA with the stated goal of accepting all published bioimaging data. Individual researchers are left with two issues: how to transfer large volumes of data to a centralized location, and how to share images which have not yet been published. Therefore, NFDI4BIOIMAGE will focus on developing the

---

[45] https://openbioimageanalysis.org/

[46] https://elixir-europe.org/news/elixir-and-euro-bioimaging-agree-image-data-strategy

technical capacity for sharing self-hosted images as well as providing the tools and knowledge for submitting such data to the international archives.

The previous development and the current state-of-the-art in bioimage data management, as well as steps to achieve the desired state in the long term are summarized in Table 4.1.

Table 4.1: Previous, current, and prospective phases of bioimaging data management.

Timeline	Acquisition	Management	Discovery	Publication
Previous state	2005: Microscopy data is unreadable except on the proprietary software used for acquisition.	2005: Portable hard-drives used to transport images when they can be extracted from the microscope at all.	2000: Data confined to proprietary systems; closed, unpublished analysis packages are the norm. Images often exported in low-quality formats.	2015: Publishing images with an article is generally considered too difficult in comparison to, e.g., genome sequences.
Current state	**Bio-Formats** reads 150 different file formats; **OME-TIFF** is a widely accepted export format.	**OMERO** and other servers can be used to share images remotely.	Manual workflows, e.g., in **Fiji**, make up a majority of usage.	**IDR** publishes “reference images” of any size for free. (Currently, 0.2 PB)
Related survey results		*35% of respondents use OMERO*	*67% use “Point-and-Click” analysis tools*	
In-progress	Development of a common format, **OME-NGFF**, for cloud-native storage.	**RDM4mic / GerBI-GMB** bring users of OMERO in Germany together.	**BIDS for Microscopy** standard proposed; **JPipe** attempts to capture workflows.	**BioImage Archive** being developed at EBI for “all published data”. Storage will scale to tens of PBs.
Related funding	***CZI-EOSS***[47] *(2022-2023) for support of a cloud format.*	***I3D:bio*** *(2022-2024) for improved OMERO support.*	*GerBI has signed the* ***Wellcome Leap*** *MARFA*[48]*.*	
NFDI Objective	All bioimaging data is structured in open formats produced by or minimally exported from acquisition systems (TA1).	Simplified deployment of RDM servers to increase uptake with standardized APIs to reduce lock-in (TA2).	Searchable index of linked and enriched cross-domain metadata (TA3); provenance of processing steps for reproducible analysis (TA4).	Expand infrastructure offering for pre-publication data and facilitate submission (TA1-3).
Long-term goal	A common, international standard is natively produced by all hardware and can be submitted to all research repositories.	Institutes can readily choose between RDM servers or pure cloud-storage and reasonably expect community tools to work reliably with their data.	A growing body of standardized workflows have been developed and published to community workflow managers.	Federated storage solutions in collaboration with cloud providers like EOSC accelerate sharing of pre-publication data.

---

[47] https://chanzuckerberg.com/eoss/

[48] https://wellcomeleap.org/leap-establishes-global-network/

4.2 Metadata standards

Another finding of the 2021 NFDI4BIOIMAGE community survey quantified a common issue in bioimaging. Although an overall high self-assessed knowledge about the meaning and importance of metadata was reported, in practice, metadata annotation is experienced as a hurdle and a large fraction of participants admitted to using free-text for annotation (Figure 6).

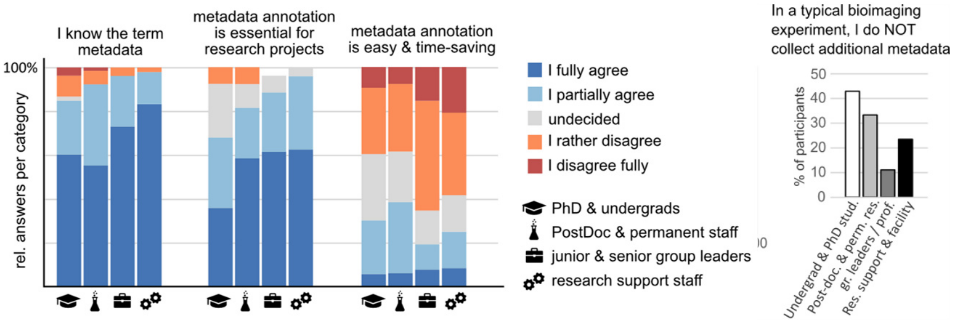

Figure 6: Survey results of 55 undergrads, 66 postdocs and permanent research staff, 27 group leaders, and 48 facility staff regarding their desire for and use of metadata methods. Participants were asked to state their opinion about the presented statements for their field of bioimaging research. Color-coded graphs show the relative fractions of each answer item per group. The right-hand panel shows the relative fraction of persons within these groups stating that they do not collect metadata beyond the automatically generated metadata in a typical bioimaging experiment. Taken from [19]

The survey results are well in line with the local experience at the co-applicant institutions and ongoing collaborative activities at the international level. They show a demand for better guidance and resources for the community to facilitate metadata handling. In particular, the standardization of metadata formats and schemas needs to be improved to foster interoperability and reusability via machine-readable annotations. Next to the sole “grammar” for metadata representation, the evolution of discipline specific vocabularies is an ongoing process and needs alignment within and between research domains. The use of these standardization elements can be boosted by user-friendly, easy-to-use annotation tools.

So far, the *de facto* metadata standard within the bioimaging community has been the OME Data Model that has been continuously maintained and updated to meet evolving needs. The core of this model was developed from 2001 to 2004 and published in 2005 [30]. Updates have been added to handle High Content Screening Data [22] and extensions to the model have been published and used for fluorescence lifetime, optical tomography and other advanced modalities [31]. In 2018-2019, support was added for multi-resolution data, as used in whole slide imaging [32]. Such yearly releases of the OME Data Model were successful in maintaining the utility of the model, but the rate at which new imaging modalities arose has often shown the OME Data Model to be either too inflexible or incomplete. These include advanced forms of light sheet microscopy, super resolution microscopy, and many others. To keep pace with this development, free form annotations using key-value pairs initially loosened the specification of the OME Data Model [23]. This step successfully captured the emerging metadata and stimulated use but did not provide the community a clear method for defining their own vocabularies. The Resource Description Framework (RDF) specified by the World Wide Web Consortium (W3C), a standard for expressing metadata, enables modelling that can faithfully record the design and implementation of bioimage acquisition systems.[49] The OME Data Model was translated to the Web Ontology Language (OWL)[50], a semantic modelling language for RDF, by collaborators at RIKEN [33] and subsequently extended in the 4D Nucleome Project [34] for developing common metadata specifications for super-resolution imaging. Ongoing efforts like BBMRI-ERIC's ISO 23494 standard for biobanking information [35] are eagerly awaiting a compatible representation of the OME metadata.

This work provides an ideal foundation for the specifications that NFDI4BIOIMAGE proposes to deliver. By standardizing on RDF, OME-OWL-based metadata can be combined with other specific vocabularies like for quality control (QC) or domain-specific ontologies, ultimately becoming one part of a larger collection of semantic references, the "NFDI knowledge graph". A richer and more accurate representation of imaging system parameters and protocols will be possible. Usage of these more advanced metadata structures will not be mandatory but will nonetheless be available for technology developers to properly document the acquisition and even analytic protocols that they use for their data allowing for the first time, an accurate representation of the wide range of imaging systems that are used across bioimaging. The intent is that such a package of bioimaging metadata plus binary data can be considered a "Fair Digital Object" (FDO) (Figure 8, TA1). As a self-contained research data bundle, imaging-enabled FDOs can be combined with other FDOs and transferred between institutions despite differing RDM infrastructures via the interoperability tools provided by TA3.

Conversion of existing datasets into such FDOs will allow the community to quickly make use of these developments, but the strong desire is to see developers of bioimaging equipment from academia and industry enable openly annotated datasets directly, thereby reducing the burden on users. Services developed by the technology-focused TAs, especially TA1 and TA3, will further reduce users' burden by simplifying annotation and search. At the same time, a targeted programme of user training in TA5 will strive for broad adoption of these standards, leading to a steady increase in the amount of well-annotated data available to the public.

4.3 Implementation of the FAIR principles and data quality assurance

Quality. A primary motivation for the move to a more open, flexible RDF-based metadata standard was the ability for partnered initiatives to provide their own extensions to the metadata standard. An initial extension by members of the 4DNucleome Image Data Working group has been adopted by BioImaging North America (BINA) and is on route to adoption by the Quality Assessment and Reproducibility for Instruments and Images in Light Microscopy (QUAREP-LiMi) initiative, a global network of imaging experts including commercial vendors and members of

---

[49] https://www.w3.org/TR/rdf-primer/#intro

[50] https://www.w3.org/OWL/

bioimaging infrastructures. This extension of the OME Data Model delivers tiers of metadata for capturing ever more precise quality measures of image acquisitions [36]. With several co-applicants among the founding members of QUAREP-LiMi, NFDI4BIOIMAGE will leverage the expertise, disburse tools, and guidelines as an essential contribution to raising both the quality of image data as well as the quality of the annotations themselves.

Findability. The choice of RDF as the basis for the metadata standard strongly enables many of the FAIR principles (as shown in Table 4.2 below). By having a clear mechanism for attaching identifiers to meta(data), RDF inherently covers both principles F1 and F3. Various storage regimes proposed by NFDI4BIOIMAGE — locally hosted servers, OME-NGFFs stored in the cloud, or deposited data with assigned digital object identifiers (DOIs) — provide a globally unique, remotely accessible identifier which can be used by these mechanisms. TA1 is tasked with defining a collection of rich vocabularies (F2) which the community will be encouraged to use. Finally, TA1, TA2 and TA3 will work together to build searchable resources for consortium metadata (F4), supporting the standard query language of RDF, SPARQL[51].

Accessibility. All NFDI4BIOIMAGE resources are expected to be accessed over the HTTP protocol and all software components emerging from consortium projects will be made available under open source initiative (OSI) approved licenses[52] regarding this as standard in the research community (A1.1). The authentication and authorization procedures will need to be adapted to the provisional cloud storage provided by TA2 (A1.2). Here, we will follow the AAI patterns defined by partnered consortia (NFDI4Immuno). Every effort will be made to keep metadata accessible in the SPARQL endpoints even if the data is no longer available (A2). Further, we will encourage all participants to obtain DOIs for their images which is the strongest guarantee that the metadata will be permanently retrievable.

Interoperability. Metadata interoperability is again assured through the choice of RDF as the syntax language (I1). The interoperability of the binary image data, however, is a critical issue for the bioimaging community. Several separate APIs are used across the RDM servers available in the community. Each enables *accessibility* for users of the platform, but not *interoperability*, i.e., clients can likely access only one particular platform. This limitation arises from the fact that the most common image exchange format, TIFF, is not fit for purpose when dealing with large, n-dimensional data. And though there are great n-dimensional binary formats for local or cluster access, remote HTTP-access remains difficult. We therefore propose the use of the NGFF format which is compatible with HTTP as well as cloud protocols (S3, GCS, etc.) working within TA1 to make the binary format as interoperable as the metadata itself and via the TA6 Data Format Ambassador to propose the format to the wider NFDI. All metadata vocabularies directly suggested by NFDI4BIOIMAGE will be RDF based and therefore compatible with the core

---

[51] https://www.w3.org/TR/sparql11-overview/

[52] https://opensource.org/licenses/

metadata (I2). Where vocabulary boundaries must be bridged either to other domains or to other non-RDF frameworks, conversion tools will be developed by TA3 (I3).

Re-use. The goal, of course, of all other FAIR activities is to enable and encourage the re-use of bioimaging data, both that made available within the NFDI4BIOIMAGE consortium but also at large. Though no licensing requirements are foreseen as a part of the initiative, guidelines developed by TA1 will be distributed by TA5 and TA6 to begin a cultural change towards the sharing of more data, in line with guidelines of the international image resources like IDR and BIA which are strongly preferring Creative Commons licenses (CC-0 and CC-BY) for data submissions (R1.1). With a consistent framework for incorporating metadata, each Task Area (1-4) will ensure that provenance for actions related to acquisition or discovery are recorded (R1.2). Finally, though the bioimaging community is still in the process of identifying potential standards, NFDI4BIOIMAGE will work to support the most prominent developments in the community (R1.3). The current forerunner is the Recommended Metadata for Biological Images (REMBI, [6]), co-authored by members of GerBI-GMB, and which heavily makes use of the OME Data Model.

Table 4.2: Many of the FAIR principles are covered by the use of RDF as the basis for the metadata. TAs are responsible where gaps persist.

Responsibility for FAIR principle	RDF	Task Area					
		1	2	3	4	5	6
F1. (Meta)data are assigned a globally unique and persistent identifier	M	D	D				
F2. Data are described with rich metadata (defined by R1 below)		✓		✓			
F3. Metadata clearly and explicitly include the identifier of the data they describe		✓					
F4. (Meta)data are registered or indexed in a searchable resource		✓	✓	✓			
A1. (Meta)data are retrievable by their identifier using a standardized communications protocol	M	D					
A1.1 The protocol is open, free, and universally implementable	M	D					
A1.2 The protocol allows for an authentication and authorization procedure, where necessary	M		D				
A2. Metadata are accessible, even when the data are no longer available				✓			
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	M	D					
I2. (Meta)data use vocabularies that follow FAIR principles		✓		✓	✓		
I3. (Meta)data include qualified references to other (meta)data				✓	✓		
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			✓	✓	✓	✓	✓

Symbols: ✓: both metadata & data, M: metadata only, D: data only

4.4 Services provided by the consortium

The spectrum of services that we foresee targets user groups from a typical, non-managed state for their bioimaging data to large managed installations, a reflection of the members of the consortium itself. NFDI4BIOIMAGE's objective is to capacitate users in their own RDM rather than to provide them with ready-to-go solutions. In particular, we will not provide a long-term centralized storage. Where the technical infrastructure is lacking, provisional resources will be offered via TA2 while the wider scientific community together with funding agencies identifies appropriate mechanisms and solutions for the storage of data at bioimaging scale.

Community support. For those new to RDM, detailed guidelines and training modules for the proper management of the data will be the first point-of-call, curated into a searchable index (TA5). This approach in TA5 continues and expands on work by German BioImaging to provide these materials to imaging core facilities and synergizes with the cross-discipline RDMkit[53] resource developed by ELIXIR Europe. For emerging issues, the Help Desk provided by TA5 and TA6 will offer personalized support, e.g., via moderating requests in established community exchange platforms like image.sc. The Data Stewards of NFDI4BIOIMAGE will handle persistent, more complex issues including on-premise support if necessary to, e.g., support a dialogue with institutional stakeholders including research and IT.

Usable (meta)data tools. For researchers farther on the path to RDM, the next barrier is the formatting of their bioimaging data. Community members have built services for the conversion of proprietary formats into a common format (Bio-Formats, see above). Though a common specification lies at the heart of making bioimaging data and components FAIR, experience shows that without tools for simplified annotation and curation of bioimaging data, the barrier to entry will remain such that data will not be made FAIR. TA1 and TA3 will build and host usable tools to ease the preparation as well as discovery of well-annotated datasets, building on but simplifying the use of RDF's powerful SPARQL query language. Instances of these services will be made centrally available or can additionally be deployed separately by participants, and the optimal balance between local and shared resources will be developed over the course of the project.

Provisional Cloud. Interested research groups with properly curated metadata may still be stymied in their efforts to properly manage data if sufficient IT infrastructure is not available locally. NFDI4BIOIMAGE partners have allocated several hundred terabytes of provisional storage to accelerate RDM activities. This storage will equally be available to all participants throughout the funded period (TA2). This infrastructure-as-a-service will ease the installation of data management platforms. Towards the end of the grant, an evaluation of available strategies for the hosted data, including those listed below, will be performed with data owners.

Data migration. Critical to prevent data from being locked into any one location is the ability to migrate data simply and efficiently between locations. Users of the provisional cloud resource who have successfully acquired institutional resources can migrate their data to local storage. Users with institutional resources who must move to a new institution can migrate their entire data collection, overcoming one of the primary downsides of systems like OMERO. The cloud storage provided by TA2 can be used as an exchange mechanism with bundles of bioimaging data temporarily migrated there and shared via standard S3 mechanisms, fostering sharing early in the bioimaging data lifecycle. The completeness and re-identifiability of the exchanged

---

[53] https://rdmkit.elixir-europe.org

information as defined by TA1 frees users from individual implementations and creates a single, FAIR bioimaging ecosystem. For multimodal researchers with a diversity of datatypes beyond bioimaging data, the migrated data can be integrated into a single data structure defined by TA3 for all research outputs.

Streamlined data submission. As an alternative to migration within the consortium, services will be provided for the transfer of previously curated bioimaging data to a central archive. In concert with EBI's BIA, users without an institutional archive for bioimage data wishing to preserve their published data will be encouraged to upload them centrally via tools provided by TA3. With a low-overhead workflow from the curation of pre-publication data through to publication in a central archive, NFDI4BIOIMAGE data will serve as a model for encouraging others to share early and often.

Distributed analysis. With data remotely accessible and migratable, whether in the central archives or on one of the distributed storage mechanisms within the consortium, distributed analysis becomes not only a feasible but a powerful tool at researchers' disposal. Workflow management systems like Galaxy can optionally access data remotely or download it to a temporary local cache based on the size and frequency of use. Developments in the Galaxy community have integrated imaging data and workflows more centrally into a dedicated platform (Galaxy Imaging)[54].

Long-term vision. Though our goal with NFDI4BIOIMAGE is to facilitate the sharing of more consistently prepared, pre-publication data, it is clear that the long-term vision must be larger. As with many NFDI consortia, we are closely following the developments of EOSC and in particular EOSC-Life regarding the future of large-scale cloud computing resources. Consortium members at EMBL have several running projects for the development of features for OME-NGFF, e.g., specifically related to the integration of imaging data into the COVID Data Portal[55]. As participants, whether locally or via EOSC, begin to have more cloud storage, we foresee a federation of such data. The services and the open-source software that underlies them will have prepared the community for a new scale of sharing bioimaging data.

---

[54] https://imaging.usegalaxy.eu

[55] https://www.covid19dataportal.org/imaging

5 Work Programme

Making imaging data FAIR and thus enabling real benefit for the progress of research as a whole is the goal of each single measure, service, tool, teaching and networking activity we propose in our work programme. As described in section 3, the team of NFDI4BIOIMAGE is composed of experts in the field of bioimaging and related fields of expertise. These experts agreed on a common work programme of NFDI4BIOIMAGE, a concerted action of six Task Areas (Figure 7).

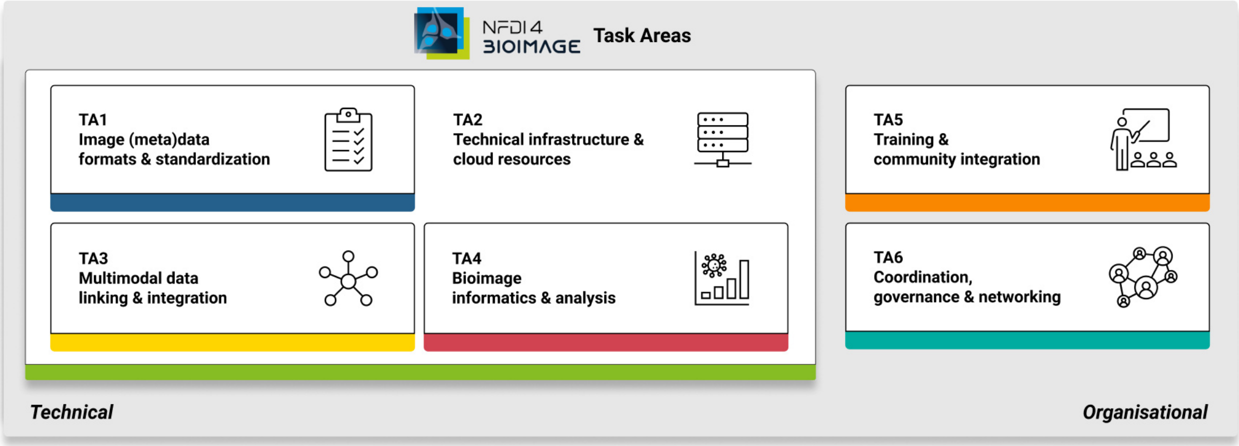

Figure 7: Schematic representation of the Task Area structure, showing the dependency of all technical task areas on the technical infrastructure that will be provided.

In **Task Area 1** “Image (meta)data formats and standardization” the harmonization and further development of data formats and metadata standards is a central measure as this is key to guarantee interoperability with other data types and within distinct data management systems, requiring both machine-actionable and human-readable annotations. A recently developing international activity of the bioimaging community is the development of next generation file formats (NGFF) for imaging data, in which we take part. NGFF are promising candidates for a globally harmonized data format enabling among others cloud-readiness for large-scale image data. Consequently, this topic has the potential for becoming a cross-cutting topic within the NFDI. We consider this TA as the conceptually most advanced and innovative of our work programme from which we expect, in the course of the project, input that may serve as guidance for the other task areas. **Task Area 2** will provide “Technical infrastructure and cloud resources” and will serve as a central enabler for the consortium and the community, incl. “sandbox” environments for testing of new tools and services. TA2 will provide access to image analysis workflows that require high compute resources in collaboration with TA4. **Task Area 3** “Multimodal data linking and integration” will work on the integration of image data with other data types and the interfacing of image-data-specific data management systems with more generic data management environments. We moreover integrated specific cross-consortial collaborations on common data structures into dedicated measures within TA3. **Task Area 4** “Bioimage informatics and analysis” aims to foster reproducibility of image analysis workflows, in particular by making popular individual tools more interoperable and to make state-of-the-art method deployment accessible to the whole bioimaging community which will be done in collaboration with TA2. TA4 will also collaborate with Task Area 5 on training materials to support the successful application of new and advanced analysis methods. **Task Area 5** "Training & community integration" will in general focus on in-depth training and exchange with the community by providing access to a cross-platform database for resources and training materials as well as in-person events. They also aim to prepare training materials for the tools and services that will be provided by the other Task Areas. The Data Steward team will be led by TA5 for community support on all levels. The overall project coordination will be done by **Task Area 6** "Coordination, governance and networking". This includes the executive management of the whole project as well as all outreach and networking activities.

The people and the tasks of our work programme are highly intertwined and function as an ensemble. We have implemented this strong cooperative spirit in the work programme at several levels: first, two co-spokesperson of complementary expertise share responsibility for each TA; second, different sites contribute together staff capacity for solving the same specific task. Thus, tasks are addressed from different angles. This principle is realized both at the level of the DaSts and at the level of the PTPs. As an example, staff assigned to UOS, GerBI-GMB, DKFZ, LIN, HKI, EMBL, UKON, UFZ and JGU will contribute together to the specific demands of HCS. On the other hand, the FTE assigned e.g. to UFZ, a PTP site with broad expertise in bioimage RDM, will work primarily in TA3, but devote part of its time also to tasks in TA1 and TA4. Similarly, DaSts, distributed at several sites, will work collaboratively and according to their expertise to answer user-specific requests (TA5). Thanks to the modern opportunities for remote working, we believe that this model will be highly feasible and will prevent capacity to be absorbed by contingent requirements arising at the workplace of any NFDI4BIOIMAGE member.

