

## INVITED REVIEW

# A practical guide to bioimaging research data management in core facilities

Christian Schmidt<sup>1</sup>  | Tom Boissonnet<sup>2</sup>  | Julia Dohle<sup>3</sup> | Karen Bernhardt<sup>3</sup>  |  
 Elisa Ferrando-May<sup>1,4</sup>  | Tobias Wernet<sup>5</sup>  | Roland Nitschke<sup>5,6</sup>  |  
 Susanne Kunis<sup>3</sup>  | Stefanie Weidtkamp-Peters<sup>2</sup> 

<sup>1</sup>Enabling Technology Department, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>2</sup>Center for Advanced Imaging, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

<sup>3</sup>Center of Cellular Nanoanalytics, Integrated Bioimaging Facility iBiOs, University of Osnabrück, Osnabrück, Germany

<sup>4</sup>Department of Biology, University of Konstanz, Konstanz, Germany

<sup>5</sup>Life Imaging Center, University of Freiburg, Freiburg, Germany

<sup>6</sup>CIBSS and BIOSS – Centres for Biological Signalling Studies, University of Freiburg, Freiburg, Germany

## Correspondence

Christian Schmidt, Enabling Technology Department, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

Email:

[christian01.schmidt@dkfz-heidelberg.de](mailto:christian01.schmidt@dkfz-heidelberg.de)

Stefanie Weidtkamp-Peter, Center for Advanced Imaging, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany.

Email: [stefanie.weidtkamp-peters@uni-duesseldorf.de](mailto:stefanie.weidtkamp-peters@uni-duesseldorf.de)

[uni-duesseldorf.de](http://uni-duesseldorf.de)

## Funding information

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Grant/Award Number: 462231789

## Abstract

Bioimage data are generated in diverse research fields throughout the life and biomedical sciences. Its potential for advancing scientific progress via modern, data-driven discovery approaches reaches beyond disciplinary borders. To fully exploit this potential, it is necessary to make bioimaging data, in general, multidimensional microscopy images and image series, FAIR, that is, findable, accessible, interoperable and reusable. These FAIR principles for research data management are now widely accepted in the scientific community and have been adopted by funding agencies, policymakers and publishers. To remain competitive and at the forefront of research, implementing the FAIR principles into daily routines is an essential but challenging task for researchers and research infrastructures. Imaging core facilities, well-established providers of access to imaging equipment and expertise, are in an excellent position to lead this transformation in bioimaging research data management. They are positioned at the intersection of research groups, IT infrastructure providers, the institution's administration, and microscope vendors. In the frame of German BioImaging – Society for Microscopy and Image Analysis (GerBI-GMB), cross-institutional working groups and third-party funded projects were initiated in recent years to advance the bioimaging community's capability and capacity for FAIR bioimage data management. Here, we provide an imaging-core-facility-centric perspective outlining the experience and current strategies in Germany to facilitate the

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Journal of Microscopy* published by John Wiley & Sons Ltd on behalf of Royal Microscopical Society.

practical adoption of the FAIR principles closely aligned with the international bioimaging community. We highlight which tools and services are ready to be implemented and what the future directions for FAIR bioimage data have to offer.

#### KEYWORDS

bioimaging, data stewardship, FAIR, OMERO, research data management

## 1 | INTRODUCTION: CORE FACILITIES AS ESSENTIAL PARTNERS IN RESEARCH AND DRIVERS OF FAIR IMAGE DATA HANDLING

Imaging core facilities create synergies by centralising procurement, maintenance and accessibility of high-end imaging instruments.<sup>1–3</sup> Many core facilities also perform their own research on method development. Researchers use the core facility and consult with its staff to discuss their experiments.<sup>4–6</sup> Early career scientists especially rely on core facility specialists to learn about microscopy techniques and how to best apply them.<sup>7,8</sup> Besides providing training on microscopes, core facility staff regularly deal with various user questions, for example: How does the microscope choice influence data quality and quantity? How are microscopy files structured? How do you retrieve data from the instrument, and where do you store it? Can images be used in an electronic lab notebook?, etc. Following data acquisition, many users need support in image analysis and software access and preparing data for publication, for which they continue consulting with core facility staff. Concerns include: What information about the imaging experiment must be detailed in a methods section? How do you adequately describe image processing and analysis? How and where do you deposit data as demanded by the publisher? Lastly, helping to recover images in raw formats, for example, from former PhD students, or identifying the imaging setup of a former Postdoc's work are among the challenges faced by core facility staff. Notably, many questions relate to the complex topic of research data management (RDM), a crucial part of research in general. RDM refers to all activities to organise, store, share, document and preserve research data, which are mainly, but not only, digital data nowadays. A professional RDM strategy helps to ensure scientific quality and reliability, protect intellectual property and increase the efficiency of scientific work by ensuring reproducibility and allowing reusability of research data. Research communities have now widely accepted the FAIR guiding principles for RDM.<sup>9,10</sup> Funding agencies and policymakers support them, and they have become part of good scientific practice. In this article, we focus on RDM aspects for handling bioimaging data, especially on adopting the

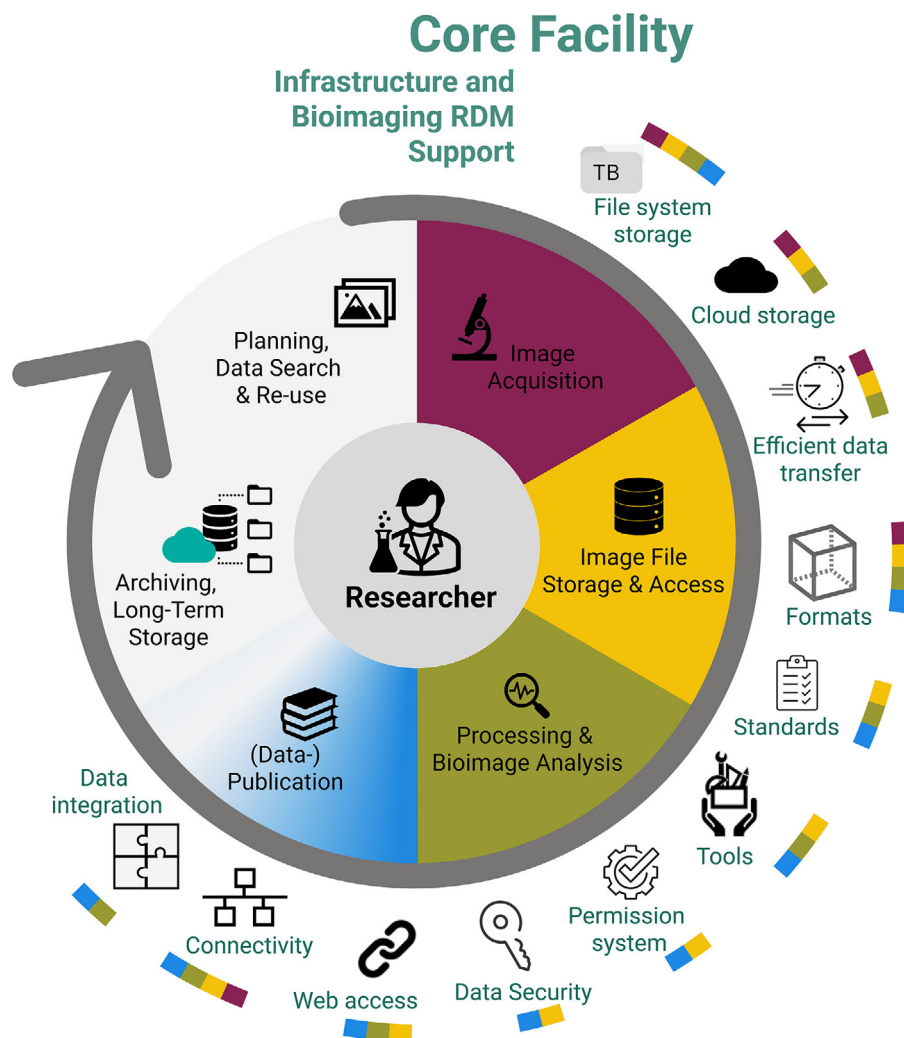
FAIR principles for this data type. While we include generally applicable aspects of RDM throughout the text, we mainly refer to 'bioimaging RDM' or 'bioimage data management' in this article.