Table 5.0: Overview of task areas

Task Area	Measures	Responsible Co-spokesperson(s)
TA1: 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	S. Kunis, S. Weidtkamp-Peters, *J. Moore*
TA2: 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	B. Grüning, M. Blank-Burian
TA3: Multimodal data linking and integration	M3.1: Automatic sample tracking and sample metadata acquisition M3.2: Specify a data structure for multimodal research projects M3.3: Combine data from different (imaging) modalities M3.4: Integrate imaging data and metadata with other data types and omics data in a multimodal graph database (MUGDAT) M3.5: Solution for archiving and publishing multimodal data	W. Zuschratter, J.P. Mallm, *T. Stöter*
TA4: Bioimage informatics and analysis	M4.1 Formalize best practice standards for users and developers microscopy image analysis tools M4.2 Image Analysis Tool Integration M4.3 Simplify the transition to cloud-based computing environments	A. Kreshuk, M.T. Figge, *C. Tischer*

TA5: 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	R. Haase, T. Zobel
TA 6: Coordination, governance and networking	M6.1 Project management, administration and coordination M6.2 Governance M6.3 Data Stewardship commissioning and flexible assignments M6.4 Communication – digital tools M6.5 Communication – groups M6.6 Communication – access M6.7 Embedding into the NFDI and coordination of activities on cross-cutting topics	S. Weidtkamp-Peters, E. May

Throughout all Task Areas, deliverables or services are shown as “DS” and milestones as “MS” (see sec 5.7).

Note: The listed co-spokespersons on page 9/10 are the mandated representatives of their affiliated Co-applicant Institutions. Additionally, the following persons are acting Task Area leaders:

J. Moore (GerBI-GMB), Task Area 1

T. Stöter (LIN), Task Area 3

C. Tischer (EMBL), Task Area 4

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

Today, bioimaging data is most frequently delivered to users in formats and with metadata as defined by the commercial vendors of acquisition systems. This means that visualization and analysis are often tied to manufacturer-specific software, complicating the combination of data from multiple sources. Community members regularly must achieve their research goals by time-consuming translation of their imaging data into a re-usable format [21]. Combined with the size and dimensionality that bioimaging data can achieve, the typical local file download of genomic data with text-based FASTQ files becomes prohibitive, leading to the common characterization of imaging data as difficult to archive. In the future, however, users should be enabled to remotely and interactively visualize their large data volumes, with ubiquitous tools for sharing and re-analysis.

The goal of TA1 is to define common mechanisms for such data exchange, starting from a consolidated concept for a common data format. This format should ensure that acquired bioimage data (multidimensional spatial measurements) as well as the analytical results can be made shareable, linkable, browsable, re-usable, and archivable. This framework and the related measures of this task area are balanced to meet the needs of NFDI4BIOIMAGE, the imaging component of other NFDI initiatives, as well as the global bioimaging landscape to maximize the long-term value to all involved.

In its first measure, TA1 will therefore focus on the development of a common specification of "FAIR Image Objects" (FAIR-IO) for the bioimaging community (M1.1) combining knowledge-graph enabling metadata with high-dimensional binary data (Figure 8). User-friendly tools and guidelines for annotating and browsing annotated data (M1.2) will foster re-use and sharing while services for cloud-native image hosting will ensure a healthy ecosystem around the core binary formats (M1.3). M1.4 aims to synchronize this community specification with the wider bioimaging community via a "request for comment" (RFC[56]) process. The value of this framework will be demonstrated via implementation of several use cases in collaboration with participants and community members (M1.5). The work of this TA involves 9 PTPs as well as several CUCs, providing a basis both for defining a common and FAIR bioimaging framework and to have it adopted in practice for bioimaging research.

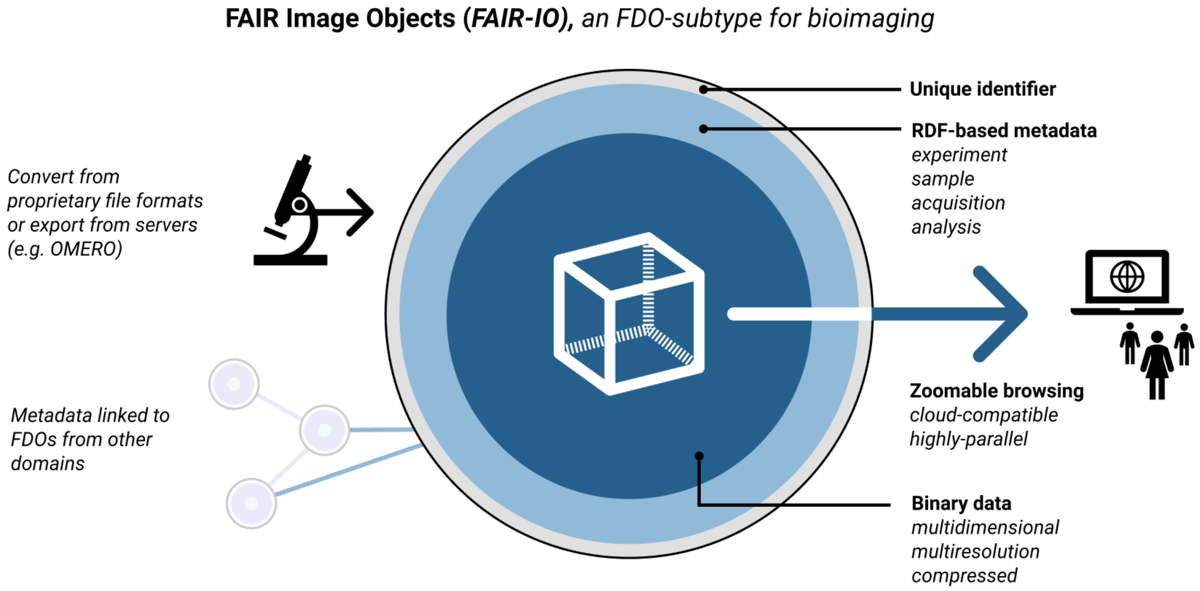

Figure 8: A FAIR-IO bundle combines the necessary acquisition and provenance metadata together with multi-resolution, chunked binary data in a single cloud-compatible format for simplified sharing and re-use. Available at: NFDI4BIOIMAGE Consortium. (2021). Section 5.1 "Task Area 1: Image (meta)data formats and standardization)" Figure 8 (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.7394675; CC-BY 4.0

5.1.1 *Measure 1.1: Define an FDO subtype, "FAIR Image Objects"*

Contributions: UOS (Kunis), GerBI-GMB (Weidtkamp-Peters, Moore), TA3, TA4, TA5, TA6

Goal: Increase the FAIRness of bioimaging data across the initiative with a usable formalization for metadata combined with n-dimensional images data

This foundational TA1 measure will define a bioimaging specialization of FDO (FAIR Data Object). Interoperable vocabularies covering a breadth of bioimaging applications combined with a cloud-compatible binary data will be defined as FAIR Image Objects (FAIR-IO). Extensible metadata specifications will be arranged in increasing levels of expressiveness without the need for the

---

[56] https://en.wikipedia.org/wiki/Request_for_Comments

development-intensive static schemas which the bioimaging community currently relies on. Starting with the publication of OME-OWL (described in sec 4.2), the framework will strive to capture the bioimage metadata in OME-OWL as well as other ontologies and RDF-based vocabularies in a web-accessible format like JSON-LD[57] ("JSON for linking data"). Through standard RDF mechanisms, datasets can be enriched with links to metadata from other domains (TA3, M3.1) and metadata can be captured for image analysis (in conjunction with TA4). This is critical for the goal of enabling community-wide searchable knowledge graphs, while maintaining developer friendliness for embedding in open file formats produced by the community (OME [20, 32]) in M1.3. We achieve the cross-linking of data in the graph-based concept by defining so-called crosswalks for vocabularies, which act as a bridge to other domains and enable the corresponding links to be made. At the same time, binary formats will be identified for the FAIR storage of large, multi-dimensional bioimaging measurements. The OME-NGFF format (also described in sec 4.2) is a primary candidate and will be reviewed for compatibility with the FAIR-IO definition. By identifying appropriate APIs, RDM services like OMERO and Galaxy (TA2 and TA3) can be made readily FAIR-IO-compatible. Further proposals for specific FAIR-IO implementations can be submitted via the process defined in M1.4 (FAIR-IO RFC). Finally, in order to make use of the same storage mechanisms, data producers require recommendations and tools for the conversion of proprietary formats and metadata (TA5). A balance must be struck between the use of vendor software and the benefit of making a dataset available to the wider community. The ultimate target of these developments is the use of an open and FAIR format by all microscopy vendors. Therefore, making use of the communication mechanisms provided by TA6 (e.g., M6.5, M6.7) and engaging with commercial partners of our participants and partners like UHD, the EuBI Industry Board and QUAREP-LiMi, we promote collaboration with industry to gather the requirements to meet this goal.

M1.1		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 1.1.1	Prepare roadmap with initial vocabularies for internal & external communication	■									
DS 1.1.2	Review and release OME-OWL ontology	■									
DS 1.1.3	Establish OME-NGFF compatibility with FAIR-IO		■	■							
DS 1.1.4	Evaluation of FAIR-IO API support for participating RDM servers			■	■						
DS 1.1.5	Guidelines for the export of meta(data)				■						
MS1-TA1	Draft the framework as a whitepaper (incl. API, recommendation, etc.)				■						

5.1.2 *Measure 1.2: Increase usability of RDF metadata for bioimaging*

Contributions: UOS (Kunis), MPIEB, UoC, TA3, TA4, TA5, TA6

Goal: Foster usage and user acceptance of metadata linkage with FAIR-IO

NFDI4BIOIMAGE as most other NFDI consortia is committed to the development of a domain-spanning knowledge-graph, based on or in the spirit of the W3C's RDF. Despite the advantages

---

[57] https://json-ld.org/

of having metadata as an extensible and FAIR graph of facts, most bioimaging researchers currently lack experience working with such structures. To this end, NFDI4BIOIMAGE is committed to developing guidelines, tools, and services to encourage scientists from a new domain to adopt knowledge graphs (M6.7). The primary output of this measure will be the development of a standalone and componentized service (standalone as well as plugin for RDM systems) to simplify and facilitate users' interaction with this technical concept. Users will be guided in the easy-to-use annotation of datasets. Subsequently, datasets can be visualized in context with users navigating the graph as they would navigate webpages. Datasets can also be validated against community-specified metadata criteria such as emerging minimum standards (e.g., REMBI, [6]). Such datasets can be enriched with links to metadata from other domains (TA3) as well as capture metadata relating to image analysis (related to M4.1). The more connected the data, the easier it is to identify related data. Intelligent forms will be available that are automatically generated from one or more ontologies. These services will build on existing applications such as MDEmic [7] (UOS) and OMERO.forms (CUC @MPIEB). For cross-standard compatibility, we will integrate an ISA service to convert the RDF-based metadata into an ISA format on demand (UoC). This Measure works according to the agile principles[58], so that the development process can take up and implement new requirements from the community and from the NFDI sections (collected in M6.7). In addition, tailored guidelines with training concepts and illustrative examples for use and implementation are developed in cooperation with TA5.

M1.2		Y1		Y2		Y3		Y4		Y5	
DS 1.2.1	Deliver annotation service (e.g., OMERO.forms, MDEmic)			▨	▨	▨	▨				
DS 1.2.2	Integrate feedback from the NFDI sections	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
DS 1.2.3	Enabling subset validation/classification with concrete examples (e.g., REMBI)							▨	▨		
DS 1.2.4	Integrate ISA into services for cross-standard compatibility (e.g., ARC)							▨			

5.1.3 *Measure 1.3: Enable cloud-native image access*

Contributions: GerBI-GMB (Moore, Weidtkamp-Peters), UOS (Kunis), UGOE, HAWK, UHD, TA2, TA3, TA6

Goal: Provide chunked, multiresolution, cloud-compatible image data storage and sharing

The power of imaging data lies in the spatial-context of acquired measurements which is also a burden. Even with performant, non-cloud binary formats like HDF5[59], users can be left downloading TBs of monolithic archive files to preview an image. Without special infrastructure, the complexity and size of the data is not conducive to sharing. Several strategies like the OMERO server exist and are used by NFDI4BIOIMAGE consortium members to make data available remotely. These, however, require dedicated servers and personnel. With the rise of cloud computing and storage, the prospect of low-overhead sharing arises. To make this possible, a

---

[58] https://www.agilealliance.org/agile101/12-principles-behind-the-agile-manifesto/

[59] https://en.wikipedia.org/wiki/Hierarchical_Data_Format

cloud format is needed for the retrieval of portions of a multi-dimensional image at user-defined resolution levels. To this end, OME-NGFF was based on a cross-platform data format, Zarr[60], “for the storage of chunked, compressed, N-dimensional arrays.” In Zarr, large data volumes are split into individual, ordered “chunks” which are stored separately in the cloud, simplifying parallel retrieval. Chunks can be trivially found in the cloud by their index, removing the need for any server infrastructure.

This measure will focus on making this cloud-native format usable within and beyond NFDI4BIOIMAGE. The interactive viewing of images from community image-analysis platforms (napari in Python [37], the MoBIE plugin for Fiji in Java [38], and Vitessce [39]) in JavaScript) will be implemented for all participant use cases. Especially relevant is the ability to share links to large (tera- to petabyte scale) images with the ability to view them at progressively greater resolutions without the need to download the entire image (TA2). Other new data types within NFDI4BIOIMAGE will additionally require the modification of these ecosystems to support their specification and visualization. For example, the storage of images with lists of values rather than single numerical values per pixel (“vector-based images”) as are common e.g., in spectroscopy or fluorescence lifetime imaging (FLIM) will be covered by this measure with support from HAWK. Through the office of the “Data Formats Ambassador” (M6.7), this format will be proposed as a cross-cutting concern to other NFDI consortia and we expect further new data types to be suggested, which will be handled as possible in this measure. A critical component of moving the community towards a common format are the conversion tools to take a broad-spectrum of input formats and convert them as entirely as possible into a scalable format like OME-NGFF (GerBI-GMB, UHD, UGOE, TA3). Participants with hardware experience will lead the way simplifying integration of later commercial hardware. With the diversity of input formats, users will be supported with working converters or help on implementing local conversion strategies as possible. Finally, as more groups including vendors and users alike begin producing the common format natively, the stability and performance of the overall imaging standard must be maintained to make it a viable option for cost-critical acquisition and long-term storage. We will define a technology compatibility kit (TCK) including benchmarks to ensure that change requests do not degrade the trust in the format.

M1.3		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 1.3.1	Establish recommendations for the viewing of cloud-based data		▨	▨							
DS 1.3.2	Extend format and viewers tools for vector images				▨	▨					
DS 1.3.3	Integrate feedback from NFDI sections	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
DS 1.3.4	Provide converters for consortium members			▨	▨	▨	▨				
DS 1.3.5	First release of TCK for conformance testing							▨	▨		

---

[60] https://zarr.readthedocs.io/en/stable/

5.1.4 *Measure 1.4: Formalize FAIR-IO RFC process*

Contributions: GerBI-GMB (Moore, Weidtkamp-Peters), UOS (Kunis), TA6

Goal: Establish a process for the consensual evolution of maintainable bioimaging technical specifications, balancing industrial and research concerns

The core FAIR-IO concept (M1.1) represents the product of a global community of bioimaging specialists. Ongoing developments of the wider community need to be recognized and collaboration facilitated towards bioimaging's common goal of long-term maintainable data. To this end, a "Request for Comment" (RFC) process will be defined and submitted to the community for review, modified based on feedback, and finally formalized by bioimaging bodies such as GBI's image data working group, in which NFDI4BIOIMAGE members play an active role. A prototypical community process has evolved for the development of the OME-NGFF [21], and as such will need to be coordinated with the Open Microscopy Environment (OME). Critical components of a successful process will include technical review periods and a transparent voting mechanism that clearly defines roles for industrial, academic, and open-source participants. The goal of this process will be to balance the needs of acquisition software (simply put, fast persistence of large volumes of data) and those of data consumers (long-term storage, frequent access, FAIR standards). Requirement gathering will be the first step. Requirements should be captured from external actors concerning specification additions and modifications but equally concerning the process of M1.4 itself. These requirements will provide an initial roadmap with both the specifications outlined in M1.1 as well as those surveyed from the community. Beyond these technical interactions, there is a clear potential for broadening the scope of both FAIR-IO and the RFC process to other communities. There are several other communities also working on the specification of imaging formats, e.g., climate/geoscience, neuroimaging, and health. Frameworks with their own processes exist in several of these (respectively, CfConventions, BIDS, DICOM). Together with the Data Formats Ambassador (M6.7), this measure will reach out to other image-based communities to evaluate the potential for a common base specification covering a variety of n-dimensional binary data.

M1.4		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 1.4.1	Call for proposals: community & vendor requirements			▨							
DS 1.4.2	Formulate specification roadmap				▨						
DS 1.4.3	Request for comments on the process				▨						
DS 1.4.4	(Online) workshop with other imaging communities & publication of the findings					▨					
MS2-TA1	Release of process document						▨				

5.1.5 *Measure 1.5: Implement community requirements and use cases*

Contributions: UOS (Kunis), GerBI-GMB (Moore, Weidtkamp-Peters), UoC, HAWK, INP, MPIEB, IPHT, UGOE, UFZ, UHD, TA3, TA5, TA6

Goal: Integrate use-case-based requirements from the community

The FAIR-IO concept will be extended to include the specific requirements of the community and applied on a diversity of use cases from current and future community sites. The extension and application is based on preliminary work from M1.1, M1.2, M1.3, and M1.4, and is a valuable test of the usability, adaptability and extensibility of the concept. In this measure, we have collected the participant use cases which will be directly funded for testing the FAIR-IO concept. These will be carried out separately as functionality becomes available and participants have their local RDM positions filled. We also support the integration of FAIR-IO in software packages that are currently widely used in the community. Here we are implementing the requirements for visualization in viewers with regard to vector-based data (HAWK). In addition, in collaboration with TA3 we are implementing FAIR-IO concepts into the image data management platform OMERO according to the guidelines formulated in M1.1.

The main part of use cases covers the development of sub-domain specific meta(data). Sustainable integration will require specifications of syntactic and semantic structures as well as sub-domain specific guidelines for metadata annotations and for data conversions and storage recommendations. At some institutions, the appropriate integration of bioimage data into other domain specific structures is being addressed (1.1.5.1, 1.1.5.4, 1.1.5.8). Wider use cases (1.1.5.5, 1.1.5.6, 1.1.5.9) attempt to implement the FAIR-IO concepts locally, with participants engaging in specific development tasks of other measures for their own needs.

Table 5.1.1 List of use cases at participant institutions that will be implemented

Use case	Participant	Domain / Consortium	Request
1.1.5.1	INP	Plasma science / NFDI4Phys	plasma-mds integration; cross-linking; FAIR-IO integration
1.1.5.2	IPHT	Biophotonics / NFDI4Chem	Integration of technical metadata from newly developed microscopy setups
1.1.5.3	MPIEB	Evolutionary biology	Metadata Microbial population, evolutionary cell biology
1.1.5.4	UFZ	High-content screening / NFDI4Biodiversity, NFDI4Microbiota	cross-linking; FAIR-IO integration
1.1.5.5	UGOE	Multiscale Bioimaging	RDM for FLAMINGO platform (light sheet)
1.1.5.6	UHD	Neuroimaging	RDM; Conversions
1.1.5.7	HAWK	Medical imaging	Vector-based images integration
1.1.5.8	UoC	DataPLANT	cross-linking; FAIR-IO integration
1.1.5.9	JGU	High-content screening	Standardization of cell painting assay data
1.1.5.10	FZJ	High throughput scanning of whole human brain sections	Establishing compatibility of a large-scale light microscopy repository with an OMERO instance

INP as an important cooperation partner in the fields of plasma science will develop an ontology for plasma imaging based on the Plasma-MDS metadata schema, enabling exchange of metadata with OMERO and integration into elabFTW. The coupling of OMERO with INPTDAT (TA3, M3.5) and of the respective metadata standards and ontologies could also highlight the benefit of integrating domain-specific (plasma) and resource-specific (bioimage) metadata and systems, particularly regarding the re-usability of published datasets for ongoing research. This requirement is formulated in particular by NFDI4Phys for the bioimage-centred part in biological physics and the application-oriented research in low-temperature plasma physics. The ongoing QPTDat project, which is working on a knowledge graph for plasma technology, addresses another common point of interest. This knowledge graph is to be used here, and further development is planned within the framework of NFDI4Phys.

IPHT will develop hardware-specific metadata in a FAIR-IO-compatible format (ontology).

MPIEB will develop an ISA-based controlled vocabulary for the description of microbial population biology and evolutionary cell biology, including attributes crucial for the proper interpretation of data collected in an evolutionary science context, such as time-series microscopy data from growing cell or bacteria cultures, fitness-assays, genotype-phenotype association studies, and colony morphologies. In addition, existing metadata standards from within MPIEB institute and from other Max-Planck partners will be collected for integration into the FAIR-IO framework, facilitating the re-use of existing OMERO tools and further data cross-linkages.

UFZ will structure into FAIR-IO bundles high-content screening data currently stored in OMERO and cross-linked with lab sample management systems. Due to the parallel nature of NGFF, such bundles are ideal for workflows for analytical high content systems (multispectral imaging flow cytometry, imaging of small organisms and imaging of bioassays). With associated metadata and research results, the bundles can then be transferred to a repository if not submitted simultaneously to the IDR/BIA and similar resources from NFDI4Biodiversity and NFDI4Microbiota.

UGOE participants have a strong expertise in light-sheet microscopy with a dedicated hardware platform, Flamingo. This platform should be made compatible with the NFDI4BIOIMAGE specifications, and as a result more participants should be able to make use of the platform which would boost the acceptance of Flamingo as a shippable light sheet microscope. Where a local installation of such a combined hardware/software solution is not possible, remote-control access to the microscope will be established for collaboration between light microscopists and data scientists.

For **UHD**, as with many institutions, the current situation is such that without a common solution, a host of solutions including file format conversion, metadata standardization, and sharing tools would need to be developed locally. This allows UHD to function as a broad use case for the use of FAIR-IO. The presence of a RDM role at an institutional level — covering training, data curation, and data management — will serve as a bridge to NFDI4BIOIMAGE as a whole, preventing insular solutions to develop. A focus on full conversion of particular formats like Nikon's ND2 into FAIR-IO will solidify the current neuroimaging pipeline, strengthen the embedding of NFDI4BIOIMAGE meta(data) into other standards like BIDS, and involve commercial vendors more deeply.

HAWK will help define formats for describing vector-based formats for hyper-spectral imaging (HSI) which currently lacks a standard. Hyper-, or multi-, spectral imaging collects more than one value per pixel of an image. Storing this data poses a unique challenge, expanding the normal understanding of the common image data type. Beyond the binary storage and visualization aspect, metadata must be specified to explain the measurement quantities (e.g., for the manual macroscopic annotations of regions on the physical glass slides, and for the cell-level or region annotations performed with software on the digitalized slides) as well as the quality of the acquired data. These activities will enable the evaluation of segmentation and classification algorithms by researchers as well as an objective comparison of the performance of different methods.

UoC will work to further enable the embedding of FAIR-IO structures in a general context and to create cross-domain contexts as well as develop FAIR-IO export and access conventions for data management platforms to foster the exchange of FAIR-IO with third-party tools such as laboratory notebooks and laboratory sample management. These conventions are also needed by TA3 to integrate FAIR-IO in the used directory structures (M3.2). In particular, the transformation into ISA-formatted, ontology-linked metadata describing the raw and processed data stored in ARC will be implemented in this collaboration (M1.2). This is a highly requested functionality especially from the DataPLANT consortium, which are integrated in this collaborative work.

JGU will focus on the harmonization of public data sources towards a standardized representation of cell painting assay data. Datasets will be curated into FAIR-IO-compatible formats following the existing cell painting studies in IDR (idr0016, idr0036, etc.) and used for training deep learning models. Cross-link cell painting assay data to other open data sources such as ChEMBL[61] and PubChem[62] is being pursued. A central resource for looking up publicly available data cross-linked to other data sources would, for example, support the drug discovery process. Curated datasets may be submitted to the IDR as part of the project, which is pursued in collaboration with the DeBioData consortium.

FZJ is establishing a large-scale distributed repository for high-throughput scanning of whole human brain sections with light microscopes, which is implemented on the supercomputing infrastructure at the Jülich Supercomputing Center (JSC) and will expose parts of the data as public releases to the EBRAINS research infrastructure for remote visualization and download, as well as to the Canadian Open Neuroscience Platform (CONP). In the frame of NFDI4BIOIMAGE, FZJ aims to adopt folder structures and image formats for maximal compatibility with BIDS-Microscopy as well as the specifications established by this consortium.

---

[61] https://www.ebi.ac.uk/chembl/

[62] https://pubchem.ncbi.nlm.nih.gov/

The goal is being able to expose the distributed repository as an endpoint to an OMERO instance installed in Jülich, and demonstrate the feasibility of such a connection.

Finally, in collaboration with the above-listed participants (see Table, we will extract the key use case success stories of their use cases into a form that can be provided to TA6 for documenting the value of the NFDI4BIOIMAGE RDM strategy. In addition, we have identified the following CUCs that will benefit from the measures in TA1:

- Max-Delbrück Center, Berlin: interest in RDM concepts in general and collaboration with microscopy companies in particular to establish metadata principles of image data acquisition (M1.1)
- R. Kurre (UOS, iBiOs): establish standardised workflows for data management of lattice light-sheet data (M1.1, M1.2)
- S. Rehwald (University Duisburg-Essen): establish a RDM infrastructure, which comprises the functionalities of eLabFTW, OMERO and a repository for data retrieval and archiving (M1.1)
- W. Möbius (Max-Planck-Institute for Multidisciplinary Sciences, MPINAT): restructure RDM for Electron Microscopy

Guidelines and best-practices for the optimal implementation of use-case related solutions will be documented and shared with TA5. In the course of the project and through interaction with other consortia and interest groups, further use cases will emerge, which we will work on implementing in this measure.

M1.5		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 1.5.1	Integrate FAIR-IO in third party software							■	■		
DS 1.5.2	Extend FAIR-IO by community requests					■	■	■	■	■	■
DS 1.5.3	Document domain specific guidelines for meta(data)					■	■	■	■	■	■
DS 1.5.4	FAIR-IO export (service/third party compatibility)								■	■	
DS 1.5.5	Report on implementation in use cases										■

5.1.6 *Dependencies and Interactions*

The definition of core metadata and data formats must by their nature closely interact with the other NFDI4BIOIMAGE task areas, but every effort will be made to allow the efforts to progress independently of one another. While the formal specification proceeds, this will be achieved by initial agreements on intermediate formats with an emphasis on using FDO-implementations wherever possible. Similarly, services as they currently exist (e.g., OMERO) can be deployed immediately by TA2 while work on the updated services proceeds. TA3's challenge of linking to and being embedded in external and multimodal standards will be facilitated by building the needed base layers and endpoints into the core model. In collaboration with TA4, existing image analysis software must be updated to work with new structures like NGFF. Existing datasets as provided by IDR and BIA can be used while consortium-specific datasets are generated in concert with participants and the community. This task area will provide TA5 materials for training and outreach and iteratively improve the material based on their feedback. Processes as defined in this task area will work in concert with TA6's plans for the distribution of standards, coordination of development efforts, and the curation activities of all data stewards.

5.1.7 *Risks and Mitigations*

Risks when developing specifications for consumption by a diverse community include: lack of buy-in, over-generalization, poor maintainability, rising complexity and costs over-time. Participants in this task area have a record for defining and maintaining specifications (e.g., OME-XML) over multiple decades. In a field where the primary drivers of new and incompatible formats are the needs of commercial vendors, a clear balance must be struck between commercial and research interests. The focus on usability (M1.2, M1.3) and an open-process (M1.4) should serve as the primary mitigations as outlined in the RDM strategy. Finally, there is a certain risk that such a framework will not be accepted by the wider community. We try to counteract this with targeted training and further education programmes in TA5. Lack of engagement in the process will be mitigated by extensive outreach (conferences, etc.) aided by the Networking Coordinator (TA6).

TA1 is working under the assumption that all FDOs can readily be converted into an RDF graph. Currently, however, the definition of fair data objects is somewhat loose. To the extent that formalizing such definitions is within the scope of the NFDI, we would work with the sections in M6.7 to spell out these expectations in order to have a better understanding of how to transfer metadata between FDO implementations. If such transformations are not possible, individual converters must be built between implementations as necessary for NFDI4BIOIMAGE constituents. Additionally, interoperability with non-RDF vocabularies will be mitigated by constructing necessary mappings between standards. These mappings are version dependent, for example when exporting vendor metadata, and their update may be delayed. Overlap with other processes (BIDS/BEP, etc.) will hopefully be mitigated through the NFDI participation, with the potential of working towards a hierarchical model of research standards. The Research Data Alliance (RDA) will be investigated as a potential home for this process (sec 3.4). The RFC process outlined in M1.4 should provide enough flexibility that participants can enact changes to the specification as necessary in time for their own implementations at any time.

The total number of deliverables in this TA poses an issue as each participant will have differing requirements and timelines for implementation. On the one hand by collecting all the deliverables in a single measure we simplify the tracking of percentage complete, providing a global metric as to the overall feasibility of the framework spelled out in TA1. Finally, the overall goal should be to improve the RDM situation at each participant site and collect bottlenecks and future improvements for the NFDI initiative.

Trust in the resulting frameworks will be built by M1.3 through the availability of a wide-selection of implementations all tested against the technology compatibility kit (TCK) as has been seen in the example of NGFF. The primary risk in the development of truly usable software is the time and effort needed to iterate with users. We will begin by building on existing, open-source solutions and encourage others to participate in these communities in order to build a network of developers. We will make use of M5.4 to involve these developers in workshops and request feedback regarding their RDM needs. Finally, new platforms will be evaluated for inclusion based on the long-term maintainability.

5.2 Task Area 2: Technical infrastructure and cloud resources

In the last decade, many scientific domains have been transformed into data-driven disciplines. Data-driven research has expedited the globalisation of research. Research projects increasingly and crucially depend on the exchange and processing of data and have developed platforms tailored to their needs, e.g., OMERO, Jupyter, and Galaxy, though many do not have access to the necessary technical infrastructure to run them.

To rise to the challenge of big data-driven computational research, German national and/or regional funding initiatives have invested in HPC, HTC and distributed compute capability, including the BMBF with the de.NBI cloud. At the European level we see the formation of computing and storage providers under the European Open Science Cloud, contributing to the vision of a shared environment for researchers and professionals that fosters the reuse of data and computational processing methods across scientific disciplines and national borders. Alongside data processing, FAIR sharing of data is now heavily emphasized and supported by funders, publishers and infrastructures; and adoption of FAIR principles is recognized as a crucial component of Open Science, for reproducibility and innovative re-use of research outputs.

To increase the availability of RDM services, this Task Area will provide infrastructures based on national and regional resources to all users of NFDI4BIOIMAGE (Figure 9). Those infrastructures will be made accessible for various levels of user-expertise. Ready to use web services (SaaS)[63] will be offered but also API access to clouds to independently deploy and maintain new services (IaaS)[64] or interactive graphical applications (DaaS)[65].

---

[63] Software as a Service, https://en.wikipedia.org/wiki/Software_as_a_service

[64] Infrastructure as a Service https://en.wikipedia.org/wiki/Infrastructure_as_a_service

[65] Desktop as a Service https://en.wikipedia.org/wiki/Desktop_virtualization#Desktop_as_a_service

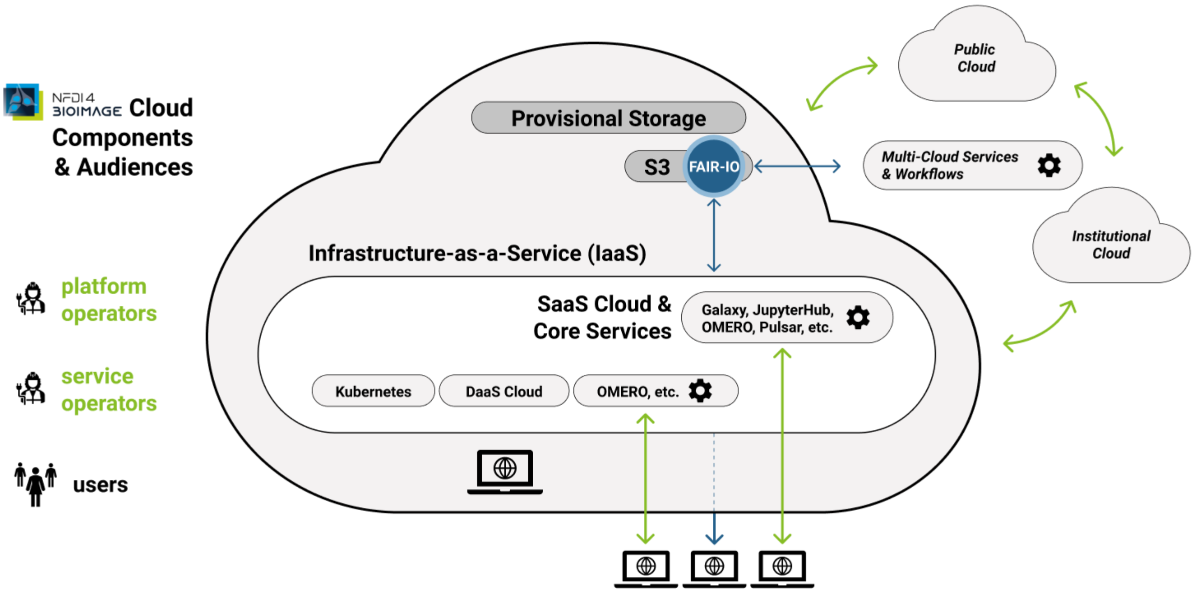

Figure 9: In the NFDI4BIOIMAGE Cloud, provisional S3 storage and core SaaS offerings will be provided by platform operators for all participants. Service operators may deploy and manage services in the IaaS cloud or in Kubernetes. Users may connect to any of the services or start virtual desktops in the DaaS cloud. TA2 will support other cloud teams to integrate with the NFDI4BIOIMAGE cloud resources and run multi-cloud services and workflows. Modified after: NFDI4BIOIMAGE Consortium. (2024). Section 5.2 "Task Area 2: Technical infrastructure and cloud resources" Figure 9. Zenodo. https://doi.org/10.5281/zenodo.13122265, CC-BY 4.0

5.2.1 *Measure 2.1: Infrastructure as a Service*

Contributors: University of Münster (Blank-Burian), ALU-FR (Grüning)

Goal: Establish infrastructures for the efficient deployment of community services

Public clouds have become a standard tool on which to deploy and operate production software. Cloud providers typically offer Infrastructure as a Service (IaaS) on demand to their users. This infrastructure consists of virtual machines, networks, load balancers, storage and more. However, due to privacy regulations and pricing, universities and other institutes have started to provide their own *private cloud* IaaS solutions, mainly based on OpenStack[66], which is the most popular open source IaaS software platform. On top of virtualized cloud hardware, Kubernetes[67] has evolved as a standard platform for operating services. Kubernetes provides an easy-to-use API to create highly available service deployments, especially suited for web applications. At University of Münster, a large multi-tenant Kubernetes platform will be available for services in NFDI4BIOIMAGE. Both OpenStack and Kubernetes platforms are typically used by service operators and not directly by end-users. Thus, the aim of M2.1 is to allow service operators to deploy and administer a platform like OMERO themselves on central resources but for the benefit of their local users. Despite running similar software stacks, all OpenStack deployments have key differences when it comes to requesting hardware resources and operating publicly accessible services. Workflows for these key tasks have to be established and documented. We will also

---

[66] https://www.openstack.org

[67] https://kubernetes.io

offer help and training for service operators on using OpenStack and Kubernetes. The whole software stack in OpenStack clouds or Kubernetes deployments must be updated regularly to guarantee security, stability and adaptability to changing requirements (e.g., GPU computing, FPGA). OpenStack releases two major versions per year and Kubernetes at least three major versions. All updates have to be tested rigorously; installed software may have to be adapted to version changes and new features have to be tested. Hardware must be maintained as disks fail or new hardware must be integrated. We will support the existing infrastructure providers and service operators in all these tasks to guarantee stable operations of this crucial infrastructure.

M2.1		Y1		Y2		Y3		Y4		Y5	
DS 2.1.1	Provide workflows, documentation and training for requesting and using of virtualized resources	🟩									
DS 2.1.2	Support stable operations of cloud services	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩

5.2.2 *Measure 2.2: Community software as a service*

Contributors: ALU-FR (Grüning), University of Münster (Blank-Burian), TA5

Goal: Software as a Service (SaaS) layer will be provided to stimulate best practice research and data management beyond the IaaS level (M2.1)

This measure will offer and maintain four important services for the community and endorse their usage, by providing simple, yet reproducible deployments and training. Those four services are OMERO, Jupyter Hub, Galaxy and Pulsar. In the second half of the project the DaaS (M2.4) will join these services for special purpose interactive GUI centered software environments. We will offer an instance of each service, continuously tested and deployed (CI/CD) for all members of the bioimaging community. In addition, the reproducible deployment recipes will be offered to provision local instances of those services, either on the offered IaaS or on private resources of the users.

OMERO was chosen as the first supported image data warehouse of NFDI4BIOIMAGE based on current usage statistics. OMERO is a complete platform for managing images, with expansion points to integrate with existing tools as well as other systems like Jupyter or Galaxy. We will offer ready-to-use integrations for different types of OMERO cloud deployments.

Jupyter Hub was designed to bring the power of computational notebooks to the end user. Notebooks are popular for prototyping, ad-hoc usage and data analysis with scripting languages like Python or R. They are often referred to as synchronous workflows that a user documents as part of a notebook. Together with rich notebook images, as developed in M2.3, containing JupyterLab, support for X11/Wayland applications, GPU accelerated remote visualization and web based integrated development environments (IDEs), Jupyter Hub can also provide a complete virtual working environment on cloud or HPC resources (M2.3). For launching these images, we will offer such an accessible and flexible cloud based Jupyter Hub for NFDI4BIOIMAGE.

Galaxy is an open, multi-user platform for data-intensive research. It allows scientists to schedule state-of-the-art data analysis tools on large-scale HPC and Cloud infrastructure through a web-based graphical interface. Researchers without programming experience are able to asynchronously run data analysis workflows on their data, share their analyses with others, and enable others to repeat the same analysis. This makes science reproducible, facilitates the sharing of data and results, and eliminates the need for users to compile and install software tools. The power of Galaxy for end-users lies in the ability to develop workflows using an intuitive graphical interface to connect tools, which is crucial for reproducible analyses. The multi-user platform facilitates sharing of developed workflows publicly or with specific colleagues. Galaxy on its own can be considered a Platform as a Service (PaaS), as it is integrating 100+ tools and making them interoperable.

Pulsar allows workflow developers or Galaxy instances to harvest compute resources from all available infrastructures in a standardised cross-cloud way and send computing jobs into those resources. It is a lightweight service that can receive tasks and execute them on requested compute resources. A task defines a set of input files, a set of dependencies (Docker, Singularity, conda, etc.) and commands to run, a set of output files, and various metadata and logging files. Pulsar stages the data from incoming tasks and hands those tasks over to a job scheduler, e.g. HTCondor, SLURM, Kubernetes. Offering a Pulsar endpoint enables research infrastructure providers a unified and rapid way to offer compute resources to a research community through a unified API

A special service provided through this Measure is a Training Infrastructure as a Service (TIaaS) to ensure scalable and no-effort training environments by helping trainers to focus on the training and not on the training environment. The infrastructure will be developed in close collaboration with TA5 and transparently documented to enable the whole community to use it for training of local scientists. Training materials will be provided to TA5 and courses will be organized together with TA5 and TA6. TIaaS is a GDPR compliant system, in which a trainer can request dedicated computational resources during a limited time. Usually a Galaxy instance is shared across all users and all users share one computational queue. In peak times, this might result in some waiting time until a job starts, which is distracting for trainers and trainees. With TIaaS a trainer can request a special queue that guarantees fast executions and dedicated resources for all trainees. In addition, since the training environment is the European Galaxy server, people can reuse the material, data, histories from the training afterwards in the same environment for their real data analysis. For the trainer TIaaS offers a GDPR compliant dashboard that indicates the process of each student in an anonymous way. In combination with the Galaxy Training Network (GTN), all trainings are continually tested on the Galaxy server, so that in the end GTN training material and TIaaS allows training with close-to-zero preparation from the trainers side, especially in remote training settings [40].

In cooperation with M2.4 this measure will in addition offer several DaaS environments that are commonly used by the community.

M2.2		Y1		Y2		Y3		Y4		Y5	
DS 2.2.1	Create SaaS deployment for OMERO for multiple platforms			🟩							
DS 2.2.2	Operate OMERO SaaS instances			🟩							
DS 2.2.3	Provide access to a cloud based shared JupyterHub		🟩								
DS 2.2.4	Provide BioImaging based Galaxy instance		🟩								
DS 2.2.5	Provide distributed resources for job scheduling via Pulsar			🟩							
MS1-TA2	Production services available for the NFDI4BIOIMAGE community				🟩						

5.2.3 *Measure 2.3: Development and extension of community services*

Contributors: ALU-FR (Grüning), University of Münster (Blank-Burian), TA1, TA3, TA4, TA5

Goal: Increase user uptake by improving deployed services

The services deployed in M2.2 are themselves large, independent community projects with their own issues and use-cases, and extension points (e.g., data types, data importers, tools). In close cooperation with TA5, community feedback and requirements for deployed services will be collected, communicated, and potentially addressed as contributions against the upstream platforms with the goal of improving usability.

We will develop rich notebook images, containing JupyterLab, support for X11 applications, GPU accelerated remote visualization and web-based components. When run via JupyterHub, they can provide web-accessible virtual working environments on cloud or HPC resources, similar to curated interactive software environments in DaaS (M2.4). We will create such Jupyter images with common tools and software, used in NFDI4BIOIMAGE for interactive data analysis and algorithm development. The images will be constantly updated and archived.

For bioimaging tools that are not yet readily available, we will support the community in creating conda packages (Bioconda) and containers (BioContainers: Docker, Singularity, rkt) enabling increased accessibility, reusability and reproducibility of those tools. In addition, packaged and versioned tools are easier to deploy in Jupyterlab and also easier to integrate into Galaxy.

Robust installations are a requirement to schedule jobs in an automatic manner across different clouds, e.g. via the Pulsar service. To enable scientists to run large-scale workflows via Galaxy we will support them, in cooperation with TA4, by integrating missing tools or complete workflows into the Galaxy framework. This will include a “request-your-tool” service, but also developer training, together with TA5, to build capacity inside the community, including the use of conda packages [41] and BioContainers [42, 43].

Galaxy features, like a Fair Digital Object (FDO) exporter of an entire workflow invocation have been raised as needed by the bioimaging community and are necessary to preserve the entire workflow provenance during data export (TA3). To this extent, work has already started inside

EOSC-life to develop an FDO implementation called RO-Crate[68]. Building on top of that specification we will develop the Galaxy integration as part of this measure (DS2.3.2). Other Galaxy features can be requested by the community (TA5) and if within the scope of the available resources, will be implemented. The integration of OMERO with Galaxy will happen in cooperation with TA1 and TA3. OMERO will be able to submit jobs to Galaxy, and FDOs like FAIR-IO can be passed around between both systems.

M2.3		Y1		Y2		Y3		Y4		Y5	
DS 2.3.1	Rich Jupyter notebook images				🟩						
DS 2.3.2	Workflow invocation export as FDO in public archives						🟩				

5.2.4 *Measure 2.4: Desktop as a Service*

Contributors: ALU-FR (Grüning; v. Suchodoletz), TA4, TA5

Goal: Provide remote access to powerful compute resources via virtual workstations for analysis, visualization, and training

Due to the size and complexity of bioimaging datasets and the corresponding software tools, analysis and visualization workflows require sophisticated compute resources. Suitable local workstations are expensive to purchase and require regular maintenance and upgrades. In contrast, a cloud-based solution with capable compute power co-located with the data is less expensive, since hardware is shared among researchers and operation is managed centrally. Here, we aim to create a rich cloud environment as a (sustainable) fully open source solution for powerful remote online workstations via DaaS for bioimaging workflows (M2.3). Furthermore, we will build scalable PC pool infrastructures for teaching and training, accessible from a wide range of (low performance) end-user systems (M2.2), and offer general desktop environment replacements for researchers, enabling the migration of all computing activities into the cloud (M2.3). Finally, we will enable remote visualization in various other setups (required in other communities and other NFDI consortia) to avoid copying vast amounts of data and having efficient WAN remote access to resources. Such a remotely-accessible cloud environment is ideally suited for imaging core facilities. Imaging core facilities aim to provide image analysis in concert with image acquisition. This typical use case scenario is given by the Life Imaging Center (LIC) at ALU-FR. The head of LIC, R. Nitschke, is the founder of the global QUAREP-LiMi network, co-founder of the former GerBI-GMB network and an active member of GerBI-GMB, ELMI, BINA, DIN and ISO and thus a pioneer in the promotion of sharing of bioimaging data and advanced bioimage analysis tools. The LIC closely collaborates in networks like EuBI, GBI, NEUBIAS and the EXC CIBBS and livMatS, CRCs 850, 992, 1453, where data sharing between collaborators and remote access to compute resources are highly demanded. The described DaaS environment will be made available to the LIC to set-up and run appropriate high-capacity GPU

---

[68] https://www.researchobject.org/ro-crate/

processing infrastructure tightly coupled to common OMERO- and/or Galaxy-based large data storage to allow remote high-quality data visualization and processing using well-established (commercial and proprietary) software tools under Windows- or Linux-OS.

In addition to the rich DaaS environment dedicated to bioimage data provided to LIC, an online demonstration and teaching platform at ALU-FR for interactive bioinformatics workflows showcasing the typical desktop environments deployed in plant research labs will be established in close cooperation with DataPLANT. We will also use these very different applications to benchmark our cloud environment.

The development of an open-source DaaS solution available for OpenStack has been started already at ALU-FR[69] and is of high relevance for the broader research community. Thus, our contribution will be aligned with the activities of DataPLANT and coordinated with other efforts.[70] As such, it holds the potential for becoming a cross-cutting topic of NFDI.

To generalize the concept of open source DaaS environments and to be able to support a broad range of such use cases with a maximum of flexibility in terms of application and operating system support, a complete DaaS solution is desired. However, the existing commercial solutions are quite expensive and not available for OpenStack. Therefore, we aim at several objectives starting from establishing a sound overview on the state of the art and ongoing activities on open source DaaS solutions as well as available technologies, leveraging cloud platforms. It should offer a 3D, video capable graphical software environment on standard open source cloud installations like OpenStack.