The above-mentioned experiences have stimulated core facilities to initiate activities to establish bioimaging RDM practices. One of the first examples in Germany is the open discussion group Research Data Management for Microscopy (RDM4mic), focusing on the data management system OMERO (OME Remote Objects).<sup>11,12</sup> With support from GerBI-GMB, these activities led to additional projects. The Information Infrastructure for BioImage Data project (I3D:bio, <https://www.i3dbio.de>) started with a focus on facilitating the implementation of OMERO at universities and research institutions in Germany.<sup>13</sup> Moreover, more than 20 German institutions formed a consortium dedicated to bioimage data management within the National Research Data Infrastructure (NFDI).<sup>14</sup> This consortium, NFDI4BIOIMAGE, aims at fostering new standards for image data handling and metadata curation (<https://nfdi4bioimage.de>) and is embedded within a network of consortia (<https://nfdi.de>). These and more projects (e.g., FoundingGIDE, <https://cordis.europa.eu/project/id/101130216>) aim to align strategies for bioimaging RDM while accounting for local and discipline-specific needs. Based on the experience from these projects, this article

- summarises current challenges for bioimaging RDM from the perspective of core facilities (PART I),
- provides action items for core facilities to advance their bioimaging RDM capabilities and services (PART II),
- invites to collaboratively drive the cultural change toward expanding and acknowledging the role of core-facility-supported RDM and bioimage data stewards in scientific discovery (PART III).

## 2 | PART I: THE BIOIMAGE DATA LIFE CYCLE

Research data have the potential for scientific discovery beyond its original acquisition purpose when handled according to the FAIR principles.<sup>9,15</sup> To structure all



**FIGURE 1** The research data life cycle as applied to bioimage data. As outlined in the main text, the researcher-centred life cycle highlights six steps for conducting bioimaging experiments. Core facilities play a role in all steps of the life cycle, starting with planning an experiment. However, we have highlighted four steps in which core facilities are mainly involved as essential partners, as indicated by the colours. From publication to long-term archiving, the responsibilities are increasingly shared with external providers or generic RDM support services, as the gradient indicates. Around the life cycle, tools, systems, and other aspects relevant to researchers and core facilities for applying data management practices are positioned in arbitrary order. Many of these aspects are relevant for multiple life cycle steps, as indicated by the colour code.

relevant RDM topics throughout every step of experimental research, we use the data life cycle concept applied to bioimaging data.<sup>14,16,17</sup> We highlight issues commonly experienced in core facilities and describe our view on present and potential future roles for core facilities in bioimaging RDM. We exemplify tools and describe future directions for data handling.

(i) The first step of a typical bioimage data life cycle is planning. When FAIR image data are openly accessible in public repositories, searching and reusing existing image data can be considered at this step next to planning the following ones: (ii) data acquisition, (iii) data storage and organisation, (iv) data processing and analysis, (v) publication and (vi) data archiving (Figure 1). Many aspects are special for bioimage data since they are often large

and multidimensional and lack a uniform file format but are defined by a specific instrument.<sup>15,18–22</sup> Usually, core facility staff helps with instrument choice. As soon as new bioimage data are to be acquired, they advise users how data should be handled and moved from their quality-controlled acquisition (2.2, image acquisition) to storage location(s) (2.3, storage and organisation) over processing and analysis (2.4, image analysis), to archiving or public deposition (2.5, publication, and 2.6, archiving).

## 2.1 | Experiment planning

Research involves running experiments iteratively to evaluate and refine them but planning early helps to anticipate

potential issues and relevant stakeholder interactions. Problems are averted if the core facility is consulted upfront, which oftentimes does not happen, either because users underestimate the complexity of the techniques and the hurdles to be taken to generate reliable results<sup>23</sup> or simply because the core facility's capacity may be limited to devote enough attention and resources to the project.<sup>5</sup> Scientists with little bioimaging experience might insufficiently consider the consequences of the suboptimal choice of equipment, instrument settings and sources of bias.<sup>24,25</sup> These factors influence whether an image accurately represents the biological phenomenon and the necessary workload from image acquisition to publication-ready results. Improvements in planning and preparation avoid acquiring excessive and unusable data. Additionally, when users do not consider the requirements for image analysis before acquisition, image data sets are likely to be difficult or unsuitable to process for technical reasons.<sup>26,27</sup>

### 2.1.1 | Role of a core facility for experiment planning

Experienced core facility staff can help to acquire high-quality imaging data and avoid storage-intensive imaging modalities when they are not needed,<sup>28</sup> for example, favouring a high-content plate-based acquisition over repeated acquisitions with classical slides. Core facility staff can help identify and overcome technical feasibility concerns,<sup>2</sup> for example, by guided instrument choice after per-project consultation and by estimating required storage volumes (2.2, data acquisition). Ideally, users draft a data management plan (DMP).<sup>5,29</sup> However, researchers and core facility staff seem to have limited usage of DMPs in practice.<sup>7,30</sup> DMP templates that have been approved or provided by the core facility, if available within an institutional DMP tool, can guide scientists to questions regarding image data handling appropriate to their local IT environment. Core facility staff are familiar with the institutional resources for bioimaging, including acquisition machines, IT resources for storage and data transfer, workstations for processing and analysis and services by academic libraries such as providing persistent identifiers (PIDs). Thus, investing in DMP templates can help structure project consultations, alleviating the workload in the long run and taking the prior experience of the researcher planning an experiment into account. Writing a DMP does not mean that everything is fixed. A DMP is regarded as a living document adapted as necessary in a transparent manner.<sup>31</sup> Tools for writing a DMP can be found online ([https://rdmkit.elixir-europe.org/data\\_management\\_plan](https://rdmkit.elixir-europe.org/data_management_plan)).

Core facility staff can teach researchers about the potential for data reuse. Specialised repositories like the BioImage Archive (BIA) or the Image Data Resource (IDR) allow for metadata-based findability, access to, online visualisation, and downloading of published bioimage data.<sup>32–34</sup> Preceding own experiments, researchers can search for data with similarities to their own project and suitable bioimage analysis tools to test one's own analysis workflows (2.4, image analysis). Searching data in repositories and archives also educates about the importance of metadata for finding and retrieving data after publication. Both technical and biological metadata documentation can be planned before an experiment.<sup>35,36</sup> Core facility staff are experts in technical metadata, but it is out of the core facility's scope to provide all research domain-specific metadata standards. However, users can learn from core facility staff about community-established guidelines. Highlighted examples are the Recommended Metadata for Biological Images (REMBI),<sup>37</sup> the NBO tiered guidelines<sup>38</sup> and recommendations for certain fields and modalities like MITI for multiplexed tissue imaging,<sup>39</sup> MIHCSME for high-content screening,<sup>40</sup> 3D-MMS for 3D volume microscopy<sup>41</sup> or the Brain Imaging Data Structure (BIDS).<sup>42</sup> Tools are available to support metadata annotation.<sup>43–45</sup> Tabular metadata collection formats are often used.<sup>7,40</sup> Optionally, core facilities can enforce metadata enrichment, especially for large data otherwise hard to manage. A mandatory metadata checklist could become part of the core facility's usage policy, for example, as recommended for high-content screening (HCS) data within the Dutch bioimaging network.<sup>40</sup>

As data management planning and data handling best practices are accumulated locally over time, they can become a useful resource for new facility users. Due to their mostly well-established cross-institutional networks, core facility staff can also guide towards third-party resources where needed (2.3, storage and organisation, and 2.4, image analysis). Many federated resources exist that researchers may not be initially aware of. Some examples from the European area are the European life science infrastructure ELIXIR, the German Network for Bioinformatics (de.NBI), and the developing European Open Science Cloud (EOSC). Core facility professionals disseminate new and updated community-approved practices, leverage peer-to-peer support or moderate help requests in forums like <https://forum.image.sc>. Table 1 summarises important aspects of data management planning.