As DaaS is a sought-after technology, various technological frameworks already exist (both commercial and to a rather limited degree in Open Source). These solutions can provide the means and ground truths to check certain expectations and features against, before delivering similar services from an OpenStack cloud. Starting from this baseline we will provide a summary of State-of-the-Art in DaaS regarding relevant hardware and software drivers as well as protocols, provide a gap analysis and a community requirements document (in close cooperation with TA4 and TA5) to clarify the tasks defining the envisioned DaaS. Within this analysis, we will define the required rendering capabilities (3D, video, etc.) and possible transport options to the researchers. For the infrastructure side the necessary resource reservation, advance planning of demands and suitable scheduling mechanisms are to be selected or developed into a DaaS orchestration model for OpenStack. This would include to script the necessary API calls to set up and start the required resources and provide the necessary user session selection interface.

---

[69] The PePP project, https://www.hnd-bw.de/projekte/pepp/, will prepare basic scheduling of DaaS resources in OpenStack. Resources provided equal to three person years.

[70] Such an infrastructure is envisioned for general teaching purposes as it allows to offer scalable remote desktop infrastructures independent of the location of the researchers taking part in courses.

Further, we include hardware evaluation and technology watch and work on the list of required driver components (check of availability, suitability). Depending on the outcome, missing components will be developed taking a (previously created) matrix of encoders, decoders and other rendering requirements e.g. for 3D into consideration (check for multiplexing options - how many VMs of a certain configuration or requirement can share a GPU). Definition of suitable (desktop framebuffer stream) transport protocols including hardware backed optimizations on server and client side (including detailed analysis of state-of-the-art protocols from DS 2.4.2 with respect to bandwidth, latency, quality/perception). Finally, a module/component for fast remote access in OpenStack is to be created (How to configure a suitable graphical rendering interface? How to attach the respective user to their interface?) In some use cases, alternate DaaS clients (beside browser-based access) might be desirable. To complete the user remote interaction researchers should be enabled in future versions to interact with the VM on non-visual channels (non-video/display channels) like allowing file-sharing between VM/DaaS and a user's remote client or the transport of audio from/to an user in a bi-directional channel. Further it might be desirable to integrate devices attached to remote hardware interfaces into the DaaS/VM (e.g., remote pass-through of USB devices, printer support, etc.).

M2.4		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 2.4.1	Establish a technical baseline for general user remote access (existing solutions)	🟩									
DS 2.4.2	Create a basic VNC (virtual network computing) based/backed DaaS (lacking certain performance, displaying features) on the cloud resources provided in NFDI4BIOIMAGE including orchestration of resources		🟩	🟩	🟩						
DS 2.4.3	Develop enhanced (hardware backed) video framebuffer transport model for the remote visual cloud access					🟩	🟩				
DS 2.4.4	Additional remote interaction channels (USB access, two channel audio, file transfer)							🟩	🟩		
DS 2.4.5	Further features of advanced OpenStack optimization: Suspend and resume our sessions (including switching remote clients in between)									🟩	🟩

5.2.5 *Dependencies and Interactions*

TA2 will provide computational infrastructure and services via IaaS, SaaS, TIaaS, and DaaS to all members of the consortium and the imaging community. Specifically, with TA1 we will create robust OMERO deployments, either public or private depending on the community needs. Together with TA1 and TA3 we will work towards an integration of FDO into different services like Galaxy and OMERO to create a generalized exchange format between services. Cloud-based and simplified access to DaaS will be developed in concordance with TA4 (M4.3). To cover the demand on increased batch-processing needs we will further work with TA4 to port tools and workflows to the cloud to make them accessible via Galaxy. Moreover, TA4 will help us to determine the hardware and software requirements for effective use of cloud resources from interactive environments, like the Galaxy infrastructure (M4.3). To support the community training needs we are running TIaaS together with TA5 and will offer admin and developer training to onboard the community to new concepts of cloud computing and containers.

5.2.6 *Risk and mitigation*

Operating complex systems like clouds and engineering on system level requires an uncommon skill set. Hence, finding qualified people and sustaining them is very challenging. We will offer continuing education to render the positions attractive, qualifying for careers in the wider scientific IT environment.

A further risk lies in a too steep entry level for service operators. We will mitigate this by providing better training for service operators and trainers in TA5, work with cloud providers to simplify service operations and assist in service setup and operations. Not all analysis software may be portable to clouds and containers. M2.4 provides a Desktop-as-a-Service environment, where all other software can be run.

By enabling one-click deployment of complex platforms like OMERO, one may give service operators an incorrect sense of security, leading to incomplete management of the platforms. We will take steps to clearly communicate the expectations for operators upfront and attempt to monitor critical situations (lack of security updates, insufficient storage, etc.) and notify the responsible parties. In the event that data must be transferred out of the provisional cloud (e.g., end of grant, change of institution), the export tools of TA3 will allow researchers to freely migrate their data.

Risks concerning the DaaS are that additional financing and wider community contribution cannot be secured in M2.4, so that some deliverables have to be dropped. Furthermore, the necessary rendering and encoding hardware may not become available with a reasonable open source driver support and/or pricing for general large scale application. To avoid lock-ins all three potential hardware-vendors (AMD, NVIDIA, INTEL) will be considered and contacted early to ensure vendor support. Not all deliverables might get finished if complex problems (excessive adaptation required because of API and code changes in the Linux kernel, bug hunting) arise. At least a basic DaaS (DS 2.4.2) is a reasonable goal to achieve to provide at least support for some use cases.

5.3 Task Area 3: Multimodal data linking and integration

Answering fundamental research questions generally requires input from different experimental approaches that deliver a variety of data types. Thus, modern research projects rely on combining and jointly exploring multimodal datasets. In this TA we first propose an automated system for uniquely identifying biological samples for tracking their whole lifecycle, linking to a multitude of measurements and data, and capturing relevant metadata immediately when produced (M3.1).

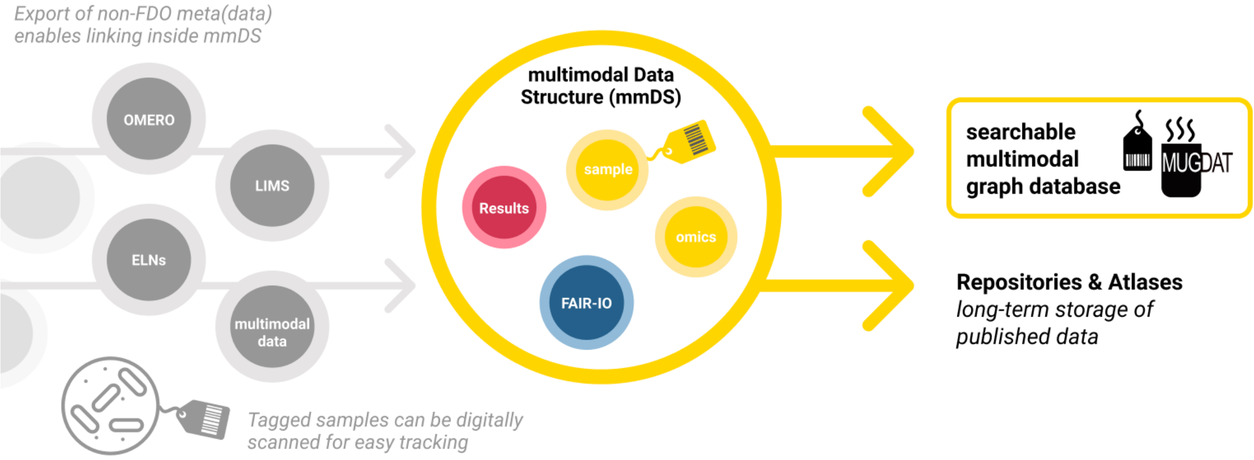

Figure 10: Data flow to obtain a multimodal data structure (mmDS) with an overarching graph database (MUGDAT). Taken from: NFDI4BIOIMAGE Consortium. (2021). Section 5.3 "Task Area 3: Multimodal data linking and integration" Figure 10 (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.10389955, CC-BY 4.0

We will then address the issue of the management of multimodal data in M3.2 by conceiving and building a data organization system, the multimodal data structure (mmDS) that stores all measurements from an experiment in a single directory with descriptive files and metadata attached. Next, we will implement mechanisms for the import and export of data from mmDS from other (domain-specific) data management systems (M3.3). Finally, in M3.4 we will develop an overarching database, the multimodal graph database (MUGDAT) that enables searching over multimodal datasets according to sample type, acquisition modality etc. exploiting ontologies defined in TA1 (Figure 10).

The focus of the TA is on creating a system for multimodal data management that is suited for the inclusion and integration of large and complex bioimage data. An example for the frequently requested integration of bioimaging data with sequencing data is illustrated by the Quantitative Biology Center in Tübingen [44]. Here, openBIS, an integrated ELN&LIMS system, and OMERO were linked and metadata visualized in a project browser. In addition, spatially resolved omics studies were recently highlighted as a data management challenge while guidelines were suggested to address these issues [45]. Similarly, the integration from different modalities such as Mass spectrometry imaging (MSI), (Imaging)-Flow-Cytometry, Multiparameter Image Spectroscopy or omics datasets (proteomics, genomics, metabolomics among others) will be dealt with in this task area. Our measures will thus follow the experimental workflow of (i) sample generation – the sample being the common denominator, (ii) data structure to assemble and organize multimodal data, (iii) provide interfaces to other systems for automated synchronization and data transfer with an over-arching database (MUGDAT), and (iv) archiving solutions according to the FAIR principle.

The resulting structure will use unique identifiers to cross-link FAIR-IO-formatted datasets (see TA1) to similar data structures from other modalities. We will develop supporting software tools for initializing a new mmDS directory, converting to and from both BIDS and ARC, and accessing mmDS from Python, i.e. adding data, receiving data and/or metadata, performing minor searches. Such a well-specified data organization makes automatic data processing and cloud computing (see TA2 and TA4) much easier. In addition, the data structure can be easily packaged for archival or transfer to a collaborator at another institute (M3.3).

5.3.1 *Measure 3.1: Automatic sample tracking and sample metadata acquisition*

Contributors: LIN (Stöter, Zuschratter), DKFZ (Mallm), UFZ, UHD, INP, IPHT, TA1, TA5

Goals: Test and provide tools for automated, 2D matrix code-based sample tracking and sample naming convention to facilitate data linkage and synchronization

Since many read-outs and measurements can be obtained from one biological specimen, it is imperative to implement a convention for sample naming and tracking that will be used as a central entry point for the anticipated multi-modal data structure and ultimately links samples with data. Eventually, all acquired data will then be organized under the predefined specimen. A 2D matrix code for uniquely identifying, tagging and tracking samples will be used. At LIN, scanners, printers, paper and DataMatrix codes have been tested that are compatible with all major operating systems and work well under low light conditions in microscope rooms. However, there is still need to elaborate on a suitable code scheme that is easy to handle, e.g. many predefined codes that can be printed and handed out at once before they are used. Yet, the code should be meaningful and interpretable by humans using naming conventions for different sample and assay types that need to be defined. Various sample scenarios (i.e., source, types, protocols, acquisition) provided by the PTPs UFZ, UHD and IPHT will be considered and discussed for finding suitable naming schemes. In addition to sample tracking, the processing workflow needs to be integrated in the proposed central entry point. The protocol for instance must be linked to the sample and the sample must be linked to the raw data. Hence, the sample ID needs to be integrated into the metadata model (TA1), into ELNs, image databases and potentially other software systems. This is to be realized through an overarching API interface, bundling the other software APIs developed in this TA. We will work on the sample integration with the OMERO image database together with all PTPs listed above. UFZ and INP will concentrate on linking sample IDs with ELNs/LIMS and the INPTDAT platform. Guidelines for linking sample IDs with ELNs, protocols and meta-data will be developed in collaboration with TA5.

M3.1		Y1		Y2		Y3		Y4		Y5	
DS 3.1.1	Guidelines on hardware tested and used for automated sample tracking	■									
DS 3.1.2	Providing code schema and naming conventions	■	■								
DS 3.1.3	Tools and guidelines for linking sample IDs with ELNs, protocols and meta-data		■	■	■						

5.3.2 *Measure 3.2: Universal data structure for multi-modal research projects*

Contributors: LIN (Stöter, Zuschratter), DKFZ (Mallm), UHD, FZJ, UoC, TA1

Goal: Specify a data structure for file system level data management of multi-modal raw research data

After measuring biological samples by, e.g., microscopy, all generated data should be stored and organized in a structured way. For microscopy data and many other modalities specialized databases and storage solutions exist, e.g., OMERO, openBIS or Bisque in bioimaging; XNAT and other PACS (picture archiving and communication systems) in medical imaging; and gGenestackODM[71] or site-specific data management systems in the omics domains. However, for multi-modal research projects there is no one-fits-all solution due to the diversity of data types, data sources and file formats. The least common denominator for organizing and storing data is in a directory hierarchy with files on the file system, augmented with relevant and descriptive meta-data. This type of data organization is also the one most users are accustomed to. To implement this common denominator for multimodal scientific studies, we propose building on, adopting, extending, and also contributing to two existing data structures, the Brain Imaging Data Structure (BIDS, [46]) and the Annotated Research Context (ARC)[72] standards developed by the neuroscience community and the DataPLANT NFDI consortium, respectively. Such organisation of files with descriptive and consistent naming is essential for keeping track of data, assuring completeness and facilitating analysis. Our vision is to enable all data from a multimodal scientific study to be bundled together in one place, the multimodal Data Structure (mmDS), rather than solely being distributed across modality-specific data management and storage solutions. mmDS will be built on BIDS and ARC as underlying directory structures, i.e., mmDS can be viewed as flavors of and used on top of either BIDS or ARC. At the same time, we aim at resolving the shortcomings of these standards, maximize their benefits and harmonize the use of both. The mmDS will integrate imaging and non-imaging modalities, capturing all necessary metadata and linking it with the corresponding research data. It will follow file format and metadata standards defined in conjunction with TA1. We will also develop supporting software tools, for bootstrapping, conversion, and interfacing with mmDS.

BIDS: The BIDS standard has been developed continuously by the neuroscience community over the last decade and has become an accepted standard in that field. It suggests a well-defined

---

[71] https://genestack.com/news/blog/wellcome-sanger-institute-adopts-genestack-omics-data-manager/

[72] https://nfdi4plants.de/

directory structure per subject or sample measured, naming conventions for directories and data files, sidecar meta-data files, and standard file formats to be used. Its main focus is functional and anatomical MR imaging of the brain. Via BIDS extension proposals (BEPs) the standard can be adopted and extended by another science community. Software tools called BIDS Apps help in automatic validation, quality control, and data conversion. Currently, BIDS is being extended for microscopy in the BIDS extension proposal BEP031 to which our PTP Timo Dickscheid (FZJ) is actively contributing. We want to continue these efforts and push for BEP031's finalization. We plan to initiate a new BEP for omics data analytical measurements and assays which so far are absent (see Measure 3.3 and 3.4) and thus strengthen our BIDS community involvement. Consistency across BEPs is asserted by the rotating BIDS steering group, in which we pursue a voice with our PTP T. Dickscheid.

ARC: The ARC standard is a feature-rich directory hierarchy designed for plant biological research and developed within the NFDI DataPLANT initiative offering the necessary extensibility for interoperability with the bioimaging domain. It intends to capture and organize the whole research data lifecycle and forms a homogenized RO-Crate compliant entry point. ARC's focus is on bioassay targets stored in tabular form. Based on the ISA model (ISA-Tab), it packages research data from analytical measurements or assays and metadata, but also workflows and computation results in a corresponding directory structure. Quality scores are used to assess the "FAIRness" of datasets. Supporting software tools SWATE and ARCcommander[73] allow for easier metadata annotation and data management. So far, the ARC data structure captures omics data very well, but lacks support for bioimages. This is a highly requested community need within DataPLANT. We want to foster collaboration with DataPLANT by implementing bioimaging support into ARC and align on file format and meta-data specifications provided by TA1 and PTP UoC. We also want to adopt ARC's quality scoring for bioimaging data. Based on PyBIDS[74] and ArcCommander we will provide an API and CLI (command-line interface) to assist in initializing a new and managing an existing mmDS directory, converting to and from both BIDS and ARC, and accessing mmDS from Python, i.e. adding data, receiving data and/or metadata. These tasks will be carried out by PTPs UoC and UHD for ARC and BIDS respectively. To amplify user-adoption we want to develop a data collection tool at LIN, which as early as the data acquisition allows files to be added remotely to a stored mmDS with study specific metadata, configurations and validations. For all specifications, processes and tools we will provide documentation and training material to TA5.

M3.2		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 3.2.1	Specifications for easy-to-use data structure		▨	▨	▨						
DS 3.2.2	BIDS and ARC extension for bioimaging and multimodal data organization		▨	▨	▨	▨	▨	▨	▨		
DS 3.2.3	API/CLI interface tools			▨	▨	▨	▨	▨	▨		

---

[73] https://github.com/nfdi4plants/arcCommander

[74] https://bids-standard.github.io/pybids

5.3.3 *Measure 3.3: Combine data from different (imaging) modalities*

Contributors: LIN (Stöter, Zuschratter), DKFZ (Mallm), UoC, FZJ, INP, TA1

Goal: Enable mmDS-import/export of meta(data) from and to specialized and domain-specific data management solutions, such as OMERO and XNAT among others

Transfer of data between domain-specific data structures and mmDS is required to benefit from the advantage of combining data from different modalities in one place. Importing and exporting from and to mmDS also allows shipping a self-contained research dataset to a different institution or collaborator where no or only some of the domain data management solutions are available. Importantly, where possible, e.g., for OMERO, only metadata will be captured along with references to large files to avoid duplication of data.

By interfacing and synchronizing with domain-specialized solutions, we can maintain access to features such as data viewing, metadata queries and basic tools like histograms, ROIs, vector annotation or figure-editing offered by those solutions. Moreover, existing multimodal data, which are not compatible with domain-specific data management solutions like OMERO, should be cross-linked to other domain-specific data management solutions by MUGDAT (Measure 3.4).

We will implement above functionalities for OMERO and selected domain-specific data management solutions by writing an interfacing layer between the domain solution's API and our mmDS API. An individual interface per data management solution implemented as a plugin or module can be realized as an isolated small development project. Due to the overall complexity of this measure participants from UoC, FZJ and INP will be involved. Specifically, the import and export functions will be implemented by these participants in their respective field / focus domain.

M3.3		Y1		Y2		Y3		Y4		Y5	
DS 3.3.1	OMERO-mmDS importer/exporter			▨	▨	▨	▨				
DS 3.3.2	OMERO-mmDS synchronizer			▨	▨	▨	▨				
DS 3.3.3	Domain specific databases-mmDS importer/exporter				▨	▨	▨	▨	▨		
DS 3.3.4	Domain specific databases-mmDS synchronizer				▨	▨	▨	▨	▨		

5.3.4 *Measure 3.4: Integrate imaging meta(data) with other data types and omics data in a multimodal graph database (MUGDAT)*

Contributors: LIN (Stöter, Zuschratter), DKFZ (Mallm), MPIEB, IPHT, INP, TA1, TA4

Goal: Establish an overarching database for searching and retrieving data and metadata from mmDS datasets, the MUGDAT database

This measure will make metad(data) from FAIR-IO and other FDOs, which have been synchronized as described in M3.3, searchable and retrievable according to their modality, specimen or other characteristics. To this end, we will develop MUGDAT, a multimodal graph database that uses SPARQL/knowledge graph technology for searching RDF-based metadata

(IPHT, LIN, MPIEB). A key information in the indexed datasets is the sample / donor characterization. In a human or animal disease model context this may include GDPR non-critical data such as mutation status, flow cytometry profiles, treatments. It is important to note that this information cannot be used to identify an individual but is nonetheless crucial within a project for efficient data analysis. For sensitive data, such as sequencing or medical imaging data, only links to established secure and protected storage will be included. Here, we will provide guidelines on how to add meta(data) from these repositories and also how to preserve the link between bioimage data and, e.g., sequencing data. Examples here could be linking regions of interest (ROIs) from a laser capture microdissection experiment and FASTQ files from RNA / DNA sequencing. As described above these metadata will be stored in JSON-LD files or imported via flexible and user-friendly ISA-tab-based approaches in the multi-modal data structure (mmDS). It is important to note that we also aim for a solution that fosters the combination of data from (i) multiple centres, (ii) diverse disciplines and (iii) analysis states e.g. raw, pre-processed and quality-controlled data. MUGDAT will be provided as a webtool which is independent of converters or domain or system specific software. MUGDAT will preserve the relationship between datasets and provide all metadata as well as the file structure. As MUGDAT uses ontologies defined in TA1 for metadata description, it can be used to facilitate multimodal data analysis by any site using FAIR-IO or other FDOs supported by the mmDS (meta)data structure developed in M3.2 and M3.3. With MUGDAT, there will be a straightforward connection to other NFDI consortia such as NFDI-Neuro, NFDI4Immuno and NFDI4Chem and the chance for an interdisciplinary, community-driven approach for multimodal data integration. Several members of NFDI4BIOIMAGE take part in the aforementioned initiatives and are thus ideal partners to develop an overarching database for knowledge graph searches. While we aim to facilitate multimodal data analysis, we plan to complement the information stored in MUGDAT with a description of the applied data analysis pipelines (as described in TA4). TA4 will provide an interface to enable researchers to manage experimental procedures, raw and processed data and computational analyses.

M3.4		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 3.4.1	Software Solution to extract metadata and file information from mmDS to build MUGDAT					▨	▨	▨	▨	▨	
DS 3.4.2	Provide a web-based tool to search MUGDAT					▨	▨	▨	▨	▨	▨
DS 3.4.3	Link MUGDAT to data analysis for starting data processing and feedback results and data analysis procedures / pipelines					▨	▨	▨	▨	▨	
MS4-TA3	Preview of the MUGDAT metadata base									▨	

5.3.5 *Measure 3.5: Solution for archiving and publishing multi-modal data*

Contributors: LIN (Stöter, Zuschratter), DKFZ (Mallm), FZJ, INP, MPIEB, UFZ, TA1, TA2, TA4

Goal: Facilitate data transfer from mmDS to local and public archives

To share datasets in an orchestrated manner across groups and collaborators, standardized routines from mmDS to endpoint repositories are needed. All collected metadata will be included. Best practices for internal archiving will be formulated in order to help users manage precious raw and preprocessed datasets. In connection with data repository tools, the mmDS environment can be prepared for data transfer to public repositories. Apart from bioimage archives such as EBI's BIA or the IDR as mentioned in TA1, we aim to couple OMERO / mmDS with selected domain-specific repositories such as INPTDAT, the German Human Genome/Phenome Archive (GHGA) and tissue atlases such as the Human Cell Atlas or EBRAINs facilitated by FZJ. GHGA could be used for spatial sequencing data and thus combining both omics and imaging datasets as a prototypic multimodal example. For the aforementioned repositories, tools and guidelines will be provided that allow an automatic upload. In case of INPTDAT, the database initiator at INP will couple our data structure to the repository. These steps can also serve as templates for users to develop their own tools for any given public database.

M3.5		Y1	Y2	Y3	Y4	Y5
DS 3.5.1	Best practices to set up an internal domain-specific repository for archiving multimodal data				🟨	🟨
DS 3.5.2	Tools and digital workflows, e.g. packaging and publishing of multimodal BIDS/ARC directories to internal / open repositories				🟨	🟨
DS 3.5.3	Provide submission portal to upload multimodal data to selected public archives				🟨	🟨

5.3.6 *Dependencies and Interactions*

TA3 will interact with essentially all other task areas as outlined below. In general, intense discussions are needed to streamline our efforts and agree on common metadata and data structures. Data linkage to FAIR-IO and embedding in OMERO is a central theme that needs close interaction with TA1, especially with regard to the development of APIs, metadata structure and sample annotation. Our data structure needs to be compatible with the infrastructure guidelines and tools as outlined in TA2. For this reason, we have proposed a clear folder structure usable at local and centralized infrastructure, and in concordance with TA4 that can be utilized for cloud computing. We will closely adhere to the needs for efficient data analysis which includes importing and exporting data, metadata accessibility and linking datasets for automated pipelines as outlined in TA4. To train users to exploit our tools and software packages the support of TA5 will be essential - also in connection to CUCs as described below. Data stewardship will play a key role in establishing and testing our proposed integration scheme. This will be coordinated with TA6. Tracking and automated sample annotation is a general need across disciplines. Thus several members of our consortium, including UHD and the CUC at UDUE, will benefit from the implementations described in M3.1 in the context of larger consortia. The integration of several microscopic systems and modalities can be challenging especially with unusual formats or systems with limited commercial support. One example would be electron microscopy. Therefore, the MPIEM will test our multimodal interface to manage imaging data and facilitate analysis sustainably. Here, the work of data stewards will be highly beneficial to support the CUCs. In the context of spatial transcriptomics the link between sequencing and imaging data as well as sample annotation, e.g., ROIs, tissue and sample type, have to be recorded and linked to other datasets. mmDS and MUGDAT will provide this functionality which will be utilized by the group of M. Gerstung at the DKFZ.

5.3.7 *Risk and Mitigation*

The success of our proposed measures depends on the acceptance of the community and easy and straightforward implementation. We believe that in a highly interactive and interdisciplinary field our deliverables can simplify project and sample management as well as data analysis. It will be imperative to have many use cases to show their power and feasibility. Thus, we will collaborate with TA5 to deploy our system and put it to the test by the community. Furthermore, we have decided to use existing, well-established tools and extend their capabilities but still preserve bi-directional compatibility. The integration of these accepted data types and formats will be highly beneficial for researchers well beyond the bioimaging community.

There might be limitations to integrate (meta)-data from some resources due to complexity or data protection regulations. The interfaces also need to be maintained and constantly adjusted to the requirements of the community and new data formats. As described above we will provide a modular and highly flexible approach for multi-modal data integration. We expect that this mitigation can be administered by the community itself due to our preparatory, well-documented work and adaptable implementation that can be leveraged by the planned networking and workshop efforts in TA5.

Due to the very specific requirements for public repositories an automated upload needs well annotated and complete datasets. Most of the databases will only host specific datasets, e.g., either imaging or sequencing related. Here we will lobby for cross-referencing of datasets or even better to manage linked datasets from independent readouts in the same repository. The proposed work here will be capable of facilitating both scenarios.

5.4 Task Area 4: Bioimage informatics and analysis

For the majority of microscopy experiments, images are acquired to be analyzed. The biological conclusions are made on derived data, which can only be made FAIR if the analysis methods themselves are FAIR. Herein lie the main aims of TA4: ensure reproducibility of analysis, make popular individual tools more interoperable, and make state-of-the-art method deployment accessible to the whole bioimaging community. With the explosive growth of microscopy image volumes brought by the advances in instrumentation, making analysis FAIR is paramount for open science and true accessibility of Big Data in microscopy.

The NIH Center for Open BioImage Analysis has recently run a survey on the community experiences and needs for the future [14]. According to this survey, the vast majority of image analysis is performed using "Open-source point-and-click" software, e.g., Fiji [15, 47], Icy [48] or CellProfiler [49]. Our own survey confirms that Fiji is the most used tool in the community both among bioinformatics & data science professionals and non-expert survey participants [19]. Such tools provide powerful algorithms and are easy to use. Still, they are often reliant on manual adjustment of parameters that are not automatically stored and preserved for repeated and reproducible analysis. For example, in Fiji, there exists the possibility to write all commands as a macro and there is even a macro recorder that stores manual parameter adjustments. Unfortunately, such macros are rarely provided with analyzed data. One explanation for this could be that the typical analyst of bioimage data is a person with a background in the biological sciences with limited to average computational expertise [14]. This means that many researchers performing excellent image analysis do not have the training and habit to write code and do not use version control tools to track their progress. It is not feasible or even desirable to limit image analysis to just data scientists, as knowledge about the subject matter is a key aspect of any meaningful analysis of microscopy data. Therefore, it is vital that the tools and community strive towards the consistent reproducibility of analysis pipelines without requiring advanced programming skills. To achieve this task, we aim to introduce GUI widgets and APIs into popular analysis applications to accurately capture analysis pipelines. TA participants represent some of the most popular analysis tools, e.g., Fiji and Ilastik. We will also identify and engage with developers of other applications. This part of TA4 will be in close cooperation with TA5, as the dissemination of developments is of utmost importance to make a difference for the community.

Another important aspect is the large number of available tools with their respective strengths and overlapping application areas. This means that multiple tools are often used for specific tasks like segmentation, tracking, and object classification in a single project. Additionally, novel algorithms are typically developed for a specific tool and are hard to implement for users who want to create other parts of the analysis in a different framework. Therefore, there is an urgent need to facilitate the combination of tools by developing Standard Operating Procedures (SOPs) for different tools via APIs. The APIs can either work by the tools communicating directly with each other or by having a common data input/output (I/O) structure. An example of how the former can be achieved is how Fiji can be integrated with other developments using the Java API. An example of the latter is how ilastik [10] and Fiji can exchange information using the HDF5 file format. We will work with other TAs to ensure that the developments are in line with the (meta)data standards established in TA1 and TA3. Also, the dissemination of these SOPs will be made in close collaboration with TA5.

A crucial advantage of reproducible settings and programmatically connected pipelines lies in their capacity for scaling. Modern microscopes can easily produce terabytes of data for a single experiment, moving manual or semi-manual analysis out of reach and requiring substantial computational resources (Figure 11). Furthermore, with the on-going AI revolution in the Computer Vision domain, there is a strong interest in deep learning methods for microscopy [50]. The deployment of most deep learning-based tools requires both expertise and GPU resources, missing in many experimental labs and even imaging facilities. However, these resources are readily available through cloud computing services promoted by the EOSC initiative, both from commercial providers and from academic initiatives such as de.NBI[75]. In NFDI4BIOIMAGE, TA2 will be concerned with cloud resource provision, while TA4, working in close collaboration, will aim to make these resources accessible to researchers without substantial computational background by enabling, simplifying and automatizing the remote/cloud deployment of popular image analysis tools. To make sure that the developments for cloud computing become accessible and known to the community, we will work with TA2 towards development of cloud infrastructure and with TA5 for dissemination. For the entire project time of TA4, we will be in close contact with TA6 to integrate our work with both the NFDBIOIMAGE consortium and the NFDI community as a whole.

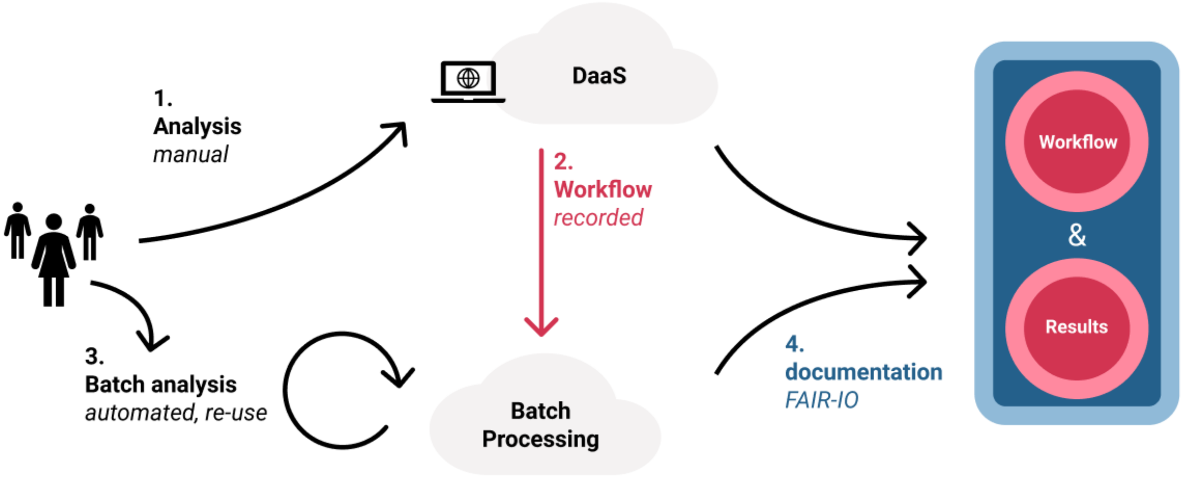

Figure 11: FAIR pipelines for image analysis to balance the current status quo of hardly recordable manual processing (point-and-click) with the goal of increasing reproducibility of workflows. Taken from: NFDI4BIOIMAGE Consortium. (2024). Section 5.4 "Task Area 4: Bioimage informatics and analysis" Figure 11. Zenodo. https://doi.org/10.5281/zenodo.13121527; CC-BY 4.0

---

[75] http://denbi.de

5.4.1 *Measure 4.1: Formalize best practice standards for users and developers of microscopy image analysis tools*

Contributions: HKI (Figge), EMBL (Kreshuk), TUD (Haase), TA5

Goal: Create SOPs and support tools for users and developers of bioimage analysis software

For end-users, we will look at a number of the most commonly used tools in the bioimage analysis community and formulate how these tools are best used to comply with the FAIR principles. Measures like writing or recording of macros instead of pointing and clicking and using a version control tool, may seem obvious to computer scientists. However, the large ecosystem of tools for image analysis and the variety of people who use them does not always mean that this can be considered possible or practical. We will therefore compile such a guide for the general community. On the software development side, we will improve the possibility and automation of parameter tracking in some of the most used tools in the community. In more detail, we envision three directions of work: i) improve reproducibility of Fiji, the most popular image analysis software, by the development of Fiji Time Machine; ii) develop reproducibility extensions for popular tools, such as napari or Icy plugins and the corresponding libraries for tracking all project parameters; iii) make multi-tool pipelines reproducible by further development of the JIPipe software - a graphical programming tool that shares a lot of functionality with ImageJ and KNIME [51] while still being able to perform efficient batch processing. JIPipe is in an excellent position to bridge the gap between point-and-click tools and coding by its intuitive interface and the widespread familiarity of ImageJ in the community. The Fiji Time Machine will be contributed by TUD (R. Haase). The main motivation behind its development is that Fiji and its plugin ecosystem are in a constant and non-synchronous state of development. Thus, it is challenging to reproduce workflows combining former versions of plugins and ImageJ core even with excellent documentation. Even if all plugin versions are given, it might require hundreds of manual downloads to recreate the environment. The Fiji Time Machine will cover the traditional distribution scheme for Fiji via update sites and the upcoming scheme via Maven repositories.[76] Reproducibility improvements in other software will be developed by EMBL (A. Kreshuk's group) which hosts the team behind the popular ilastik software [10]. Our aim with this measure is to create easily reusable components for popular image analysis tools that can be added to existing or newly developed plugins in napari, Icy or CellProfiler to keep track of the experimental parameters and save them in a shareable and reusable form. As part of a reproducible image analysis pipeline, it is important that all original information regarding the data is preserved. Therefore, we will keep in contact with TA3 and their work on metadata integration to ensure that metadata is not lost or corrupted as data passes through the pipeline. This is also an area of research where our collaboration with CUCs will be important, especially Microverse in Jena and MDC in Berlin. These CUCs perform multimodal imaging and image analysis in their core facilities

---

[76] https://maven.apache.org/index.html

and across many sub-projects. They are therefore ideal partners to evaluate tools and concepts developed in TA4 as well as to give feedback for further developments.

M4.1		Y1		Y2		Y3		Y4		Y5	
DS 4.1.1	JIPipe SOP example collection of project-based image analysis tool development (HAWK, and the CUCs MDC and Microverse EXC Jena)	■	■	■	■	■	■	■	■		
DS 4.1.2	Launch of the Fiji Time Machine					■	■				
DS 4.1.3	First version of reproducibility extensions for plugins-based tools released							■	■		
MS3-TA4	Release of the SOP document, taking into account all developments of this Measure								■		

5.4.2 *Measure 4.2: Image analysis tool integration*

Contributions: HKI (Figge), ISAS (Chen), JGU (Czodrowski), UFZ

Goal: Create APIs on either the level of data or direct communication between tools

Given the number of available tools for image analysis, their different strengths and the steady development of new image modalities, it is not realistic to see a convergence in the number of tools in the future. With this in mind, we will develop tools and work with use cases and participants to facilitate building pipelines that consist of multiple tools. One example of this would again be JIPipe, which allows running Fiji, Python and R scripts besides its built-in functionality. In other cases, such as the interfacing between Fiji and Ilastik, the tool interaction is best done via data I/O functions for each tool. As long as one of the two tools that should communicate is open-source, it is possible to adjust I/O so that the tools can communicate in this way. The developed I/O implementations will conform with the guidelines developed in TA1 as this will ensure that final and intermediate results can be used for reproduction and validation of the image analysis pipeline. To facilitate the transition to cloud storage and computation, cloud compatible formats such as NGFF will be prioritized. See M1.5 and M6.7 for more details on the interaction between the development of format standards and the other TAs in NFDI4BIOMIAGE.

J. Chen (ISAS) is the author of the popular Allen Cell and Structure Segmenter (ACSS) tool[77] for deep learning-based light microscopy image analysis. Within this measure, we will ensure that the neural networks of ACSS are compatible with the community development of neural network specification format (bioimage.io), while ACSS itself becomes interoperable with other popular tools at the neural network level. Furthermore, we will establish a benchmark set for validation of 3D data analysis pipelines, including the necessary infrastructure to perform pipeline comparisons.

UFZ will also ensure that the pipelines for analysis of high content systems, e.g. multispectral imaging flow cytometry, imaging of small organisms and imaging of bioassays, follow neural network specification formats (bioimage.io) and are compatible with the data format integration that we are developing with TA1.

---

[77] https://www.allencell.org/segmenter.html

Similarly, JGU (P. Czodrowski) will work on ensuring interoperability of their analysis pipelines for cell painting assays with a number of standard tools, e.g. Fiji or CellProfiler. This work would ideally mean that the pipelines can be accessed in a standardized way from other tools using specific APIs, but interoperability at a minimal level will be achieved on the level of data integration. To ensure data compatibility, JGU is also active in TA1.

M4.2		Y1		Y2		Y3		Y4		Y5	
DS 4.2.1	JiPipe example collection of integration of different tools deployed and evaluated by CUCs Microverse and MDC			🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
DS 4.2.2	Creating a unified workflow for multimodal vector image data						🟥	🟥	🟥		
DS 4.2.3	ACSS follows the emerging community neural network specification					🟥	🟥				
DS 4.2.4	Analysis pipelines for high content systems follow the emerging community neural network specification							🟥	🟥		
DS 4.2.5	Analysis pipelines for cell painting data follow the emerging standards for interoperability with the most common community tools							🟥	🟥		

5.4.3 *Measure 4.3: Simplify the transition to cloud-based computing environments*

Contributions: EMBL (Tischer, Kreshuk); UFZ, TA2, TA5

Goal: Facilitate upscaling desktop algorithms to cloud computing environments

This measure will work closely with TA2 to enable or substantially simplify execution of common image analysis workflows on remote cloud resources. Here, two approaches will be taken: interactive use of tools through a DaaS, and batch analysis, using pre-defined parameters or pre-trained models. The former approach is in active development in the EOSC-Life project providing containerized image analysis tools in a desktop environment accessible through a web browser, using an a local DaaS implementation[78]. The virtual desktop promotes interoperability by offering many tools in a common environment and its modular container architecture allows easy expansion by addition of new image analysis software as needed. By providing a familiar environment, the virtual desktop eliminates most hurdles biologists face when moving their work to remote resources. This measure will continue the containerization of image analysis tools, submitting them to the central NFDI4BIOIMAGE DaaS.

For cloud-based batch analysis, we will turn to popular workflow management tools and make common image analysis software available as nodes. For this, our first choice is Galaxy — a widely-used and well-supported visual workflow tool from the domain of bioinformatics. Galaxy is user-friendly and familiar to many biologists, once nodes have been created the composition of the workflow is very straightforward. The composed workflow can easily be deployed on clusters and, importantly for our use case, on scientific cloud computing resources available to the NFDI4BIOIMAGE community. Furthermore, Galaxy workflows can be deposited in the WorkflowHub[79]. We will create Galaxy nodes for popular image analysis tools, e.g. ilastik,

---

[78] https://band.embl.de

[79] https://workflowhub.eu

CellProfiler, StarDist, CellPose, and a Fiji macro executor and support them with extensive documentation and training materials, targeting biologists with a limited computational expertise.

Another important goal of this measure is concerned with accessibility of deep learning-based tools. While such tools are continuously pushing the frontier of accuracy, scale and throughput in bioimage analysis, their deployment remains out of reach for most biologists. The BioImage Model Zoo[80] has recently emerged as a community initiative to make deep learning-based models for microscopy FAIR. It is backed by teams behind *ilastik*, ImJoy [52], CSBDeep/FIJI [53], DeepImageJ [54] and ZeroCostDL4Mic [55]. The model specifications in the Zoo follow a common standard which makes them interoperable between the tools of the community and easily accessible to biologists. We will build on the model deployment libraries provided by the Zoo and extend them to be runnable as Galaxy nodes. Due to the „black box“ nature of deep learning models, reproducibility issues are particularly important for analysis pipelines which include neural networks. We will join the efforts to define and refine a common model standard that will fully track training parameters and provenance of the training data. Furthermore, together with TA5 we will develop training materials to ensure the AI-based models are used responsibly, teaching biologists how to validate the results and recognize implicit biases in trained networks. UFZ will contribute with the focus on integrating workflows for analytical high content systems into our cloud-based analysis system. The focus will be on data analytics using deep learning classifiers. UFZ will more generally work on data formats and transfer for this type of high-content data within TA1 and TA3 and, in TA4, will ensure that the pipelines are compatible with the cloud analysis environment that we are developing together with TA2.

M4.3		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 4.3.1	Deliver virtualized software to NFDI4BIOIMAGE users via DaaS			■	■	■	■				
DS 4.3.2	Galaxy nodes implemented for popular bioimage analysis tools					■	■	■	■		
DS 4.3.3	Galaxy node implemented to run deep learning-based methods for image analysis			■	■	■	■				
DS 4.3.4	Implementation of example pipelines for high content system analytics using the Galaxy infrastructure							■	■		

5.4.4 *Dependencies and Interactions*

The goals set in TA4 are naturally anchored to the other TAs of NFDI4BIOIMAGE, as all analyses are based on data. For example, the work of TA1 on (meta)data standards will be vital as the IO-processes of each analysis pipeline have to conform to these standards. The standardization of formats will help in the proposed work on data APIs facilitating the combination of multiple tools into one analysis pipeline. Similarly, the work on next-generation file-formats (NGFF, TA1) will be of high importance for the work we will do to make cloud computing accessible to a greater portion of the research community. For cloud computing, a close collaboration with TA2 will also be necessary, where TA2 will approach the system administration side and resource provision side

---

[80] https://bioimage.io

of the problem, while we will work on supporting the user workflows to run in a cloud environment. The metadata integration developed in TA3 will need to be supported by analysis tools so that metadata does not get lost when data is passed through an analysis pipeline. We will, together with TA5, take measures making our work known within the community. One very concrete step is that we will contribute to the online resources provided by TA5 providing clear descriptions and resources for the work performed in TA4. Information passed on to TA5 will include our own software and documents and will work as a portal to many different resources, like software tools, data repositories, and tutorials. For the linked software, we will also provide details on what strengths it has with regard to the SOPs. In addition, we will continuously be present at virtual platforms such as the image.sc forum[81] as well as at conferences such as "From Images to Knowledge with ImageJ & Friends"[82] or "Image-based Systems Biology"[83]. We will naturally also disseminate our findings in the NFDI community during the community meetings to identify further cross-cutting topics where we can exchange experiences with members of other consortia through a close cooperation with TA6.

We will closely work with a number of CUCs and use cases at our participant institutions to better understand and fulfill the needs of the community. C. Chalopin (HAWK) requires a unified workflow for pre-processing and analysis of multimodal vector image data, will be included in M4.1 and M4.2 to fulfill need both for new tools and tool integration. The EXC „Balance of the Microverse" is another CUC where we will interact mainly with the core imaging facility of the cluster to ensure that our tools developed mainly in M4.1 and M4.2 are compatible with their needs. Of special interest for TA4 are the novel imaging modalities and imaging in novel experimental model systems. Examples of novel techniques that are being developed in the cluster are super-resolution microscopy, organ-on-chip models and imaging of microfluidic droplets. The cluster will also be a good candidate to test the methods for upscaling to cloud computing described in M4.3. We will interact with the CUC from MDC in M4.1, M4.2 and M4.3 as they have an established structure including analysis of many different modalities of image data and methods. All CUCs will be regularly updated on the progress of work directly connected to their needs through bilateral meetings in person or online. We will also ensure they are informed about events, e.g., workshops or hackathons, of interest to them organized by TA5.

5.4.5 *Risk and mitigation*

The obvious risk for developing SOPs, software and APIs is that the technology development in general advances in such a way that software is obsolete before launch and guidelines are outdated. The SOPs will continuously be revised and updated to comply with the latest developments. The SOPs will also not be specific to a certain tool and will be transferable to new

---

[81] https://forum.image.sc/

[82] https://www.janelia.org/you-janelia/conferences/from-images-to-knowledge-with-imagej-friends

[83] image-based-systems-biology.com

tools and techniques by being general rather than applied to specific tools. JIPipe is supported by the HKI for sustainability and will be Open Source for continued community development as new needs arise.

Changes in the base tools may invalidate any API. Therefore, we will as far as possible focus on well-established tools using standard languages like Python or Java. All code will be Open Source with the intent to have a community around the initiative that maintains the APIs. For data I/O integration, our interaction with TA1 will ensure that our methods are compatible with the file-format standards, for example FAIR-IO that are being developed within the consortium.

The upscaling to cloud computing naturally requires access to a cloud computing infrastructure. In the case of delays in the TA2 developments, we will use existing server infrastructure at the Co-applicant institutions to develop and test our tools.