## 2.2 | Data acquisition

Ideally, when users collect data, they have planned their experiment enough to ensure the data are comprehensive



**TABLE 1** Common issues and solutions regarding bioimage data management planning.

Common issue	Proposed solution
Underestimating the complexity of bioimage data handling	Consulting and taking advice from core facility staff. Using DMP templates that have been approved or provided by core facilities.
Reusing existing data is not considered	Data search and review in public repositories and institutional archives can become part of experiment planning. Reusing data from repositories can help users familiarise themselves with the data types and test workflows before acquiring their own data. Suitable archives can be found on re3data.org or on FAIRsharing.org. Bioimaging-specific archives include, for example, the BioImage Archive (BIA) and the Image Data Resource (IDR). Institutional archives may be another data source.
Lack of adaptation strategies for changing demands throughout the experiment	Writing a data management plan and using it as a living document to adapt where required, and record changes.
Suboptimal instrument choice	User training and guidance on instrument choice and setup, as well as recommendations on sample preparation. Consulting the core facility and taking advice on instrument and software choice.
Image analysis and data processing are not planned prior to the acquisition	Consulting with core facility staff, image analysts and data stewards before starting a new project. Searching example data and protocols that have established the desired procedure previously to test image analysis a priori. Planning suitable data structures and metadata for analysis and running a pilot experiment before adjusting and scaling up.
The data volumes, storage space and data transfer requirements are not considered before the acquisition	Gathering comprehensive information about typical image sizes of the core facility's instruments. Highlight limits to memory load and data transfer with large files. Estimating expected data volumes and required storage and ensuring resource availability.
The importance of metadata collection is overlooked	Preparing a metadata annotation checklist for the data even before acquisition, for example, based on REMBI. Optionally, enforce annotation as part of a usage policy.
Lack of a strategy for the publication of image data	Considering data publication, candidate public repositories, and reviewing submission guidelines. Defining authorship and licensing criteria with collaboration partners early in the project.

and reliable by recording all the relevant (meta)data that describe the experiment and the (biological) sample. The complexity of this task may not be fully appreciated upfront, for example, when applying an advanced microscopy technique to an unknown sample for the first time. Reliability presupposes valid quality measures to ensure an appropriate instrument setup and calibration, for which most researchers rely on their core facility. Some scientists state that they would not share and reuse data because they do not trust the quality of a dataset they did not acquire.<sup>46</sup> While not standard practice today, a record of the quality measures (e.g., a QC and calibration protocol) would optimally be linked via a persistent identifier (e.g., a DOI). Most commercial acquisition software records technical metadata in nonstandard formats.<sup>47</sup> Due to proprietary file formats, this metadata – like the measurement data itself – is often only accessible with vendor-specific software or through file format conversion or translation. However, the translation fidelity differs between file formats, and the maintenance of

libraries such as Bio-Formats does not scale with the variety of proprietary formats (<https://www.openmicroscopy.org/2016/01/06/format-support.html>).<sup>47</sup> What's more, on-the-fly translation creates a computational bottleneck.<sup>48</sup> For nonbioimaging specialists, these issues are usually a challenge to overcome.

### 2.2.1 | Role of a core facility in data management for image acquisition

Many core facilities have adopted regular quality control measures and instrument calibration as a standard operating procedure – an essential contribution to data reliability and reusability. International standardisation of quality control measures and the provision mentioned above of those quality measure records would be an asset.<sup>49,50</sup> Through global collaboration, the Quality Assessment and Reproducibility for Instruments and Images in Light Microscopy (QUAREP-LiMi) initiative defines such

standards and develops protocols and tools for their practical application.<sup>51–56</sup>

To understand the technical metadata associated with an image, which encode the settings under which the image was acquired, and to avoid mistakes in data handling and interpretation, users should understand first the basic architectural principles of bioimaging files and, second, how information on, for example, optical resolution, detector gain, bit depth, etc., are essential to make the digital representation of a sample that the microscope produced interpretable.<sup>57–59</sup> This technical metadata should be available along with the shared FAIR data. Core facilities are essential in providing training on these aspects.<sup>7,8</sup>

With its fast storage component, the usually vendor-provided acquisition computer functions as the initial location to write data to disk. Using a network drive for initial storage is not recommended, as network instability could result in data corruption. Subject to individual core facility policies, users should clear the acquisition computer and transfer data after acquisition to a suitable storage location optimally through network transfer (2.3, storage and organisation), avoiding portable drives, which would increase the risk of spreading malware.<sup>60</sup>

Vendors usually optimise their proprietary software and file formats for fast writing-to-disk. Many vendors offer free versions of their software to open their file formats. However, data accessibility and interoperability are enhanced when data can be recorded in or faithfully converted to open file formats.<sup>21</sup> Bio-Formats-based conversion to, for example, OME-TIFF has been the de facto standard solution for a long time.<sup>47</sup> But new file formats are required for working with large files in cloud and object storage environments, for which the OME-Zarr format is under development.<sup>48,61</sup> Core facility staff should familiarise themselves with next-generation file formats (NGFF) like OME-Zarr, learn to use conversion tools in relevant use cases, and stay abreast of the development. Thanks to NGFF, data chunking makes large multidimensional images cloud-compatible and ready for streaming. As such, OME-Zarr allows users random access to any plane or part of large multidimensional images over the Internet without the necessity of loading the whole image. This is particularly important for modalities like light-sheet microscopy, HCS, volume electron microscopy, etc., and bears strong potential for a broadly adaptable new standard.<sup>61,62</sup> Table 2 summarises data management aspects related to image acquisition.