5.5 Task Area 5: Training and community integration

This TA will focus on the in-depth training of all community stakeholders with the goal to convey full data management competence across the entire bioimaging data lifecycle [56] (sec. 2). Awareness about and literacy on data formats and the appropriate metadata (TA1), knowledge of the deployment and utilization of modern technical hardware infrastructures (TA2), the ability for robust cross-linking to other data types from one's own and related research fields (TA3), and

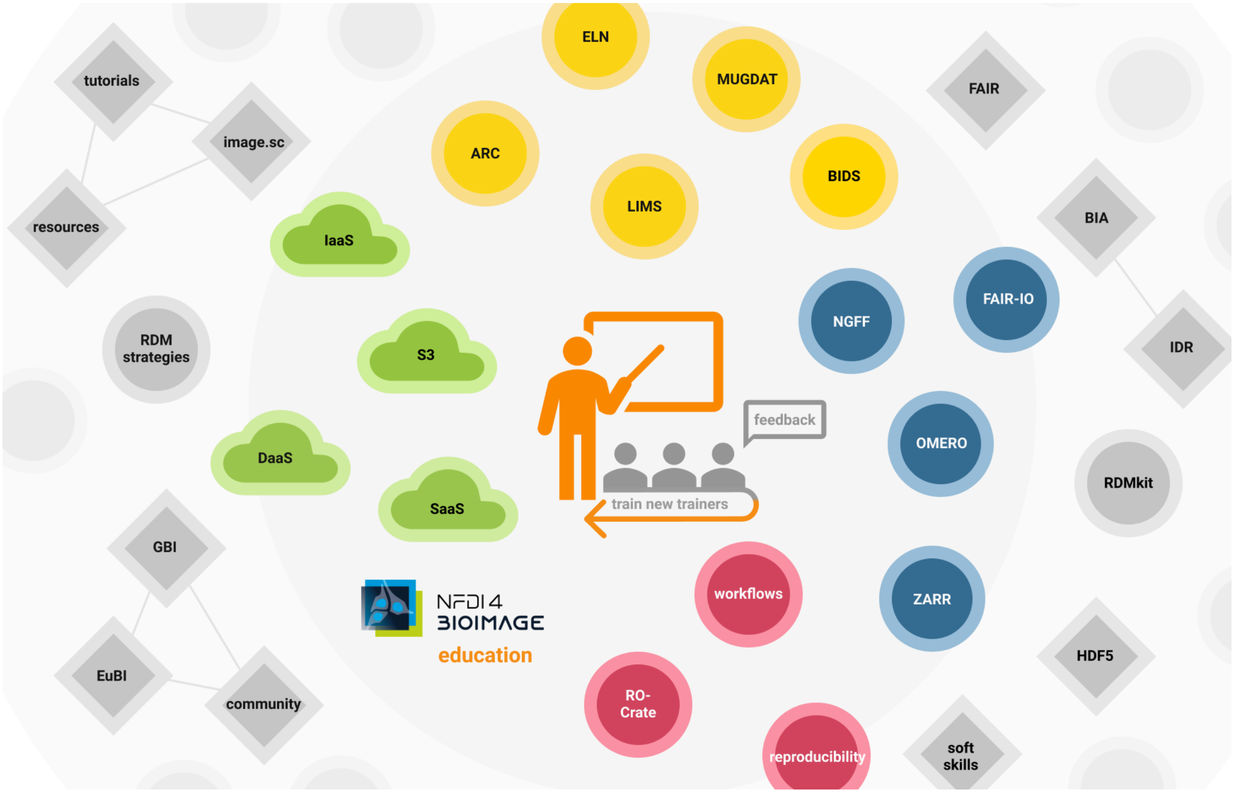

Figure 12: Provision of comprehensive training and user education on all aspects of RDM as developed and deployed by the Task Areas 1-4. The goal is to capacitate the community for professional handling of all matters along the bioimage data lifecycle. Modified after: NFDI4BIOIMAGE Consortium. (2024). Section 5.5 "Task Area 5: Training and community integration" Figure 12. Zenodo. https://doi.org/10.5281/zenodo.13121833; CC-BY 4.0

a well-trained skill-set to extract information from data with state-of-the-art analysis tools (TA4) are the cornerstones of successful bioimaging research and must be underpinned by a corresponding training programme (Figure 12).

To achieve these goals TA5 will leverage knowledge established in all thematic TAs. Our training concept consists of four measures: establish the NFDI4BIOIMAGE training portfolio (M5.1), collect and curate existing training resources in a searchable resource (M5.2), engage the community in the continuing and sustainable development of the training resources (M5.3), and coordinate the training specific aspects of the Help Desk and the DaSts team (M5.4).

5.5.1 *Measure 5.1: NFDI4BIOIMAGE Training portfolio*

Contributions: TUD (Haase), University of Münster (Zobel), TA1, TA2, TA3, TA4, TA6
Goal: Create a comprehensive training programme for the bioimaging community with NFDI4BIOIMAGE training materials

In the course of the project, TAs 1-4 will produce software tools, documentation and training material. TA5 will collect these materials (Table 5.5.1), organize and expand them where necessary to build a coherent modular training programme for NFDI4BIOIMAGE members and beyond. NFDI4BIOIMAGE will work at the forefront of bioimage RDM establishing this field in the larger scientific RDM landscape. Users will need proper training measures to acquaint themselves with novel concepts like FAIR-IO, knowledge graphs, and the use and management of AI-based analysis software just to name a few prominent examples. We will take care to include different media that target the various professional profiles: system administrators and software developers, image facility managers, image analysts, principal investigators, Postdocs, and (PhD) students. We will also leverage the experience at HHU in introducing bioimage RDM training at the undergraduate level and develop a FAIR Starter-Kit (see section 4) for researchers with no previous experience. Attention will be paid to convey not only the usage of single RDM tools, but to educate users to combine and connect them flexibly depending on the problem to be solved. Multiple surveys highlighted the importance of offering training and teaching of different formats [14, 19]. Written online tutorials and slides are important for scientists seeking materials for teaching; video recordings are increasingly important for users who need immediate advice and training. The latter is highlighted by the success of recently implemented YouTube channels such as that of the NEUBIAS community[84] and of R. Haase[85] with several thousand subscribers and about 170.000 video clicks (counted until Oct. 10th, 2021). The training formats will also include webinars, hackathons, and (in-person) workshops, where we will show how to share and distribute source code, mentioning pitfalls and highlighting success stories. On-premise workshops and individual solutions are particularly important when setting up RDM concepts (e.g., the installation of an OMERO server) since the infrastructural conditions at different

---

[84] https://youtube.com/neubias
[85] https://youtube.com/haesleinhuepf

universities vary. We will also implement and expand the "Rent-the-Trainer" concept that was successfully tested in the context of bioimage analysis by GerBI-GMB[86]. To raise the visibility of our training program, we will be present at international conferences in the bioimaging field (e.g., ELMI, Seeing is Believing, Focus on Microscopy, etc.) as well as at national RDM-related events like the E-science days, the "Tag der Forschungsdaten NRW" and general NFDI meetings. We will additionally approach relevant professional societies like the German Society for Cell Biology (DGZ) or the Society for Biochemistry and Molecular Biology (GBM). Participation will be planned together with the networking coordinator of the NFDI4BIOIMAGE Office (TA6). We also plan to organize or co-organize bioimage RDM training events, for example in the frame of the "Trends in Microscopy" Spring School, the highly education-oriented, biannual conference of GerBI-GMB. A further well-established forum is the RDM4mic group through which we can reach a significant number of imaging core facilities in Germany. At these events, we will collect feedback on the training program per se, as well as on the usefulness and performance of NFDI4BIOIMAGE services and convey it to the TA leads. Further instruments of evaluation will be the planned NFDI4BIOIMAGE surveys, as well as monitoring access to our training resource search index (see below). Survey results and search statistics will be published after 3 years and by the end of the funding period after 5 years in a format analogous to the one chosen by the UK and US bioimaging communities [14, 57]. Overall, this process is in line with the intended high-level of community participation of the NFDI4BIOIMAGE and should capacitate the community to make best use of the future infrastructure.

Table 5.5.1: Resources developed by TA1, TA2, TA3, and TA4, and disseminated by TA5

Measure	Description / Keywords	Delivered services /resources type
1.1: Define an FDO subtype, "FAIR Image Objects" (FAIR-IO)	FAIR-IO concept extension	guidelines, best-practices
1.2: Increase usability of RDF metadata for bioimaging	Metadata annotation service / validation, Knowledge graph concept	guidelines, tools, usage examples cases & training
1.3: Enable cloud-native image acces	Integration of NGFF	Collection of application examples
1.5 Implement community requirements and use cases	Application of FAIR-IO concept to different use cases	FAIR-IO promotion training material, use cases
2.1: Infrastructure as a Service	Training for service operators	documentation and training
2.2: Community software as a Service	Training Infrastructure as a Service	no effort training, provision of IT for trainings, training environment, use-cases
2.3: Development and extension of community services	"request your tool service" / Galaxy / developer trainings Jupyter notebooks	requirement definition & feedback, developer trainings Jupyter images, training material,
2.4: Desktop as a Service	Cloud-based remote online workstations, DaaS, community requirements document	community requirements document, training tools (DaaS with hardware & software)
3.1: Automatic sample tracking and sample metadata acquisition	Linking sample with data, ELNs protocols and metadata	tools & usage, guidelines, training material & use cases
3.2 Universal data structure for multimodal research projects	easy to use data structure for integration of different data types	training material

---

[86] https://www.gerbi-gmb.de/RentTheTrainer

3.4: Integrate imaging (meta)data with other data types and omics data in a multimodal graph database (MUGDAT)	make data (and linked metadata) searchable and retrievable	webtool, training material
4.1 Formalize best practice standards for users and developers of microscopy image analysis tools	Fiji Time Machine, JIPipe	SOPs, best practices, tools
4.2 Image analysis tool integration	multimodal images, JIPipe tool integration	tools, example collection of exampless,
4.3 Simplify the transition to cloud-based computing environments	Virtual desktop, galaxy nodes, deep learning	tools, training materials for AI-based toolsl

M5.1		Y1		Y2		Y3		Y4		Y5	
DS 5.1.1	Develop and evaluate a portfolio of training formats	■	■	■	■	■	■				
DS 5.1.2	Develop training concepts for various professional profiles		■	■	■						
DS 5.1.3	Collect and prepare materials from TAs		■	■	■	■		■		■	
DS 5.1.4	Prepare undergraduate bioimage RDM training and a FAIR starter kit			■	■						
DS 5.1.4	Offer trainings and individual workshops according to our portfolio			■	■	■	■	■	■	■	■
DS 5.1.5	Offer trainings and workshops at conferences			■	■	■	■	■	■	■	■
DS 5.1.6	Collect feedback on training and services of NFDI4BIOIMAGE				■	■	■		■		■
MS2-TA5 MS4-TA5	Publication of survey results and search statistics						■				■

5.5.2 *Measure 5.2: Implement a cross-site search index for bioimaging RDM training resources*

Contributions: TUD (Haase), University of Münster (Zobel, Blank-Burian), UKON, EMBL (Keppler), GerBI-GMB (Hanne)

Goal: Enable RDM stakeholders to efficiently discover new and existing learning materials

We will build a comprehensive search index of diverse RDM training resources for the bioimaging community. Several organizations such as GerBI-GMB, EuBI and GBI collect links to educational materials on their websites. Individual researchers maintain YouTube channels, teaching basics and recent advances in bio-image data science. Communication about new methods happens increasingly online via social media channels and virtual meetings, a trend that was boosted by the Covid-19 pandemic. Many generic-level resources for RDM training have been established including, e.g., forschungsdaten.info (UKON) or ELIXIR-RDMkit.

However, there is no central entry point for life scientists working in imaging and bio-image analysis to find resources for state-of-the-art RDM training resources and materials. To address this need, we will develop a meta search index that collects training and teaching materials for existing and reliable online sources. Users should be presented with options similar to established travel booking web portals to search and book hotels and flights from multiple agencies at the same time. Here, they can find and download microscopy, bioimage data analysis and RDM training materials. With this resource we want to cover all three aspects of bioimaging-based research to share a comprehensive view of the bioimage data lifecycle, from acquisition to archiving, with the community. This search index will be integrated into the NFDI4BIOIMAGE website, but importantly, it will be accessible from all websites that deliver content for mutual benefit. To this end, we will stipulate agreements that permit all partners to link reciprocally to each other's materials. We will collaborate on this topic amongst others with A. Keppler (Director Bio-Hub EUBI-ERIC and GBI Coordinator at EMBL) and J. Hanne (Managing Director GerBI-GMB) and will join efforts with TA2 to implement the necessary technical solutions and with TA6 on legal issues with materials. The training material developed within M5.1 will be included in the results. This will initially consist of materials provided by TA leads, e.g., allowing us to define common metadata in collaboration with TA1. Together with TA6, we will then explore bioimaging and bioimage-related resources like bii.eu, QUAREP-LiMi, GerBI-GMB, EuBI, GBI, Microscopy Australia and others. We will also investigate including more generic RDM tools, exploiting our connections to https://forschungsdaten.info (UKON), https://rdmkit.elixir-europe.org (HHU), tess.elixir-europe.org and the NFDI at large. We believe that if these latter generic tools and concepts are presented in the context of a resource for bioimage data, they will reach researchers in our community more effectively and make the benefits of RDM applied to bioimaging more apparent. We will strive for a search index that is technically easy to maintain, and we will also engage the community to ensure sustainability of this resource (see M5.3).

M5.2		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 5.2.1	Cross-site meta-search index	■	■	■	■	■	■				
DS 5.2.2	Collection of material for the search index	■	■	■	■	■	■	■	■	■	■
DS 5.2.3	Collection of material from TA1, TA2, TA3 and TA4			■	■	■	■	■	■	■	■
MS1-TA5	Launch of the cross-site search index with first set of materials				■						

5.5.3 *Measure 5.3: Community-driven process and re-usable tools for curation and long-term viability of training materials*

Contributions: TUD (Haase), University of Münster (Zobel)

Goal: Maintenance and sustainability of training resources via community engagement

For storing the training materials produced by the consortium, we plan to use existing online platforms like zenodo.org, f1000.com, figshare.com, as well as institutional platforms. Our search index will then contain links and/or DOIs to storage locations of materials and is thus a collection of materials, annotated with metadata. To guarantee the long-term quality of this information, technical solutions will be developed, e.g., automatic detection of broken links, curator notifications and deprecation of materials. We will then establish quality assurance procedures comparable to the state of the art in scientific publishing. As inclusion in the search index is open, we will implement sanity checks that must be passed to make materials visible in the public search. The procedure will be similar to the sanity check on the bioRxiv preprint platform[87]. Furthermore, automatic routines will ensure that materials behind provided links do not change

---

[87] https://www.biorxiv.org/

over time. We will motivate the community to provide DOIs instead of universal resource locations (URLs). The aforementioned aspects will be documented in the form of meta-curation guidelines. Furthermore, a peer-review process will be established to identify materials of high quality. As the review process is open and transparent, experts from the RDM field can build their reputation by reviewing materials and thus, ensuring continuous improvement of materials available to the community. Reviews can also be liked by the community and commented on. Successful examples of such open and public review schemes can be found on f1000.com, stackoverflow.com and image.sc. Those materials can then be highlighted in the search results. The review process and the editorial mechanisms of our search index will be published for transparency. Furthermore, we will strive for making the processes and committees as open as possible, welcoming external collaborators from other NFDIs and beyond to make the founded structures sustainable and available also after the funding period.

M5.3		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 5.3.1	Meta-curation guidelines for searchable index community materials				■	■	■	■	■	■	■
DS 5.3.2	Publish review process and editorial mechanisms for quality check of material					■					

5.5.4 *Measure 5.4: Data Stewardship and Help Desk Activities*

Contributions: University of Münster (Zobel), TUD (Haase), HHU (Weidtkamp-Peters), UKON, TA6

Goals: Provide direct community support for bioimaging RDM with central entry point and tailored data stewardship project assistance

The NFDI4BIOIMAGE Data Steward Model foresees allocating capacity in the form of FTE time to the Help Desk and to direct project support. Importantly, it is a hybrid model, i.e., FTEs are not devoted exclusively to support duties but are also involved in project work in the TA to which she or he is assigned. We will thus build a constantly operational support team with a broad spectrum of skills that will further develop as the project progresses and should cover all types of requests, from researchers of all career stages, IT-support staff, facility staff, and RDM staff. While TA6 (M6.3) supports the team at the administrative and coordinative level (e.g., process for the selection of use cases, distribution of funds), TA5 will lead its' practical operations supervising the iterative exchange with CUCs and supporting proper documentation to generate guidelines, data management plan (DMP) templates considering the needs of bioimage data, best practices, success stories, and novel training materials. Thus, the team plays a central role for practical implementation of the consortium's deliverables.

The community support offered by the DaSts consists of two main activities: The Help Desk and the tailored CUC support actions. For the Help Desk, entry points will be the NFDI4BIOIMAGE website and other suitable online support platforms like image.sc. The team will develop a concept for a low-threshold service. It will arrange a "shift-plan" for availability at the HelpDesk according to the DaSt time assignment of each team member relative to his or her project work. Mid-term, the team will be able to first respond to every new support request within one working day. Regular online “walk-in-hours” will facilitate personal exchange and the identification of appropriate support measures. We follow the principle of *capacitating the community* rather than *taking over* tasks for it. A support action might require as little as pointing to the suitable training materials, or identifying contact persons at an institution. Other requests might require short-term assistance (hours to days), e.g., helping to draft a research DMP, moderating the request in expert platforms like image.sc, and facilitating the exchange of community members among each other. More advanced issues that would absorb DaSt capacity for weeks to a few months, will be formally regarded as CUC applications which might receive on-premise support (the CUC process is described in sec 3.5 and TA6, M6.3).

The pilot CUCs acquired during preparation of the proposal and several use cases at the co-applicant institutions will serve as immediate test cases for the DaSt team for assessing and then improving its procedures (Table 5.5.2).

Table 5.5.2: Identified test use cases for the DaSt team

Use case	**description**	**connection to Task Areas and required Data stewardship expertise**	**estimated person-month assignment**
CUCs at non-participant and non-applicant sites			
MDC Berlin (J. Steinkötter et al.)	metadata standards, structured storage of large image datasets together with image analysis tools like code, machine learning models etc.	TA1, TA2, TA4, TA5 RDM / Databases (OMERO / ARC / Infrastructure) /	6 months
JGU Mainz (C. Meesters et al.)	DaaS / connecting iRODS - OMERO / offline processing (on HPC)	TA2, TA4 Databases (OMERO / iRODS), Image Analysis, HPC-Clusters / Galaxy	4 months
UDUE (S. Rehwald)	Interaction of EPICS-ElabFTW-OMERO	TA3 ELNs / Data linking / OMERO	4 months
QBiC Tübingen (S. Nahnsen)	Integration of bioimaging data with sequencing data	TA3 OMERO / openBIS	4 months
Use cases at co-applicant and participant institutions			
UKON (L. Becks)	standardized storage and annotation for HCS data / independent long-term storage / data linking / AI & ML with HCS Data	TA1, TA2, TA3, TA4 HCS data storage / metadata handling / data linking and analysis	6 months
UOS (R. Kurre)	standardized workflow for lattice-light-sheet microscopes (Betzig lab), data-processing and storage	TA1, TA2, TA4 large image data-handling / metadata handling / analysis pipelines	4 months
DKFZ (M.Gerstung)	mmDS implementation for spatial transcriptomics data	TA3	4-6 months
Balance of the Micro-verse Excellence Cluster (A. Brakhage)	infrastructure for bioimaging data and analysis / ensure the reproducibility of high-performance image analysis pipelines,	TA2, TA4 large image data-handling / analysis pipelines	6 months

Special roles in the Data Stewards Team: Three positions (at TUD, HHU and UKON) share leadership responsibilities in the DaSt team.

The DaSt at TUD will chair the team and focus in particular on RDM support for bioimage analysis workflows in collaboration with the bioimage analysis technology development group at PoL/TUD and the University‘s compute center ZIH, a DFG-funded national high performance compute center (NHR) for the life sciences. They will collect and guide the evaluation of HelpDesk requests. The identification of use cases for bioimage analysis will be based on the reusability of publicly available datasets, tools and workflows. The DaSts at HHU and UKON will collaborate on the alignment and cross-fertilization of our consortium with existing institutional RDM initiatives such as RDM competence centers, INF-projects of local or transregional CRCs, or data stewardship activities of CRCs and clusters of excellence (EXC). The goal is to identify existing best practices, seek synergies and mutual support with respect to bioimage data handling. HHU and UKON will lead the collection of existing data management plans (DMPs) and test their suitability according to requirements for describing bioimaging data with appropriate CUCs. HHU will connect bioimaging data steward support with the generic RDM initiative fdm.nrw together with the UDUE (S. Rehwald, CUC). At HHU, the RDM competence team and the data steward of CEPLAS EXC (D. Brilhaus) will be important partners for exchange. In addition, the DaSt at HHU will be responsible for collaborating on the Elixir RDMkit for bioimaging together with J.M. Burel (OME). UKON will leverage its leading role in the context of bw2FDM to integrate bioimaging-specific RDM knowledge into the German platform forschungsdaten.info. Furthermore, UKON hosts a pilot DaSt use case in the area of High Content Screening for eco-evolutionary dynamics research (L. Becks). UKON will lead the collection of CUC documentations and steer the formulation of bioimaging-specific data stewardship guidelines.

M5.4		Y1		Y2		Y3		Y4		Y5	
DS 5.4.1	Help Desk in action (collaboration with TA6)		■	■	■	■	■	■	■	■	■
DS 5.4.2	Bioimaging-specific guidelines for Data Stewards in the community (TUD, HHU, UKON)	■	■	■							
DS 5.4.3	DaST team interacting with test use cases (table 5.5.2)		■	■	■	■	■				
DS 5.4.4	DaST team interacting with new use cases			■	■	■	■	■	■	■	■
DS 5.4.5	Interaction and exchange with local DaSt's and RDM teams	■	■	■	■	■	■	■	■	■	■
DS 5.4.6	Contribution of "Bioimaging" as research domain to ELIXIR-RDMkit and forschungsdaten.info and regular revision	■	■		■		■		■		■

5.5.5 *Dependencies and Interactions*

Within the consortium TA5 serves as a central hub to make any output generated by the technical task areas available for the community and at the same time collects feedback from the community to initiate further improvement of services and tools. Here are a few examples about these services and tools from the different TAs:

TA1 will provide TA5 with tools and guidelines to easily annotate and search annotated datasets, making it easier to re-use and share data. TA2, which takes care of the technical infrastructure, will provide us with TIaaS as well as system administrators and developer training. TA3 will develop tools for data linking and multimodal data integration and TA4 will define and develop best practices, image analysis tools and cloud base analysis solutions. These tools, developed by TA1, TA2, TA3, and TA4, are communicated to the community by TA5 in the form of workshops, guidelines, and other materials. All materials will be available and searchable in our index. To ensure high quality training materials, and to make sure we are always working on topics that are of importance to the community, TA5 will feed back the needs and inspirations of the community to the single TAs. Providing this feedback loop is an essential task of TA5. TA5 will be supported by TA6 to organize all administrative aspects of the DaSt activities and organizational issues of training and workshop events. TA5 will interact and exchange on our concept of data stewardship with all national and international initiatives with similar concepts including other NFDI consortia like DataPLANT.

5.5.6 *Risk and Mitigations*

We see a risk in the parallel and potentially redundant development of resources for training materials considering that there are already quite a few, like biii.eu, TeSS, forschungsdaten.info etc. As mitigation, we plan to collaborate closely with these partners and ideally find a common way forward that leads to a unique, bioimage RDM-specific resource with shared participation. Our DaSt team faces some risk to develop solutions for problems that are not important to the community. Hence, we will implement surveys to exchange frequently with the community about its needs. Furthermore, as short term project decisions may be taken by the individual institutions there is the risk that DaSt time is dedicated to singular issues. Constant exchange between the DaSts, the other TAs and in particular TA6 will make sure that time is well-invested for developing solutions valued by the community.