## 2.3 | Storage and organisation

Users face various challenges regarding the requirements for storage efficiency, interoperability, computational per-

formance, accessibility and data transfer, particularly for handling data from advanced bioimaging techniques like light-sheet microscopy, HCS or whole slide imagers.<sup>22,60</sup> Effective storage and organisation of large-scale data require hardware infrastructure, workflow optimisation and sharing protocols. Network availability and bandwidth are pivotal for facilitating large file transfer and access. High-speed networks and efficient data transfer protocols minimise latency and maximise throughput, particularly in distributed environments. Specific prerequisites may depend on project-specific factors like collaboration requirements, data volumes and the computational tasks performed on the data. Storage technology for various applications has evolved significantly over the past decades,<sup>63</sup> and a large heterogeneity of storage modalities is found in the research landscape. While individual research groups still often rely on NAS servers or hard drives for data storage, diverse institutional offers might exist, for example, file systems, tape systems for cold storage, distributed or virtual storage and object storage. Aspects regarding storage reliability technologies (e.g., RAID or erasure coding) are to be considered.<sup>60,63</sup> The complexity behind the growing trend towards working ‘in the cloud’ is sometimes underestimated by scientists without in-depth IT knowledge.<sup>64,65</sup>

Besides providing sufficient storage space and network capacity, the storage modality must be appropriate for the data type to allow more than merely holding the data in a certain location. Image data formats and software can influence how efficiently large multidimensional images, for example, 3D time-lapsed multichannel images, are accessed. Especially when working with data through a network, large N-dimensional arrays may require a whole file download, depending on the format. Bioimage data, like other research data, are frequently accessed throughout a project (also referred to as hot data, as opposed to archival data, referred to as cold data). For processing and analysis, regular computer RAM (random access memory) capacities readily fall short.<sup>48,66</sup> Central services (e.g., Nextcloud instances) may be inadequate for sharing large data sets. In bioimaging, data volumes can reach terabytes within one or a few acquisition sessions – exceeding most standard file-sharing solutions’ capabilities.<sup>22</sup>

Research processes encompass numerous steps involving various data types. Data conversion may be necessary for data integration across different platforms. Processing and analysis software produces output files in diverse formats, including images, tables, texts and annotation files. Compressed images are generated regularly for presentations or reports. This multifaceted nature of data handling contributes to potential data fragmentation, making it challenging to maintain a comprehensive overview. Keeping track of all copies and linked files while

TABLE 2 Common RDM issues and solutions for data acquisition.

Common issue	Proposed solution
Quality control and instrument calibration are not documented	A core facility can make the QC measures and records of instrument calibration accessible for the users or provide links to published protocols that are applied for QC at the respective instrument.
<i>Ad-hoc</i> metadata documentation during acquisition is not transformed into structured annotations	Raise awareness to document technical metadata and validate its completeness. If new metadata items are collected, the DMP and structured annotations (e.g., in OMERO) should be amended.
Users leave data on acquisition machine computers	The core facility policy should enforce data transfer from the computers. Orphaned data can be moved, uploaded to a data management system like OMERO offering a period for users to handle the data, or be deleted.
Sample preparation metadata is not added to acquired image files as structured annotations or linked metadata files	Sample preparation can be linked to data, for example, by linking to ELNs, or public protocols in the metadata. For example, in OMERO, the dataset description or the Key-Value Pairs could contain such links.
Technical metadata record correctness is not validated	Sometimes, technical metadata is incomplete or not correctly conserved in file format translations. Essential metadata for the experiment should be checked for newly handled formats.

navigating through processing, sharing, figure drafting and publication can be an error-prone task.<sup>67,68</sup>

Most imaging facilities allow users to access and use the instruments after initial training and are, hence, termed ‘self-service facilities’. In other cases, imaging experiments are performed by facility staff, who have control over data acquisition and data provision to users. Especially in self-service core facilities, the prevailing practice is to transfer data to users’ storage locations immediately after acquisition, rendering users solely responsible for organising and storing the data.<sup>5</sup> Some research groups specify their data management individually per project. Hence, data structuring and labelling can vary significantly within institutions, and this variability hampers data reusability. Repeatedly, new core facility users must recapture images when attempting to restart previous projects because the original data are no longer identifiable.

In our experience, only a small number of users, usually with a background in bioinformatics or computer science, have a deeper understanding of storage in a way as outlined above. Most users ask for sufficient space and data security but understandably rely in full on their IT department or core facility for any details. This creates a challenge when neither the user nor the IT staff know how these details can affect working with the researchers’ specific data types.

### 2.3.1 | Role of a bioimaging core facility in storage and organisation

Full-service or hybrid-service facilities, but also self-service core facilities, are confronted with storage provision and

data organisation challenges. To this end, some core facilities favour gaining full autonomy with respect to their IT resources.<sup>28</sup> One can purchase commercially available packages of acquisition systems, analysis software and workstations, including storage. However, besides a vendor lock-in risk, any additional hardware and software offered by the core facility creates a maintenance overhead that requires a sufficient budget.<sup>2,5</sup>

Most facilities inevitably rely on central resources. In our experience, it is important to seek strategic alignment at the institutional level to reduce costs, facilitate interoperable solutions and share maintenance responsibilities. Besides researchers, relevant partners regarding storage, electronic lab notebooks (ELNs), or Laboratory Information Management Systems (LIMS) are other core facilities, IT departments, academic libraries, and central administrations. Core facilities can estimate the storage needs based on the user number, the instrument bookings and the average file sizes generated per system. Creating custom in-house storage solutions might seem appealing, but a fragmented landscape of solutions can impede usability and interoperability.<sup>69</sup> A bioimaging-specific RDM platform like OMERO, combining storage and organisation, is an asset acknowledged by scientists, although such platforms have not been broadly implemented so far.<sup>7</sup> Several solutions were developed with differences in discipline-specificity versus generalisability.<sup>11,12,70–79</sup> For example, while OMERO or XNAT could (but not necessarily should) be installed by a core facility alone, a more generic system like iRODS (<https://irods.org>) is instead an option as an institution-wide strategic decision for overall RDM of all data types. The most widely adopted platform for bioimaging in Germany is OMERO<sup>7</sup> which can serve as a long-term

**TABLE 3** Important considerations about storage and organisation for bioimage data.

Common issue	Proposed solution
Fragmentation of data	Core facility staff can allocate data to storage and linkage pipelines or workflows. A DMP can help to keep track of data.
The file formats impede data integration	Data can be converted into suitable formats with the help of core facility staff to avoid loss of information regarding data and metadata.
Hurdles to data integration due to the design of the infrastructure. Different data types/modalities cannot be accessed and analysed together	Using a holistic approach that addresses hardware infrastructure, network capacity, workflow optimisation and sharing protocols. Avoiding isolated local systems and ensuring the compatibility of RDM solutions or defining bridges between different solutions. Standardisation of integration processes can avoid confusion and inefficiency. Ensuring that the performance and infrastructure scalability do not become a bottleneck for data integration.
Performance bottlenecks when accessing data	Core facility staff can help close the gaps in mutual understanding between IT and researchers in terms of the architectural structure of image data and the resulting infrastructure requirements.
Lack of central (bioimaging) RDM infrastructure	Core facility staff can educate users to adopt community-approved RDM strategies in file systems or can contribute to aligning strategies at the institution to save costs, facilitate interoperable solutions and share maintenance responsibilities.

storage depending on the total data volumes. OMERO's (and other platforms') advantage is that data are not frequently moved or copied, but users work collaboratively on the data in the same location and export only when necessary.<sup>80</sup> OMERO's user management allows defining data ownership and access rights, which helps avoid orphaned data. If required, connecting an OMERO server via a high-speed cable connection to image analysis workstations can meet user needs for fast computing with large data files.<sup>81</sup> Importantly, finding data in any location depends on metadata. OMERO allows adding metadata as structured annotations while making the instrument metadata accessible, too. In our experience, an OMERO installation on a flexible resource management backend, like a virtual-machine-based Open Stack environment with scalable mounted storage, is an asset for core facilities and researchers alike.

Where a bioimaging RDM platform is lacking, core facility staff can help users to adopt community-approved data structuring in file systems. Examples are the ISA framework, including tools, data model and serialisation for (meta)data,<sup>82,83</sup> or BIDS for Microscopy.<sup>42</sup> Version control software (e.g., Git, Gitlab, DataLad) might be an option for computer-affine users.<sup>84</sup> File naming conventions with tokens can be discussed in project consultations. Where sufficient storage with fast data access is not available, users might consider using third-party, federated storage and compute resources that exist at the regional, national and the international level (e.g., Galaxy <https://imaging.usegalaxy.eu>, de.NBI <https://www.denbi.de>, ELIXIR <https://elixir-europe.org>, EUDAT <https://www.eudat.eu>, EOSC <https://open-science-cloud>.

[ec.europa.eu](https://ec.europa.eu), etc.). Table 3 summarises considerations of data storage and organisation.

## 2.4 | Image analysis

Bioimage analysis is standard practice to obtain scientific results as objectively, reliably, and reproducibly as possible. As with any method, incorrect usage of bioimage analysis methods and tools will lead to wrong conclusions.<sup>85–88</sup> Image analysis is a science in itself, and aspects to consider are outlined elsewhere.<sup>89–94</sup> Notably, data management is crucial in image analysis: accessing large data sets, sharing copies of data or results, maintaining records of processed data and their provenance and associating results with the corresponding raw data for later exploration.<sup>95</sup> This becomes even more pivotal in the context of automated analysis pipelines, where the code heavily relies on well-structured data sources. While a common approach involves using file systems with a combination of folders, naming conventions, and inventory spreadsheets to organise data, cross-referencing integrity is not assured. These challenges create a significant workload for core facility staff assisting researchers with their image analysis project.

### 2.4.1 | Role of a core facility in RDM for image analysis

While some institutions build on independent bioimage analysis facilities,<sup>92</sup> core facilities may provide support and training with image analysis. Existing pipelines and



software should be considered for reuse to lower the overall cost of software development and maintenance.<sup>96–98</sup> Beyond that role, core facility staff can advise researchers on how to document their analysis, using file structures compliant with accepted standards (ISA, BIDS or similar), recording analysis steps, for example, with the macro recorder in ImageJ/Fiji,<sup>99,100</sup> or using Galaxy Imaging<sup>101,102</sup> for robust and reproducible workflows recorded in RO-Crates.<sup>103</sup> Other examples are Bioimage Analysis Desktop (BAND) provided by the EMBL (<https://band.embl.de>) or the AI4LIFE project (<https://ai4life.eurobioimaging.eu>). An overview of resources can be found in David et al.<sup>104</sup> and in a 2023 biologist's guide to planning quantitative imaging.<sup>105</sup> These, however, do not substitute for data management support provided by core facilities.

To ensure data quality and integrity, database technologies have introduced methods to enhance information validity and accuracy. This includes concepts such as primary keys for unique item identification or normalisation processes.<sup>106</sup> Database tables used in RDM systems adhere to these data structuring principles by design. This is why solutions such as OMERO or Cytomine are well suited to handle the complexity of microscopy data: researchers can benefit from the flexibility to view and organise images according to their preferences by leveraging the object-oriented data organisation of such systems, and they can use annotations like tags, key-value pairs (dictionaries) or tables of typed data, each with specific purposes. Such structured data integrates well into image analysis pipelines, eased by the Application Programming Interfaces (APIs) of these tools.<sup>107</sup>

Our experience with OMERO is that it eliminates the need to transfer data between collaborators, decreasing the total data volume and preventing divergences of processed data versions: the data stored centrally on a server is shared via links and accessed from different physical locations and software clients. Tools that are widely used by the community like QuPath,<sup>108</sup> ImageJ/Fiji,<sup>99,100</sup> or Python scripts within Jupyter Notebooks<sup>109</sup> and others have their dedicated OMERO clients.<sup>110–112</sup> Once an analysis is complete, the results are uploaded back to the server and kept in a format defined by the OME model.<sup>113,114</sup>

The development of integrated tools in the OMERO.web client is supported by the bidirectional links between data sources (images) and derived results (ROIs, tables), improving the user experience and simplifying the process of sharing and validating results.<sup>112</sup> ROIs are displayed in the image viewer (OMERO.iviewer) and overlaid in figures (OMERO.figure). Tables are plotted, and data points can be traced back to the source to identify outliers (OMERO.parade, OMERO.mdv). Common issues and

proposed solutions regarding the dependency of image analysis on data management are listed in Table 4.

## 2.5 | Publication

Sequencing or structural biology data have been shared in public repositories for decades.<sup>115–117</sup> Only recently, sharing bioimage data has become more feasible due to specialised repositories such as BIA, or the added-value databases IDR and SSBD.<sup>32,33,118–120</sup> 'Data is available on request' statements in publications were found to be often unreliable in practice,<sup>121–123</sup> and data are not easy to find without a public metadata record available for search engines. Researchers usually focus on experiment optimisation to achieve the best result during their projects. Making this process understandable and reproducible implies providing access to the complete documentation from sample preparation over data acquisition to publication. Mistakes in properly representing microscopy images, or, in extreme cases, intentional misrepresentation, have been well documented.<sup>89,124–126</sup> A principle of Open Science is to share data as openly as possible and keep it only as closed as necessary. Where not restricted by law, open data with metadata linked to a research publication enhances trust in research quality, also indicated by higher citation impact.<sup>127,128</sup> Scientists face several challenges: How can images and the deduced findings be represented accurately in a manuscript? How to write the methods section for microscopy experiments correctly? How to enable access to the original data with little to no restraints? What access and reuse rights should be granted, and who owns the data to choose an appropriate license? These aspects can be complicated if, for example, the data producer has left, the data went through many hands, losing essential metadata, or because guidelines on image data publication are not known or lacking.

### 2.5.1 | Role of a core facility in image data publication

It is the researcher's and author's duty to present data accurately in a publication. Helping researchers follow rigorous bioimaging RDM practices from the start avoids frustration and reduces the workload at the publication stage. With improving capabilities of repositories, new journal submission guidelines, the development of new microscopy techniques and novel standards in research domains, good publication practice keeps evolving. Uncertainty about data ownership is an impediment to data sharing,<sup>7</sup> and core facility staff can help to identify who is authorised

**TABLE 4** RDM aspects around image analysis.

Common issue	Proposed solution
Tracking of data to process with folders, file names and spreadsheets	Using a bioimaging RDM platform to communicate the data status to process and exchange results
Analysis pipelines depend on a rigid data organisation in folder structures	Using established standard file organisations and naming conventions or leveraging object storage and structured annotations to interrogate data flexibly via APIs without duplicating data
Analysis pipelines each implement their own data input and output formats	Analysis pipelines access data in management platforms through an API for data input and uploading analysis output
Image files, analysis files and annotations lack links and structure	Using standard file structures or leveraging database links of RDM platforms for bioimaging (e.g., OMERO) preserving the consistency of the relations between data and analysis results
Analysis results exploration requires dedicated software installation	Data exploration from a remote location and through a browser that only requires an Internet/intranet connection (e.g., OMERO.web)

to license the data, which in turn depends on the collaborators, data privacy concerns, intellectual property rights and funding agencies' regulations.<sup>129</sup> Institutional policy can reinforce open data sharing.<sup>130,131</sup> A trend towards more open data sharing and reuse is apparent among early career researchers.<sup>132,133</sup>

Guidelines on properly presenting image data in publications exist (e.g., by Crome<sup>89</sup>), and members of QUAREP-LiMi developed comprehensive community checklists.<sup>134–136</sup> Core facilities help to disseminate such knowledge and can crosscheck the proper data representation. Only presenting 'representative examples' in a manuscript should be avoided.<sup>122,127</sup> Instead, scientists should publish the original data behind figures, for example, full data sets used to derive quantitative results by image analysis.<sup>32,122</sup> Many journals and funding agencies now require public data deposition. Data publications – even independent of classical publications – are also becoming an option for large image data sets due to the advancement of repositories (for a list, see Ref. 120). Core facility staff can aid with data formatting and repository choice.