5.6 Task Area 6: Coordination, governance and networking

This Task Area is responsible for setting up the management structure, supervising, and coordinating all the activities of the consortium. It will also lead the national and international networking. The operational arm of TA6 is the NFDI4BIOIMAGE office that will be led by HHU and DKFZ combining administrative and scientific leadership with community networking, integration and outreach (M6.1). The governance will be orchestrated according to the organizational structure and operating model (sec 3.4, M3.5) (M6.2). Furthermore, this Task Area will coordinate the commissioning of Data Stewards (DaSts) and flexible project assignments (M6.3). As it will be imperative to ensure exchange between all stakeholders, to gain visibility in the scientific community and beyond, and to provide accessibility to our infrastructure for community members, TA6 will set up a three-tiered communication strategy (M6.4 to M6.6). Finally, it will facilitate the collaboration on cross-cutting topics within the NFDI (M6.7).

5.6.1 *Measure 6.1: Project management, administration and coordination*

Contributions: HHU (Weidtkamp-Peters), UKON/DKFZ (May)

Goal: Set up the NFDI4BIOIMAGE Office (sec 3.4) consisting of the Managing Director (HHU) and the Networking Coordinator (DKFZ)

We will implement a bipartite NFDI4BIOIMAGE Office with general management responsibilities mainly assigned to HHU (S. Weidtkamp-Peters) and networking responsibilities assigned to DKFZ (E. May). The two persons can mutually represent each other and share organizational and communication tasks for a coherent and transparent project management. Together they will be responsible for monitoring and evaluating the consortium's activities and progress.

The team will set up and maintain the consortium's communication infrastructure as described in M6.3 - M6.5. It will equip the RSE team (sec 3.5) with a collaborative DevOps working environment as needed. To this end, we will build upon resources of GerBI-GMB and refine them taking into account the needs of the RSE staff working in TA1 to TA4. The website nfdi4bioimage.de will be adopted and further developed as the central platform of the consortium and is managed and maintained by the NFDI4BIOIMAGE office in collaboration with University of Münster (TA2). The office further collects job profile descriptions and supports staff recruitment at the member institutions. Overall budget management is the responsibility of HHU, who also stipulates collaboration contracts with PTP institutions for funding allocation. At proposal submission, six co-applicant institutions are members in the NFDI association. We aim to increase this number to become a „consortium according to statutes“ within the NFDI association.

The NFDI4BIOIMAGE Office will plan, prepare and document the meetings of the different governance bodies as described in section 3.5. Monitoring and evaluation will be performed to guide project progress according to the success criteria defined in section 2.2, and an evaluation plan will be submitted to the Steering Board. Monitoring includes checking for the timely provision of deliverables and completion of milestones, but also external parameters such as the number of downloads of software, of Help Desk requests, of FAIR datasets deposited in BIA and IDR by German scientists, etc., measuring the impact of the NFDI4BIOIMAGE on the community. An annual survey based on the successful 2021 NFDI4BIOIMAGE community survey (in collab. with TA5) will allow us to follow the development of bioimaging RDM literacy within the community over time. These activities will be collectively presented in the final project report at the end of the first funding phase.

M6.1		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 6.1.1	Collaboration contracts with PTP institutions	🟩									
DS 6.1.2	Project management infrastructure and procedures in place	🟩									
DS 6.1.3	Monitoring and evaluation plan		🟩								
DS 6.1.4	Final project report										🟩

5.6.2 *Measure 6.2 Governance*

Contributions: HHU (Weidtkamp-Peters), DKFZ (May) with contributions from all TA leads

Goals: Implement the governance structure (sec 3.4) and ensure an early operational state

TA6 will draft the consortium statutes and a data handling policy that complies with the members' institutional data policies and the statutes of the NFDI association. The SB will meet shortly after project start, elect three members to the BoD, and propose three external advisory members for the CB. The first MA elects the members of the CB members and will function as the kick-off meeting of the consortium, introducing the statutes and communication channels according to M6.4 to M6.6. The office will organize the RSE-Team meetings in collaboration with TA1 and the DaSt Team meetings in collaboration with TA5 (M6.3).

M6.2		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 6.2.1	Consortium statutes and data policy in agreement with institutional policies	🟩									
DS 6.2.2	Kick-off-meeting and Member Assemblies		🟩		🟩		🟩		🟩		🟩

5.6.3 *Measure 6.3: Data Stewardship commissioning and flexible assignments*

Contributions: HHU (Weidtkamp-Peters), UKON/DKFZ (May), TA5

Goals: Implement the Help Desk and the DaSt model

DaSts will form a distributed task force of bioimaging RDM experts that can be deployed to users upon request as described in section 3.5. In this measure, we will install this team in a collaboration with the TA-leads and co-leads and with the support of the NFDI4BIOIMAGE Office. The team will align its goals according to the mission of the initiative in a DaSt kick-off meeting with selected co-applicants and PTPs. The team will agree on roles and procedures. In particular, they will identify relevant national and international data stewardship activities and seek participation by mandating a representative (e.g., to fdm.nrw, RDA, GO FAIR, etc.). At the NFDI level, the DaSt team will take part in activities dedicated to community support, e.g., on topics like DMP tools.

The DaSt concept entails direct on-site support, e.g., to adapt a solution or tool to a specific local environment. These requests will be evaluated and selected by the DaSt team and the SB as described in our operational model. The DaSts are embedded in different TAs and have complementary skills. The choice of DaSt to be commissioned will be guided by the type of question to be solved. We have already identified "pilot"-CUCs for DaSt action during the preparation of the proposal (TA5 M5.4), and expect more to emerge from the community in the course of the project. Therefore we plan "flexible assignments", and request for TA6 about 190 months of DaSt-FTE capacity that are not yet allocated at proposal submission, but with an estimate of ~20% planned for the a priori identified CUCs (Table 5.2, TA5). The DoB, advised by the SB as outlined in our governance, will decide on these flexible assignments which may also be dedicated to collaborations with other consortia on common tools (e.g., GHGA, NFDI4Phys, NFDI4Patho, NFDI4Health). The broad spectrum of know-how of the DaSt-Team distributed over seven locations will enable us to react to a wide range of scenarios covering the expertise from all TAs. By distributing these flexible FTE shares among the DaSts we are able to offer full-time positions to all staff which is essential in the very competitive IT job market and given the limited number of individuals with an explicit job profile in RDM.

M6.3		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 6.3.1	DaSt team formed and representative for the Community Board elected	🟩									
DS 6.3.2	Help Desk installed	🟩	🟩								
DS 6.3.3	Annual reports on data stewardship activity			🟩		🟩		🟩		🟩	

5.6.4 *Measure 6.4: Communication - digital tools*

Contributions: HHU (Weidtkamp-Peters), UKON/DKFZ (May), GerBI-GMB (Hanne), TA5

Goal: Provide a framework for sustained and efficient information exchange inside the consortium and with the outside world

To unite our heterogenous group of stakeholders around the goals of the initiative we must ensure efficient, inclusive and transparent communication among all consortial members. We will create a virtual project space to share and collaboratively produce documents with version control, enable chat groups, use a common calendar and agile collaboration task boards. We will adopt and further develop resources provided *in kind* and tested by GerBI-GMB during the proposal preparation phase. We will elaborate a communication agreement detailing generic and specific information flow for all members. The ambiguous use of technical terms in the different areas of specialization of our members can significantly impair understanding, and must be minimized. To this end, we will leverage the wiki resource of the DFG infrastructure project “I3D:bio” (sec 3.1) to build a glossary for mutual education on the most relevant terms and concepts from all task areas. Communication pitfalls will be discussed and analysed, and the resource adjusted as needed.

A main instrument of communication serving both internal and external purposes will be our website ndfi4biomage.de. Its design and features will be tailored to the needs of the individual TAs for sharing information, providing services (see access below) and requesting feedback. These tasks will be performed together with TA5 as part of its community integration effort. The website will define the digital presence of the NFDI4BIOIMAGE and thus play a crucial role for the global visibility of our initiative and of the future infrastructure. To maximize this visibility we will also exploit social media like Twitter or LinkedIn which require dedicated capacity and expertise. We already share materials for outreach on open source platforms like Zenodo, figshare or GitHub.

M6.4		Y1		Y2		Y3		Y4		Y5	
DS 6.4.1	Internal communication framework for the consortium	▨	▨								
DS 6.4.2	Consortium glossary on technical terms and concepts	▨	▨								
DS 6.4.3	Adoption and improvement of nfdi4bioimage.de as central web-portal	▨	▨								
DS 6.4.4	Adjustment and refinement of website features and overall digital presence throughout the project phase as needed.			▨	▨	▨	▨	▨	▨	▨	▨

5.6.5 *Measure 6.5: Communication - groups*

Contributions: HHU (Weidtkamp-Peters), DKFZ/UKON (May), EMBL (Keppler)

Goal: Devise a comprehensive communication strategy that includes all relevant stakeholder groups

At project start we will define a plan for interacting with all relevant stakeholder groups in regular intervals depending on their degree of participation in and closeness to NFDI4BIOIMAGE.

Within the consortium, communication among members working on defined tasks using our digital tools follows its own dynamics and occurs if necessary on a daily basis and may include involving non-members who are informally consulted, e.g., partners in central IT-facilities or RDM departments. Regular online or in person meetings will be set by TA6 in proximity of the project milestones, in addition to the aforementioned meetings of the RSE and DaSt team meetings.

The closest external communication partners are those NFDI consortia that have potential interfaces with NFDI4BIOIMAGE. We have already identified selected topics of common interest with a number of them, and describe distinct bilateral collaborations in this proposal (see sec 3). We expect further topics to arise in the course of the project for which we will conceive new collaborative work packages. To this end we will be in regular, at least quarterly exchange with individual NFDI consortia, will invite members from other consortia to our communication channels, and organize joint meetings and workshops (TA5). Where appropriate, we will expand these interactions to address defined NFDI cross-cutting topics (M6.7).

The communication strategy includes reaching out, as a further group, to the existing manifold national and international initiatives on bioimaging and generic RDM to draw on their know-how and pursue common interests (sec 3.3). Working on conversion tools for proprietary microscopy data formats towards a unified open format supported both by academia and industry could be one joint action of NFDI4BIOIMAGE (with TA1), and international players like GBI, QUAREP-LiMi and the EuBI Industry Board. In this context, the networking coordinator at DKFZ who will implement the communication strategy of NFDI4BIOIMAGE supported by the project leadership at HHU and TA5, will take advantage of the close proximity of EMBL and build upon the recently stipulated MoU between the two institutions.

The networking coordinator will identify and collect information about important conferences to be attended either by him/herself or NFDI4BIOIMAGE members with the relevant expertise. He/she will propose a plan for presenting the results of the consortium to the international community to be discussed in SB. He/she will also ensure that all networking activities are properly communicated back to the consortium so that no opportunity for new collaborations is missed.

With our strong commitment to inclusiveness and communication we foresee making a significant contribution as a multiplicator of the NFDI and its key values both within our community as well on the international stage.

M6.5		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 6.5.1	Exchange with other NFDI consortia										
DS 6.5.2	Exchange with national and international bioimaging initiatives										

5.6.6 *Measure 6.6: Communication - access*

Contributions: HHU (Weidtkamp-Peters), UKON/DKFZ (May), UGOE

Goals: Promote and monitor the access to tools, services and avenues of participation

As a future infrastructure, NFDI4BIOIMAGE must account for mechanisms of access to all its products from its initial conception. These include not only services and tools, but also the present and future know-how that must be accessible at all times for all arising requests concerning bioimage data management. To meet the first requirement, all developed software is shared with the bioimaging community via common open-source platforms (e.g. GitHub). Secondly, our Help Desk is a low-barrier entry point that facilitates access and informal, direct interactions with the DaSt team and access to their know-how (TA5 and M6.3). The definition and assignment of clear maintenance roles during and after the set-up of the infrastructure will ensure availability of the products of the NFDI4BIOIMAGE in the long run. Consistent access policies, both to bioimaging research data and software tools, will be defined and implemented in agreement with national and international practices. The NFDI4BIOIMAGE Office will collaborate with and support all consortium members with regard to legal issues in this context including issues regarding software licenses, open access publications, collaboration agreements with external platforms, or GDPR-related questions. Handling these aspects properly requires significant time resources which is often a bottleneck for individual researchers, that shall be addressed in NFDI4BIOIMAGE with dedicated staff time. Several of these issues might also require interaction with other NFDI consortia or the NFDI directorate as part of M6.7.

A particular use case for access is further provided by the Flamingo project for shippable light-sheet microscopes, which was successfully tested in the US and will be implemented @UGOE for Germany and the wider European area beginning in 2022. While during the set-up phase, UGOE will contribute to the file formats and required metadata for Flamingo imaging (TA1), it will collaborate with TA6 for exploiting the NFDI4BIOIMAGE network as a preferred distribution channel of Flamingo instruments. The detailed procedures will be defined in the course of the project, and are an exemplary test case for granting access, in particular to the NFDI4BIOIMAGE services that will be of interest to the Flamingo user community. Furthermore, this collaboration has the potential to showcase how FAIR principles can be implemented into bioimaging projects from the very start, as already the assembly of the Flamingo microscope will require documentation including instrument metadata and the validation of instrument performance in calibration measurements to ensure reproducibility. Being grounded in the concept of knowledge sharing, reproducibility and standardization the Flamingo project is also a dissemination action par excellence.

High visibility and clear access options will also serve to attract interest groups whose links to the field of (bioimaging) research data management may not be obvious yet but bear high potential. These could come from all areas of science and also the public. We will ensure accessibility by explicitly promoting our consortium as a contact site for bioimage data management topics providing clear contact information in any information material (e.g. position papers, booklets etc.) which we will widely distribute. We aim to organize two public conferences (year 3 and year 5) to present the consortium and engage in an open exchange about bioimaging RDM with the international community. These activities intend to stimulate the potential onboarding of new consortium members in the course of the project (see sec 3.4).

M6.6		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 6.6.1	Clarify on legal aspects of sharing of materials (with M6.7)		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
DS 6.6.2	Prepare information material for different audiences			🟩	🟩	🟩					
DS 6.6.3	Define the formal workflow for onboarding of new consortium members			🟩							
MS2-TA6	First NFDI4BIOIMAGE Conference					🟩					
MS4-TA6	Second NFDI4BIOIMAGE Conference									🟩	

5.6.7 *Measure 6.7: Embedding into the NFDI and coordination of activities on cross-cutting topics*

Contributions: HHU (Weidtkamp-Peters), UKON/DKFZ (May), GerBI-GMB (Hanne), TA1, TA2, TA3, TA4, TA5

Goal: Collaboration of NFDI4BIOIMAGE with NFDI consortia and with NFDI sections on cross-cutting topics

This measure addresses the interaction of NFDI4BIOIMAGE with the NFDI as a whole, broken down into the four cross-cutting sections which have been identified by established NFDI consortia in the second strategy workshop[88].

As a key starting point, this measure will share and compare common **training and education concepts** within the NFDI. Here, we can share our experiences with integrating bioimaging RDM into study programs, in TA5, the widely acknowledged continuing education program of GerBI-GMB, or even at the undergraduate level (HHU). Our experience is that if RDM techniques are introduced to the earliest career stage students, these young researchers cannot imagine how to

---

[88] https://www.nfdi.de/querschnittsthemen-gemeinsam-bearbeiten-einrichtung-von-vier-sektionen/

manage their data without those tools. They will independently request RDM in any new environment they work in, making such training the cheapest possible investment.

Further, we would like to represent both provider and user perspectives in the **ethical, legal and social aspects** section. Providers, such as the TA2 participants and thereby indirectly de.NBI, must define “Terms of Use” (ToS) stipulating the conditions for running infrastructure and properly managing the expectations of users. Similarly, within the NFDI, data and services become distributed with service operators and users based at different universities. This currently requires data processing agreements between participating universities. This could be a key topic of the NFDI legal cross-cutting section.

At the same time, much of our user base is composed of scientific core facilities in Germany, which have their own needs. Such core facilities serve as hubs for starting standardization processes as bottom-up endeavours that take into account the needs and feedback of a broad user community. We believe that core facilities can deliver important impulses for shaping the NFDI as a user-oriented infrastructure and consider them not mere support structures but as central partners in conducting research. A cultural change is possible with the definition of new academic job profiles, adequate in salary and perspectives to the role of research infrastructures. With our data stewardship model, we propose such a new function that needs to be substantiated within the NFDI as a whole. This topic must be discussed at the level of the institutions. We see NFDI4BIOIMAGE in a leading role in this context, as bioimaging is a very widespread technology that is present at almost any research institution in Germany.

The development of **common (technical) infrastructures** for sustainable services is a key goal of TA2 and one on which we would like to collaborate within the respective Section with an eye toward enabling the reproducible deployment of containers and VMs across infrastructures. More immediately, core facilities frequently need to offer to their users virtual desktop environments, which are mainly available as proprietary solutions so far. As part of TA2, we plan to develop an open Desktop as a Service (DaaS) with potential widespread usability for the whole NFDI. The project is led by D. v. Suchodoletz, spokesperson of DataPLANT and member in NFDI4BIOIMAGE with TA2 participating in the effort via TA2. However, identifying and implementing basic services must be an inclusive, shared effort by the NFDI as a whole, therefore this measure proposes representing the highly graphical needs of the bioimaging community.

Finally, **(Meta)data, findability, terminology and provenance** is the most relevant cross-cutting topic for our consortium and represents one of our core competencies. NFDI4BIOIMAGE is committed to maximizing the re-use of technology across all research infrastructures. Our choice, for example, of RDF and by extension JSON-LD as the primary representation of metadata is precisely intended to increase the interaction with other NFDI partners, based on preliminary presentations and discussions.

Our efforts to advance next-generation file formats (NGFF) can have even more impact. NGFFs are a data structure with potential widespread use for a multitude of other data types, allowing harmonization of data across the borders of disciplines or methods. As mentioned in the RDM strategy, the bioimaging domain is particularly crippled by the lack of a common exchange format. The solution we have chosen holds potential for widespread cloud-accessibility throughout the NFDI, an option that we aim to explore with other partners of the Section as a basic service. We believe similarly that a common but generic, n-dimensional cloud-optimized format like NGFF would reduce redundant developments. To champion this concept we have defined the role and will install a “Data Formats Ambassador”, a data specialist with excellent networking skills who will discuss and disseminate the idea of and the current knowledge behind formats for remotely viewable, large scale cloud data. He/she will coordinate meta(data) and infrastructure plans and deliverables with the relevant sections of the NFDI and will work toward a strong position statement to accelerate uptake, leading to more development that will drive usability.

M6.7		**Y1**		**Y2**		**Y3**		**Y4**		**Y5**	
DS 6.7.1	Propose early career learning packages					🟩					
DS 6.7.2	Define the specialized needs of scientific core facilities				🟩						
DS 6.7.3	Compare data stewardship models with other consortia			🟩							
DS 6.7.4	Evaluate DaaS as a cross-cutting concern						🟩				
DS 6.7.5	Establish Data Formats Ambassador role	🟩									
DS 6.7.6	Represent next-generation file formats		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
DS 6.7.7	Regular coordination with the sections	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩

5.6.8 *Dependencies and Interactions*

As the networking and coordination Task Area, TA6 will serve as the hub for sharing and exchanging information between all Task Areas to ensure constant and efficient information flow within the consortium. TA6 will be in close contact with all TAs about administrative aspects of the activities to support the project progress. This Task Area provides an essential basis, on which all TAs can depend, in particular during the early project phase when the framework for interactions and exchange needs to be installed according to the measures outlined above. Together with TA5 we will prepare all events and outreach activities of the consortium. The instrument of data stewardship and flexible assignments will allow us to collaborate with other NFDI consortia on new projects we will identify during the course of the project. Contact and exchange with the NFDI association and the directorate will be coordinated by the NFDI4BIOIMAGE Office. Contributions by individual Task Areas, e.g., to cross-cutting topics sections of the NFDI will be coordinated by TA6. Overall TA6 aims to survey and manage all internal and external interactions between the stakeholders of NFDI4BIOIMAGE and the research community to ensure the project progress and success.

5.6.9 *Risk and mitigation*

DaSt support could be abused to replace lacking institutional support: the goal of the DaSt model is to render users autonomous with respect to their bioimage data management needs and requires long-term institutional commitment. We mitigate this risk through our operational model which defines how DaSt support is commissioned. That includes review by the SB with support from the BoDs and the NFDI4BIOIMAGE Office who will help in assessing the embedding of the DaSt in the local structures.

Bringing the tools developed by NFDI4BIOIMAGE to the user (e.g., via DaSt support) might be hampered by reluctance to share project details which is needed to support effectively. The mitigation consists in elaborating confidentiality agreements between the DaSt and the CUCs, while ensuring that the RDM-specific outcomes of the project and eventual published data are made accessible according to the FAIR principles.

Consortial partners might not comply with the governance, and, e.g., employ funding for goals which are not in line with the initiative, do not take part in the governance bodies, etc. The mitigation is given by the fact that firstly, all members at project start have seen and agreed on the governance model. The governance will be officially ratified at the kick-off meeting of NFDI4BIOIMAGE. Secondly, the governance will be compatible with institutional regulations and policies and will be ensured by the project management.

The consortium members might not adequately support the office by participating in surveys and evaluations. This risk will be mitigated by providing clear communication agreements to stimulate the commitment of all consortium members.

Parallel and/or diverging developments for bioimage RDM might exist in different consortia. We mitigate this risk not only by assigning one FTE to networking but by reserving also flexible DaSt capacity to the task of cross-consortial collaboration.

5.7 Milestones

Based on the measures defined in the work programme by each TA we have set the following milestones for the project as a whole:

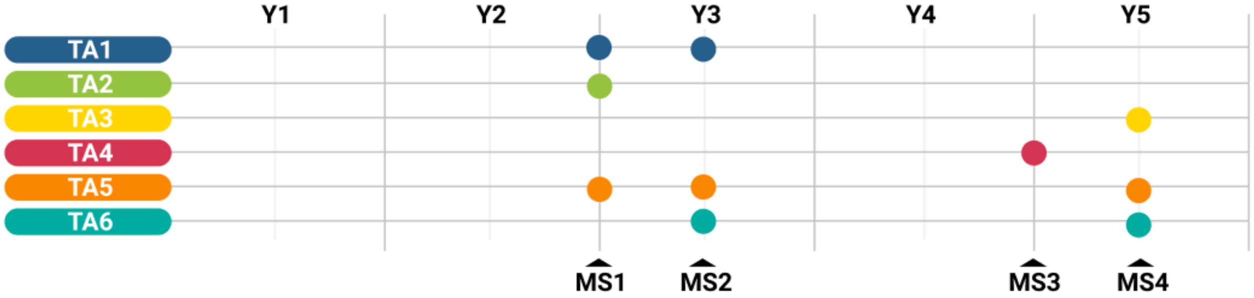

Figure 13: Overview of the project milestones, consisting of the indicated milestone parts from the Task Areas.

Milestone 1:	White paper on the FAIR-IO concept published (MS1-TA1), production services for NFDI4BIOIMAGE available (MS1-TA2) and search index for bioimage RDM training materials implemented (MS1-TA5)
Milestone 2:	Release of RFC process document (MS2-TA1), first NFDI4BIOIMAGE conference (MS2-TA6), and publication of survey results and search index statistics (MS2-TA5)
Milestone 3:	SOPs for FAIR usage of image analysis workflows available (MS3-TA4)
Milestone 4:	Second NFDI4BIOIMAGE conference (MS4-TA6), preview of the over-arching database for multimodal metadata (MS4-TA3), and publication of the search index statistics and survey results (MS4-TA5)

6 Appendix

6.1 Acknowledgements

We would like to thank Lutz Becks, Magdalene Cyra, Matthias Fingerhuth, Christian Busse, Matthias Günther, Joachim Rädler, Hans-Günther Döbereiner, Peter Boor, Rainer Röhrig, Petra Ritter, Phil Gribbon, Moritz Gerstung, Felix Hartmann, Jutta Steinkötter, Rainer Kurre, Frauke Leitner, and Johanna Bischof for networking, exchange, use case provision, and their support in the process of forming the consortium and planning the proposal.

6.2 Bibliography

1. Power, R.M. and J. Huisken, *Putting advanced microscopy in the hands of biologists.* Nat Methods, 2019. **16**(11): p. 1069-1073.
2. Franke, S., et al., *Plasma-MDS, a metadata schema for plasma science with examples from plasma technology.* Sci Data, 2020. **7**(1): p. 439.
3. Chen, J., et al., *The Allen Cell and Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images.* bioRxiv, 2020: p. 491035.
4. Viana, M.P., et al., *Robust integrated intracellular organization of the human iPS cell: where, how much, and how variable.* bioRxiv, 2021: p. 2020.12.08.415562.
5. *Crystallography: Protein Data Bank. .* Nature New Biology, 1971(233): p. 223.
6. Sarkans, U., et al., *REMBI: Recommended Metadata for Biological Images-enabling reuse of microscopy data in biology.* Nature Methods, 2021.
7. Kunis, S., et al., *MDEmic: a metadata annotation tool to facilitate management of FAIR image data in the bioimaging community.* Nat Methods, 2021. **18**(12): p. 1416-1417.
8. Nelson, G., et al., *QUAREP-LiMi: A community-driven initiative to establish guidelines for quality assessment and reproducibility for instruments and images in light microscopy.* arXiv preprint, 2021.
9. Boehm, U., et al., *QUAREP-LiMi: a community endeavor to advance quality assessment and reproducibility in light microscopy.* Nat Methods, 2021.
10. Berg, S., et al., *ilastik: interactive machine learning for (bio)image analysis.* Nat Methods, 2019. **16**(12): p. 1226-1232.
11. Ellenberg, J., et al., *A call for public archives for biological image data.* Nat Methods, 2018. **15**(11): p. 849-854.
12. Haase, R., et al., *CLIJ: GPU-accelerated image processing for everyone.* Nat Methods, 2020. **17**(1): p. 5-6.
13. Levet, F., et al., *Developing open-source software for bioimage analysis: opportunities and challenges.* F1000Res, 2021. **10**: p. 302.
14. Jamali, N., et al., *2020 BioImage Analysis Survey: Community experiences and needs for the future.* Biological Imaging, 2022. **1**: p. e4.
15. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
16. Martins, G.G., et al., *Highlights from the 2016-2020 NEUBIAS training schools for Bioimage Analysts: a success story and key asset for analysts and life scientists.* F1000Res, 2021. **10**: p. 334.
17. Rigano, A., et al., *Micro-Meta App: an interactive tool for collecting microscopy metadata based on community specifications.* Nat Methods, 2021. **18**(12): p. 1489-1495.
18. Ryan, J., et al., *MethodsJ2: a software tool to capture metadata and generate comprehensive microscopy methods text.* Nat Methods, 2021. **18**(12): p. 1414-1416.
19. Schmidt, C., et al., *Research data management for bioimaging: the 2021 NFDI4BIOIMAGE community survey.* F1000Research, 2022. **11**: p. 638.
20. Linkert, M., et al., *Metadata matters: access to image data in the real world.* J Cell Biol, 2010. **189**(5): p. 777-82.

21. Moore, J., et al., *OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies.* Nat Methods, 2021. **18**(12): p. 1496-1498.
22. Li, S., et al., *Metadata management for high content screening in OMERO.* Methods, 2016. **96**: p. 27-32.
23. Burel, J.-M., et al., *Publishing and sharing multi-dimensional image data with OMERO.* Mammalian Genome, 2015. **26**(9-10): p. 441-447.
24. Barillari, C., et al., *openBIS ELN-LIMS: an open-source database for academic laboratories.* Bioinformatics, 2016. **32**(4): p. 638-640.
25. Bauch, A., et al., *openBIS: a flexible framework for managing and analyzing complex data in biology research.* BMC Bioinformatics, 2011. **12**(1).
26. Rubens, U., et al., *Cytomine: Toward an Open and Collaborative Software Platform for Digital Pathology Bridged to Molecular Investigations.* PROTEOMICS – Clinical Applications, 2018. **13**(1).
27. Begley, C.G. and L.M. Ellis, *Drug development: Raise standards for preclinical cancer research.* Nature, 2012. **483**(7391): p. 531-3.
28. Errington, T.M., et al., *An open investigation of the reproducibility of cancer biology research.* eLife, 2014. **3**.
29. Williams, E., et al., *The Image Data Resource: A Bioimage Data Integration and Publication Platform.* Nat Methods, 2017. **14**(8): p. 775-781.
30. Goldberg, I.G., et al., *The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging.* Genome Biol, 2005. **6**(5): p. R47.
31. Degtyar, V.E., et al., *Rapid Global Fitting of Large Fluorescence Lifetime Imaging Microscopy Datasets.* PLoS ONE, 2013. **8**(8).
32. Besson, S., et al., *Bringing Open Data to Whole Slide Imaging*, in *Digital Pathology*. 2019. p. 3-10.
33. Kume, S., et al., *Development of Semantic Web-Based Imaging Database for Biological Morphome*, in *Semantic Technology*. 2017. p. 277-285.
34. Dekker, J., et al., *The 4D nucleome project.* Nature, 2017. **549**(7671): p. 219-226.
35. Wittner, R., et al., *ISO 23494: Biotechnology – Provenance Information Model for Biological Specimen And Data*, in *Provenance and Annotation of Data and Processes*. 2021. p. 222-225.
36. Hammer, M., et al., *Towards community-driven metadata standards for light microscopy: tiered specifications extending the OME model.* Nat Methods, 2021. **18**(12): p. 1427-1440.
37. Sofroniew;, N., et al. *napari/napari: 0.4.3rc3 (v0.4.3rc3).* 2021. DOI: 10.5281/zenodo.4439652.
38. Vergara, H.M., et al., *Whole-body integration of gene expression and single-cell morphology.* Cell, 2021. **184**(18): p. 4819-4837.e22.
39. Manz, T., et al., *Viv: Multiscale Visualization of High-ResolutionMultiplexed Bioimaging Data on the Web.* 2020.
40. Palagi, P.M., et al., *Fostering accessible online education using Galaxy as an e-learning platform.* PLOS Computational Biology, 2021. **17**(5).
41. Grüning, B., et al., *Bioconda: sustainable and comprehensive software distribution for the life sciences.* Nature Methods, 2018. **15**(7): p. 475-476.
42. da Veiga Leprevost, F., et al., *BioContainers: an open-source and community-driven framework for software standardization.* Bioinformatics, 2017. **33**(16): p. 2580-2582.
43. Grüning, B., et al., *Practical Computational Reproducibility in the Life Sciences.* Cell Systems, 2018. **6**(6): p. 631-635.
44. Kuhn Cuellar, L., et al., *A data management infrastructure for the integration of imaging and omics data in life sciences.* BMC Bioinformatics, 2022. **23**(1): p. 61.
45. Schapiro, D., et al., *MITI minimum information guidelines for highly multiplexed tissue images.* Nat Methods, 2022. **19**(3): p. 262-267.
46. Gorgolewski, K.J., et al., *The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments.* Sci Data, 2016. **3**: p. 160044.
47. Schindelin, J., et al., *The ImageJ ecosystem: An open platform for biomedical image analysis.* Mol Reprod Dev, 2015. **82**(7-8): p. 518-29.

48. de Chaumont, F., et al., *Icy: an open bioimage informatics platform for extended reproducible research.* Nat Methods, 2012. **9**(7): p. 690-6.
49. McQuin, C., et al., *CellProfiler 3.0: Next-generation image processing for biology.* PLoS Biol, 2018. **16**(7): p. e2005970.
50. Gupta, A., et al., *Deep Learning in Image Cytometry: A Review.* Cytometry A, 2019. **95**(4): p. 366-380.
51. Berthold, M.R., et al., *KNIME: The Konstanz Information Miner*, in *Data Analysis, Machine Learning and Applications*. 2008. p. 319-326.
52. Ouyang, W., et al., *ImJoy: an open-source computational platform for the deep learning era.* Nat Methods, 2019. **16**(12): p. 1199-1200.
53. Weigert, M., et al., *Content-aware image restoration: pushing the limits of fluorescence microscopy.* Nat Methods, 2018. **15**(12): p. 1090-1097.
54. Gomez-de-Mariscal, E., et al., *DeepImageJ: A user-friendly environment to run deep learning models in ImageJ.* Nat Methods, 2021. **18**(10): p. 1192-1195.
55. von Chamier, L., et al., *Democratising deep learning for microscopy with ZeroCostDL4Mic.* Nat Commun, 2021. **12**(1): p. 2276.
56. Schmidt, C. and E. Ferrando-May, *NFDI4BIOIMAGE - An Initiative for a National Research Data Infrastructure for Microscopy Data*, in *E-Science Tage - Share your Research Data*. 2021: Heidelberg.
57. Waithe, D., *Summary of two questionnaires designed to understand the research climate for Bioimage Analysts in the UK between 2016-2019.* F1000Research, 2021. **10**.

6.3 Abbreviations

µm	micrometer
3D	3 dimensional
ACSS	Allen Cell and Structure Segmenter
AI	Artificial Intelligence
aka	also known as
ALMF	Advanced Light Microscopy Facility
ALU-FR	University of Freiburg
AMBIOM	Analysis of Microscopic Biomedical Images
AMD	Advanced Micro Devices, Inc
API	Application Programming Interface
API	application programming interfaces
ARC	Annotated Research Context
ASB	Applied Systems Biology
BBMRI-ERIC	Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium
BEP	BIDS extension proposal
BIA	BioImage Archive
BIC	Bioimaging Center
BIDS	Brain Imaging Data Structure
BINA	BioImaging North America
BioDIP	Biopolis Dresden Imaging Platform
BioIT	bioinformatics training and community development program
BLAST	Basic Local Alignment Search Tool
BLV	Bund-Länder-Vereinbarung
BMBF	Bundesministerium für Bildung und Forschung
BoD	Board of Directors
bw2FDM	baden-württembergische Begleit- und Weiterentwicklungsprojekt für Forschungsdatenmanagement
bwFORCluster	Baden-Württemberg Cluster
BY-COVID	Beyond COVID, EU Project
CAi	Center for Advanced Imaging
CB	Community Board
CC-0	Creative Commons Public Domain License
CC-BY	Creative Commons Attribution License
CECAD	Cluster of Excellence on Cellular Stress Response in Aging-Associated Diseases
CellNanOs	Centre for Cellular Nanoanalytics Osnabrück
CEPLAS	Cluster of Excellence on Plant Sciences
CfConventions	Climate and Forcast Conventions
ChEMBL	Database of chemistry compounds
CI/CD	Continuous Implementation / Continuous Deployment

CIM	Cells in Motion Interfaculty Centre
CIO	Chief Information Officer
CLI	command-line interface
CNI	Combinatorial Neuroimaging Core Facility
CoE	clusters of excellence
CONP	Canadian Open Neuroscience Platform
COST	European Cooperation in Science and Technology
COVID	coronavirus disease
CPU	central processing unit
CRC	collaborative research centre
CRU	Clinical Research Unit
CT	computed tomography
CUC	community use case
CWL	Common Workflow Language
CZI	Chan Zuckerberg Initiative
DaSt	NFDI4BIOMAGE Data Steward
de.NBI	German Network for Bioinformatics Infrastructure
de.NBI-epi	de.NBI epigenetics project
DevOps	software development (Dev) and IT operations (Ops)
DFG	Deutsche Forschungsgemeinschaft
DGZ	Deutsche Gesellschaft für Zellbiologie
DICOM	Digital Imaging and Communications in Medicine
DINI e.V.	Deutsche Initiative für Netzwerkinformation
DKFZ	German Cancer Research Center
DMP	data management plan
DNA	deoxyribonucleic acid
DOI	Digital Object Identifier
DS	Deliverable
EBI	European Bioinformatics Institute
EBRAINS	EBRAINS Research Infrastructure
elabFTW	Open source, e-lab notebook "for the world"
ELIXIR	ELIXIR Research Infrastructure
ELLS	European Learning Laboratory for the Life Sciences
ELMI	European Light Microscopy Initiative
ELN	electronic lab notebook
EMBL	European Molecular Biology Laboratory
EMBL-EBI	European Molecular Biology Laboratory - European Bioinformatics Institute
EOSC	European Open Science Cloud
EOSC-life	EOSC Research Infrastructure for the life sciences
EPICS	Experimental Physics and Industrial Control System
ERIC	European Research Infrastructure Consortium
ERUA	European Reform University Alliance
EuBI	Euro-BioImaging
EXC	Cluster of Excellence
EXC CIBBS	Cluster of Excellence - Centre for Integrative Biological Signalling Studies
FAIR	findability, accessibility, interoperability, and reusability
FAIR-IO	FAIR Image Objects
FAIRsFAIR	Fair Practices project in Europe
FASTQ	text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding quality scores
fdm.rnw	Landesinitiative für Forschungsdatenmanagement – fdm.nrw
FDO	FAIR Digital Objects
Fiji	Fiji Is Just ImageJ
FLIM	fluorescence life-time imaging
FPGA	Field Programmable Gate Array
FTE	Full Time Equivalent
FZJ	Forschungszentrum Jülich
GB	gigabyte
Gb/s	gigabit per second
GBI	Global BioImaging
GBM	Gesellschaft für Biochemie und Molekularbiologie
GCS	Google Cloud Storage
GDPR	General Data Protection Regulation
GerBI-GMB	German BioImaging - Gesellschaft für Mikroskopie und Bildanalyse e.V.
gGenestackODM	enterprise software solution for data organization
GHGA	German Human Genome-Phenome Archive
GPU	graphical processing unit
GTN	Galaxy Training Network
GUI	Graphical user interface
HCS	High-content screening

HDF5	Hierarchical Data Format
HHU	Heinrich Heine University Düsseldorf
HIP	Helmholtz Imaging Platform
HIPO	Heidelberg Center for Personalized Oncology
HKI	Leibniz Institute for Natural Product Research and Infection Biology, Hans
HMC	Helmholtz Metadata Consortium
HPC	High Performance Computing
HSI	Hyperspectral Imaging
HTTP	Hypertext Transfer Protocol
I/O	Input / Output
I3D:bio	Information Infrastructure for BioImage Data
IaaS	Infrastructure as a Service
iBIOS	integrated Bioimaging facility
ICCAS	Innovation Center Computer Assisted Surgery
IDE	Integrated Development Environment
IDR	Image Data Resource
INM1	Institute of Neuroscience and Medicine
INP	Leibniz Institute for Plasma Science and Technology
INPTDAT	Interdisciplinary Plasma Technology Data Platform
IPHT	Leibniz Institute of Photonic Technology
ISA	„Investigation - Study - Assay“ (data model description acronym)
ISAS	Leibniz Institute for Analytical Sciences
ISO	International Organization for Standardization
IT	Information Technology
ITCF	IT Core Facility
JGU	Johannes Gutenberg University Mainz
JSC	Jülich Supercomputing Center
JSON	Java Script Object Notation
JSON-LD	JSON for linking data
KIM	Communication, Information, and Media Center
KNIME	Konstanz Information Mine
LERU	League of European Research Universities
LIC	Life Imaging Center
LIMS	Laboratory information management system
LIN	Leibniz Institute for Neurobiology
livMatS	Living, Adaptive and Energy-autonomous Materials Systems
LPI	Leibniz Center for Photonics in Infection Research
M	Measure
MA	Member Assembly
MDC	Max-Delbrück Center Berlin
MDEmic	MetaData Editor for microscopy
ML	maschine learning
mmDS	multimodal data structure
MPIEB	Max Planck Institute for Evolutionary Biology
MPIEM	Max Planck-Institute for Experimental Medicine
MSI	Mass spectrometry imaging
MUGDAT	multimodal graph database
NEUBIAS	Network of European BioImage Analysts
NFDI	Nationale Forschungsdateninfrastruktur
NGFF	next-generation file format
NHR	Nationales Hochleistungsrechnen
NIH	National Institutes of Health
OME	Open Microscopy Environment
OME-OWL	OME- Web Ontology Language
OMERO	OME Remote Objects
OME-TIFF	OME Tag Image File Format
OME-XML	OME Extensible Markup Language
OpenAIRE	Open Access Infrastructure for Research in Europe
openBis	open Biology Information System
OS	Operating System
OSI	open source initiative
OWL	Web Ontology Language
PaaS	Platform as a Service
PACS	picture archiving and communication system
PB	petabyte
PDB	Protein Data Bank
PFF	proprietary file formats
Plasma-MDS	Plasma metadata schema
PoL	Physics of Life
PTP	Participant

QBiC	Quantitative Biology Center Tübingen
QC	Quality Control
QPTDat	Quality Assurance and Linking of Research Data in Plasma Technology
QUAREP-LiMi	Quality Assessment and Reproducibility of Images in Light Microscopy
RAM	Random-Access Memory
RDA	Research Data Alliance
RDF	Resource Description Framework
RDM	Research Data Management
RDM4mic	Research Data Management for Microscopy
RDMkit	Research Data Management Kit
REMBI	recommendations on metadata for biological imaging
RFC	Request for Comment
RNA	Ribonucleic acid
RO-Crate	the marriage of Research Objects with DataCrate
ROI	Region of Interest
RSE	Research Software Engineer
S3	Simple Storage Service
SaaS	Software as a Service
SB	Steering Board
SciComp	Scientific Computing
SDS@hd	Scientific data storage, Heidelberg
SEEK	web-based cataloguing and commons platform, for sharing heterogeneous scientific research datasets
SLURM	Simple Linux Utility for Resource Management
SOP	Standard Operating Procedures
SPARQL	SPARQL Protocol and RDF Query Language
SWATE	Excel Add-In for annotation of experimental data and computational workflows
TA	Task Area
TB	terabyte
TCK	Technology-compatibility kit
TIaaS	Training Infrastructure as a Service
TIFF	Tagged Image File Format
TIRAMISU	Time resolved Raman- and metabolic Imaging spectroscopy unit
ToS	Terms of Use
TUD	Technische Universität Dresden
TUDo	Technical University of Dortmund
UB	University Library
UDUE	Universität Duisburg-Essen
UFZ	Helmholtz-Center for Environmental Research
UGOE	Georg-August-University of Göttingen
UHD	University Heidelberg
UI	user interface
UKON	University of Konstanz
ULB	University and State Library
UL-MED	Leipzig University Medical Faculty - Innovation Center Computer-Assisted Surgery
UoC	University of Cologne
UOS	University of Osnabrück
URL	universal resource location
URZ	University Computing Centre "Universitätsrechenzentrum"
USB	Universal Serial Bus
UX	User Experience
VDI	Virtual Desktop Infrastructure
VM	Virtual Machine
VNC	Virtual Network Computing
W3C	World Wide Web Consortium
WAN	Wide Area Network
WWU	University of Münster
X11	X Window System
XNAT	extensible open-source imaging informatics software platform dedicated to imaging-based research
YERUN	Young European Research Universities Network
Zarr	a format for the storage of chunked, compressed, N-dimensional arrays
ZeroCostDL4Mic	free and open-source toolbox for Deep-Learning in microscopy
ZIH	Zentrum für Informationsdienste und Hochleistungsrechnen
ZIM	Centre for Information and Media Services

6.4 Author contributions

Author contribution categories comply with the CRediT initiative (https://credit.niso.org/).

Author name	Contribution (CRediT)	Other / comment
Bauer, Pavol	Conceptualization	
Baumann, Martin	Writing – original draft	
Becker, Markus M.,	Writing – review and editing	
Becks, Lutz	Endorsement	
Beretta Carlo	Endorsement	
Blank-Burian, Markus	Conceptualization, Writing – original draft, Writing – review and editing, Resources	
Bocklitz, Thomas	Conceptualization, Writing – review and editing	
Brilhaus, Dominik	Writing – original draft	
Bumberger, Jan	Conceptualization, Writing – review and editing	
Burel, Jean-Marie	Endorsement	
Chalopin, Claire	Writing – review and editing	
Chen, Jianxu	Endorsement	
Czodrowski, Paul	Writing – original draft, Endorsement	
Dickscheid, Timo	Writing – review and editing	
Dunker, Susanne	Endorsement	Provision of a methodology use case
Ferrando-May, Elisa	Conceptualization, Writing – original draft, Writing – review and editing, Supervision, Funding acquisition	Co-corresponding senior author: elisa.may@dkfz-heidelberg.de
Figge, Marc-Thilo	Conceptualization, Writing – original draft, Writing – review and editing	
Fortmann-Grote, Carsten	Writing – review and editing	
Gerst, Ruman	Writing – original draft	
Grüning, Björn	Conceptualization, Writing – original draft, Writing – review and editing	
Haase, Robert	Conceptualization, Writing – original draft, Writing – review and editing	
Hanne, Janina	Conceptualization, Writing – original draft, Writing – review and editing, Visualization	
Hartley, Matthew	Endorsement	
Heuveline, Vincent	Endorsement	
Huisken, Jan	Endorsement	
Hülsmann, Bastian	Endorsement	
Keppler, Antje	Endorsement	
Kreshuk, Anna	Conceptualization, Writing – original draft, Writing – review and editing	
Kuner, Rohini	Writing – original draft	
Kuner, Thomas	Endorsement	
Kunis, Susanne	Conceptualization, Writing – original draft, Writing – review and editing, Visualization	
Landwehr, Matthias	Writing – original draft	
Leibfried, Andrea	Conceptualization, Writing – original draft	Coordination of the contributions from Heidelberg University
Lohmann, Jan	Endorsement	
Mallm, Jan-Philipp	Conceptualization, Writing – original draft, Writing – review and editing	
Mathur, Aastha	Endorsement	
Meesters, Christian	Endorsement	Provision of an iRODS OMERO integration use case
Mittal, Deepti	Writing – original draft	
Möbius, Wiebke	Endorsement	

Moore, Joshua	Conceptualization, Writing – original draft, Writing – review and editing, Visualization	Corresponding author: joshua.moore@gerbi-gmb.de
Nahnsen, Sven	Endorsement	Provisioning of an openBIS/OMERO intergration use case
Nitschke, Roland	Endorsement	
Pfander, Claudia	Endorsement	Consultation
Rehwald, Stephanie	Endorsement	
Serrano-Solano, Beatriz	Endorsement	
Schauss, Astrid	Writing – original draft	
Schmidt, Christian	Conceptualization, Writing – original draft, Writing – review and editing, Project administration, Visualization	Project Coordinator for the preparation of the proposal
Stöter, Torsten	Conceptualization, Writing – original draft, Writing – review and editing	
Svensson, Carl-Magnus	Conceptualization, Writing – original draft, Writing – review and editing	
Tischer, Christian	Endorsement	
Valencia-Schneider, Monica	Writing – original draft	
Vilardell Scholten, Laura	Endorsement	
Vogl, Raimund	Endorsement, Resources	
von Suchodoletz, Dirk	Writing – original draft	
Weidtkamp-Peters, Stefanie	Conceptualization, Writing – original draft, Writing – review and editing, Supervision	Corresponding senior author: Stefanie.weidtkamp-peters@hhu.de
Zentis, Peter	Writing – original draft	
Zobel, Thomas	Conceptualization, Writing – original draft, Writing – review and editing	
Zuschratter, Werner	Conceptualization, Writing – original draft, Writing – review and editing	