Beyond mentioning core facilities in acknowledgement sections, there are reasons to regard a core facility staff's work as a scientifically relevant contribution. Organising and annotating bioimage data improves processing and analysis and fosters scientific rigor, reliability and reusability. Hence, we advocate that core facility staff should become coauthors of the data publication entry in a repository (e.g., BIA) where appropriate, even if they are not coauthors of the journal publication on the overall findings.<sup>137,138</sup> This will help strengthen the core facility's position to the benefit of researchers by showing a measurable impact of core facilities not only as generic support facilities but as partners in scientific discovery. For core facility staff members, it supports building their career in scientific support infrastructures based on a track record.

## 2.6 | Archiving research data

Good scientific practice requires that the source data of a research publication be preserved over the long term, for example, at least ten years after publication, according to the German Research Foundation's code of conduct.<sup>139</sup> If the primary purpose is to hold a faithful data record, local archiving appears sufficient. But data can become trapped in institutional silos, not accessible for more research. In some cases, research data are only stored on hard drives, where data are inaccessible to others and at risk of permanent loss due to hardware failures. Since researchers move between positions while building their careers, retrieving and understanding 'old' data becomes challenging for principal investigators and colleagues.<sup>140</sup>

Data privacy restraints or economic considerations may prohibit open data publication. But even if kept behind closed doors, data archiving is more than storing data for a long time. Archiving must ensure that data can be found, retrieved, and reused upon demand. Only then will 'cold' data that are not frequently accessed remain preserved as FAIR data.<sup>141</sup> This requires data selection and cleaning, a strategy for data structuring and metadata annotation, and regulations regarding access and reuse by institutional policy. Original data must usually be conserved. Derived data may be deleted if it can be regenerated from raw data using the accurately documented analysis pipeline, highlighting the importance of reproducible image analysis again (2.4, image analysis). Many institutions provide central archiving systems through the IT department or academic libraries, but some systems may be insufficient for large bioimaging data. Special software implementations may be in place to orchestrate research data archiving, for example, Tivoli Storage Manager (now IBM Spectrum/Storage Protect) for tape storage.<sup>142</sup> Often, institutions establish long-term preservation systems compliant with the Open Archival

Information System (OAIS) reference model integrating data ingest, storage, management, access and administration (<https://www.iso.org/standard/57284.html>). A combination of solutions may be implemented in local or federated environments, for example, in EUDAT B2SAFE (<https://www.eudat.eu/service-catalogue/b2safe>). Example solutions are iRODS, Coscine (<https://about.coscine.de/en>) or the EOSC project ARCHIVER (<https://www.archiver-project.eu>). Implementing these systems is beyond the scope of an imaging core facility but should be pursued as an institutional strategy.

### 2.6.1 | Role of core facilities in image data archiving

Data archiving goes beyond the specific needs for bioimaging data. Solving this task is not primarily a core facility's responsibility. Core facility staff can support archiving in several ways: by emphasising proper archiving practices towards users, by guiding to the institutional archiving systems, and by communicating bioimaging-specific needs to research data policymakers and infrastructure providers. An example would be a strategy to identify 'cold' data in an imaging data management system like OMERO<sup>11</sup> or Cytomine<sup>70,143</sup> and transfer it to archival systems. Technical solutions for this task are under development, such as an integration between iRODS and OMERO ([https://github.com/irods-contrib/irods\\_working\\_group\\_imaging](https://github.com/irods-contrib/irods_working_group_imaging)). The development of FAIR Digital Objects for bioimaging to wrap data, metadata and accompanying research assets into shippable packages is ongoing.<sup>144</sup> As noted above, core facilities can guide metadata annotation, include quality control protocols and support data cleaning and selection. For example, open file formats like OME-TIFF or OME-Zarr are more suitable for long-term archiving than proprietary vendor formats that may not be supported in the future.<sup>134</sup> Finally, the desired best practice would be archiving data in a suitable public archive as the authoritative faithful data record (2.5, publication). Table 5 summarises issues and proposed solutions for both publication and archiving.

### 2.7 | Closing the life cycle

It has been showcased that new research can be based on existing image data from other studies.<sup>145–147</sup> The reuse potential of quality-controlled, annotated data that are findable and accessible in open, interoperable formats was demonstrated by Williams et al.<sup>118</sup> and FAIR open data are expected to advance bioimage informatics.<sup>148</sup> Especially in

the era of AI-based analysis, segmentation and hypothesis building, well-annotated and quality-controlled research data are essential for training models and obtaining reliable results,<sup>15,149</sup> and data have been reused for novel algorithm development.<sup>150,151</sup>

## 3 | PART II: ACTION ITEMS FOR CORE FACILITIES TO ADVANCE IN BIOIMAGING RDM

We explore strategies and actions for core facility managers, core facility staff and science managers tasked to advance the capacity and capability for bioimaging RDM. While one-size-fits-all solutions are unavailable, we report on considerations from practical experience within the framework of GerBI-GMB and the funded RDM projects in which we are active.

### 3.1 | Foster deployment of a bioimaging RDM platform in your institution

The complexity and sizes of bioimage data constitute specific requirements for bioimaging RDM platforms, and several systems were developed by imaging communities, often with global support. OMERO has a history of more than 20 years in development, is among the best-known and most widely used systems, and is supported by developers worldwide.<sup>7,11,12</sup> Cytomine, established in the field of histopathology, offers comparable features.<sup>70,143</sup> BISQUE is a combined image organisation and analysis platform<sup>72</sup> but appears to have a smaller user base as compared with, for example, OMERO<sup>7</sup> (<https://forum.image.sc/t/data-management-bisque/42370/2>). Originating from the Neuroscience field, XNAT (<https://xnat.org>) is a versatile image data management tool predominantly used in medical and preclinical imaging.<sup>75,152</sup> Moreover, commercial offers exist to enable microscopy data management.

In Germany, many core facilities chose OMERO as an aligned approach to implementing bioimaging-specific data management.<sup>7,13,80,81</sup> Our recommendations are, thus, largely based on the experience with setting up and introducing OMERO instances in Germany while the scheme may apply more generally.

**Review the status:** Establishing and disseminating new practices among researchers takes time, needs trust, and costs resources. At the start, core facilities need to understand where they stand and where to go.

- What are the largest RDM concerns in a facility? Pointed examples should be selected so that nonexperts can understand the nature of the issues.

**TABLE 5** Concepts and considerations for publishing and archiving bioimage data.

Common issue	Proposed solution
Copies of raw data are left on acquisition machines	The core facility's user policy defines and enforces interim data storage, transfer and deletion.
Data is only stored on hard drives together with paper notebooks as documentation	Enriching data with essential metadata to find, identify, understand and potentially reuse it. Contacting institutional archiving providers or uploading data to a repository.
Data is published or archived 'as is'	Data must be curated, selected and quality-controlled before archiving. The data to be archived can be planned upfront and noted in a data management plan. File formats and metadata should be considered. Community checklists can be used for publication.
Valuable data are only archived locally	Considering public repository deposition of the data with or independent of a research article to enable findability and reusability.
The capacity or quota of the local archival system is insufficient for large bioimage data files	Upload to a suitable public repository as the faithful data record after publication.

- Gain an overview of general concerns versus special cases. Intrafacility communication is important for this task.
- What works well? The ability to implement new solutions is underlined by previous success.
- What storage and data itinerary practices are prevailing among users?
- What data types are users concerned with? Are there cases with data privacy protection, intellectual property concerns, etc.?
- Key indicators: Current user number, number of instruments, and potential for expansion. Do external users use the facility? What is the expected yearly rate of image data volumes? Is collaborative data access across institutional boundaries required? Which services critically depend on investing in bioimaging RDM? Primary arguments thus include researchers' demands, funding agency demands, strategic decisions, and positioning from the administration (e.g., participation in large national funding initiatives like, for example, the Excellence Initiative of the German Federal and State Governments). Which high-cost factors, both on the core facility's side and on the researcher's side, could be alleviated with better bioimaging RDM infrastructure?

**Identification of stakeholders and resources:** It is imperative to invest in a stakeholder process.

- **Researchers** are primary data generators. The focus should be on heavy users and special groups of interest: Collaborative Research Centers, Research Training Groups, etc. Applying for infrastructure projects within such funding schemes is an opportunity. Early career scientists, specifically young group leaders, should be included. Strong partners are researchers acknowledg-

ing the core facility as a partner in science, not only regarding core facilities as being 'service providers'. In our experience, scientists are willing to test new solutions when expectations are managed and overpromising is avoided.

- **Core facilities for other technologies:** As is true for managing a facility, collaboration with other core facility managers is important for RDM, identifying synergies versus demand for disparate solutions. Other core facilities are partners for the metadata specification for scientific instruments and data acquisition systems.
- **Information technology (IT) departments** are responsible for implementing and maintaining technical infrastructure, including storage, backup systems, and networks. In some cases, departmental IT is distinct from the institution-wide IT. One should establish a genuine partnership, a process that cannot be started too early, as it takes time and communication. Understanding the IT partner's offers and constraints is important.
- **Institutional Review Boards (IRBs)** and legal departments: Data privacy and security experts are particularly important when dealing with sensitive data. They help adhere to ethical guidelines and privacy regulations.
- **Financial officers and grant offices:** Sustainable and continuous operation requires financing personnel and hardware/software. Especially at the beginning, extra funds to kick off new implementations might be instrumental. Which institutional and third-party funding lines exist? Core facility staff should apply as part of research grants highlighting RDM infrastructure requirements for the research objectives.
- **Scientific networks:** Colleagues with prior experience can support. One can learn from examples and understand local peculiarities as opposed to others.



Contacts within national or regional core facility networks (e.g., German BioImaging, BioImaging North America, Euro-BioImaging, France BioImaging, BioImaging UK, BioImaging NL, Microscopy Australia, etc.) are an asset.

- **Central RDM teams:** If available, local RDM and Open Science teams are valuable partners. At some institutions, general-level data stewards can be partners in implementing a new system (Part III).
- **Administration and Management:** With defined goals and established partnerships, approaching management boards can help to make the case. The management board's perspective plays a role with respect to sustainability and support. What is the future strategic plan, and how would establishing a bioimaging RDM infrastructure align with it? Outline that RDM goals could include finding tangible solutions for the important aspects:
  - Reducing costs through improved data flow: centralised maintenance of storage capacity and organisation, data security and interoperability with other stakeholder resources.
  - Building on established solutions avoiding isolated silos with lock-in problems.
  - Reacting dynamically to changing demands.
  - Facilitating scientific excellence by improving data quality, integrity, reach, collaboration and compliance with funding agency and publisher guidelines.

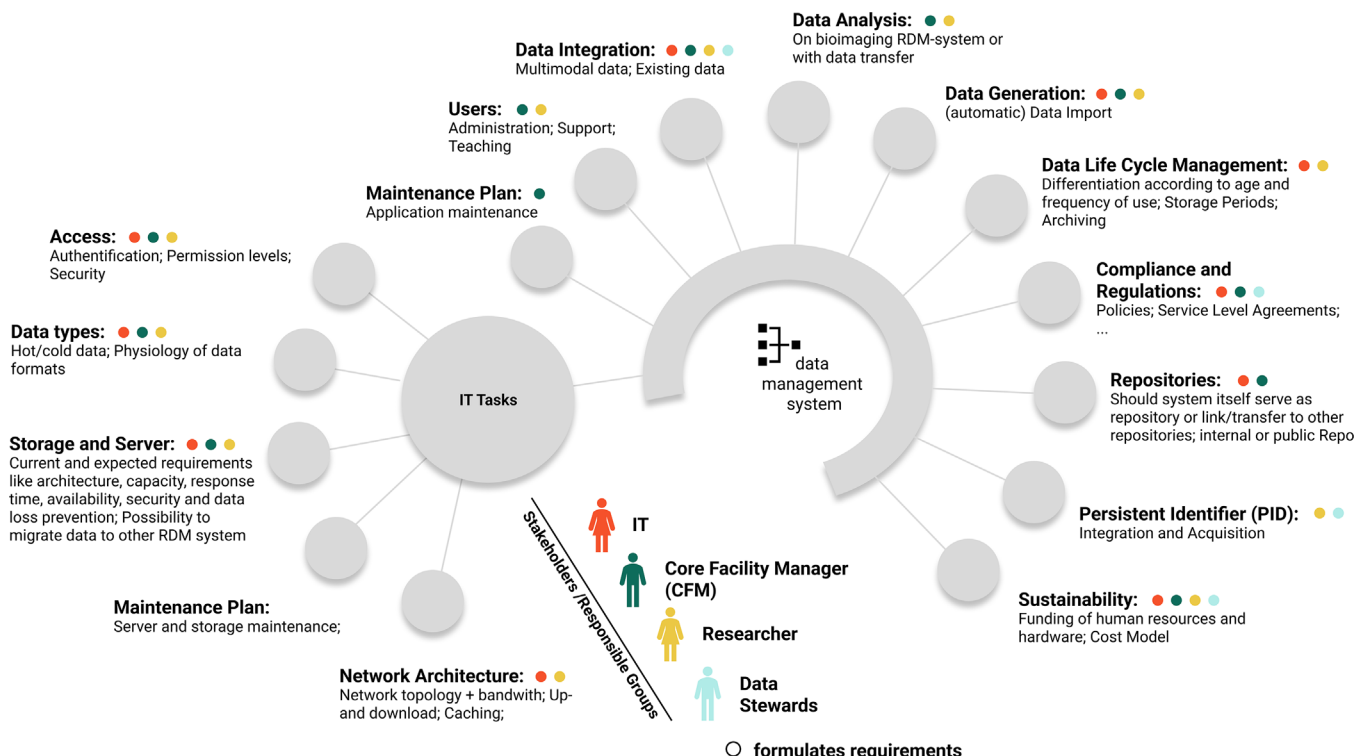
### 3.2 | Design the infrastructure concept

With the stakeholders identified, discrete aspects of a robust and scalable infrastructure can be defined. Various software options and commercial services can be evaluated, and a decision can be made based on the stakeholder process and the difficulty of tasks to solve. We propose a scheme for defining technical details and responsibilities for the topics to address (Figure 2). It focuses on operational implementation, running and maintenance. Therefore, the stakeholders in focus are only those that will be involved over the long term when the infrastructure is installed and an operational service is established.

- **Storage and server location:** Central server infrastructure is typically offered by the IT department. Software providers specify minimum requirements and depend on the expected use. These influence the installation's design (physical server or virtual machine, load balancing, network bandwidth, etc.). We recommend a virtual machine environment, for example, OpenStack, where the required storage is mounted and setups are adapted dynamically. Core facility staff can learn to manage the virtual machine and estimate initial storage needs. With

OMERO, the option to mount researchers' own responsibly maintained storage to the server has advantages (costs, storage space allocation) and disadvantages (risk of breaking links is on researchers' side, responsibility for backup and security).

- **Network architecture:** The need for frequent access to data should be discussed with the researchers and communicated to the network provider. To load centrally hosted large data via the network, several bottlenecks slowing down the process can be avoided. Using file servers mounted to a bioimaging RDM system (so-called in-place import) is an option to consider when working with large amounts of data. For image analysis, a cable-connected, remote analysis workstation with software preinstalled is an option.<sup>81</sup> A large number of simultaneous connections might challenge the system. On the other hand, microscopes generate a large amount of data that require sufficient network bandwidth for upload.
- **Maintenance plan:** Sustained support and hardware maintenance should be secured by IT. A plan to monitor performance issues must be conceived. Maintenance responsibility for the bioimaging RDM system and image analysis stations can lie with core facility staff. Responsibilities must be clearly defined to avoid misunderstandings regarding long-term operation and support.
- **Data types:** Different file sizes and structures of bioimaging data affect the performance of remote access and computing. Data integration with links to other storage environments might be required. Different storage architectures exist with different strengths and limitations.
- **Access and user management:** Depending on the configuration, remote access may be enabled over the Internet or restricted to intranet connections. The user management in OMERO allows default groups and the integration of institutional user and identity management (e.g., LDAP).
- **User integration and training:** Users need training to adopt new solutions. Immediate benefits should be highlighted (viewing images, making figures), and reusable training material can be leveraged, for example, the OMERO guide, <https://omero-guides.readthedocs.io/en/latest>, or I3D:bio's OMERO training material.<sup>153</sup>
- **Data acquisition and integration:** Best practices for uploading newly acquired and existing data can be provided. Large data might require a command-line tool-based supported upload and an appropriate network connection.
- **Data analysis:** Users' needs for image processing and analysis must be discussed. APIs often enable interoperability with OMERO. If commercial software is required or the data transfer is limiting, in-place-imported data



**FIGURE 2** Designing a bioimaging RDM concept. The figure focuses on the operational aspects of installing and maintaining the infrastructure. Different stakeholders may formulate requirements and take over responsibilities for the various tasks depending on local circumstances, configurations, and personnel availability and skill. Responsibilities may lie with different stakeholders or change over time.

might be used, or cable-connected remote stations need to be installed.

- **Data life cycle management:** Considerations include, for example, identifying unused data that should be moved to long-term archiving, or OME-TIFF/ OME-Zarr conversion before uploading to the bioimaging RDM system. Storage usage, annotations, data access, etc. can be regularly monitored, but storage quotas cannot be directly enforced in OMERO.
- **Compliance and regulations:** User management, server configurations and access influence the system's security level. For example, is the upload of privacy-protected data possible? How and when data are deleted? Who is responsible for compliance with data privacy regulations? Compliance with the usage policy, institutional data policy, and general laws have to be considered. In some cases, institutional groups like the staff council have to be involved.
- **Repositories and persistent identification:** Depending on configuration and policy, a system like OMERO can function as an institution-internal or even public repository. Academic libraries might be important partners, for example, for persistent identifiers.<sup>154</sup> Data might also be moved from an OMERO instance to a public repository to increase findability or comply with publisher requirements.

- **Sustainability:** Bottlenecks and limitations of personnel capacity, and direct costs for hardware and software must be considered from the start. Sustainability strategies should be discussed. Synergies with other stakeholders, such as central RDM teams and third-party funding, can be leveraged.

### 3.3 | Planning the process

Implementation should comprise (i) a consultation phase to establish stakeholder contacts and understand perspectives. A task team should lead the efforts once a funding plan is agreed on. (ii) A setup phase where infrastructure is installed and tested by the team to get familiar with systems and assign responsibilities. Updating OMERO, installing extensions and plugins, changing data ownership, or approving new users can be performed by an IT staff member, by a core facility member, or by a data steward, depending on availability and individual skills. Supporting annotation and best practices of image analysis with OMERO will, in contrast, rather become a core facility task. The test instance allows to make mistakes and to learn before the operational instance is installed. (iii) A pilot phase in which a limited number of voluntary users with a limited variety of data types test the

prototype operational system for applicability within their research. The trustful collaboration with pilot users and their acknowledgement is very important. For the core facility, it is a key phase to understand the dynamics of the system, to establish the training needs, and to learn to manage user expectations. The pilot phase duration is flexible but should not be too short, for example, a pilot phase of one year, during which gradually more pilot users are admitted. (iv) Launch of the operational system. Opening the infrastructure for all users is the most relevant but not the ultimate milestone. With increasing user numbers, new concerns are to be expected. No system is optimal for all cases. (v) Evaluation and re-adjustment. Reviewing how using the bioimaging RDM system has changed the previously documented user concerns provides success indicators. A successful implementation benefits the researchers and the strategic position of the core facility.

#### 4 | PART III: DATA STEWARDSHIP FOR BIOIMAGING DATA IN CORE FACILITIES

Image acquisition and analysis are typically aided by imaging experts and image analysts in core facilities (in small facilities, this is often a dual role for one person). Additionally, the size and multidimensionality of many image data formats require expertise in handling data infrastructures, which is usually acquired by core facility staff over the years. Most core facilities have established local storage facilities for image data that allow central access. At the local level, this may satisfy criteria for data accessibility and, to some degree, reusability. To achieve public accessibility and reusability, additional aspects of FAIR data management must be integrated into routine image data handling. This is to transform the theoretical FAIR framework and guiding principles into discrete, actionable practices.<sup>10,155</sup> The international bioimaging community facilitates FAIR data globally, and data stewardship is a key component.<sup>156</sup> The role of the 'data steward' has developed with various understandings depending on data types and research fields. This includes helping to optimise data values by making it reusable but also serving data preservation and research integrity.<sup>157–159</sup>

Best practices in bioimage data management are developing and must be tested and 'negotiated' in the global research community. We regard data stewards as a vital component in implementing novel practices so that the community can evaluate workflows, new tools and standards in everyday research practice. In our experience, such support increases the readiness of researchers to invest in FAIR data management. One reason for not sharing data is the perception of a lack of skills, time and technical resources.<sup>7,46,132</sup> Data stewards train scien-

tists at all career levels and, thus, document and gain experience with best practices that they can apply to new use cases. Examples of data reuse for new research have to be collected to showcase the scientific benefit of bioimaging RDM. Initiatives at the national and international levels, like Germany's NFDI, Euro-BioImaging and bioimaging communities worldwide, focus on advancing the above-mentioned implementing and iterative testing in the disciplinary and methodological research communities. They provide cross-institutional personnel capacity for bioimaging-specific data stewards who foster FAIR data sharing in collaboration with core facilities. Example studies supported by NFDI4BIOIMAGE and I3D:bio are Nöth et al.<sup>138,160</sup> or Jannasch et al.<sup>161</sup> We do not assume that individual core facilities must sustain data steward positions individually. However, data stewards will likely remain key to the research process at large. Hence, new ways for sustainable funding of data stewards at yet-to-be-determined domain-specificity levels will become important for research infrastructures.

#### 5 | CONCLUSION AND OUTLOOK

Research data management and the FAIR principles are no 'end in itself'. There is an initial risk of perceiving RDM practices primarily as a burden, apparently without immediate comprehensible benefit. However, the potential that lies in professionalising bioimaging RDM is enormous: It supports researchers to get more out of their own data, facilitating the systematic handling of (large amounts of) data, thus enabling the democratisation of the research data value if such data are publicly available. It enables the integration of different data types and, therefore, more comprehensive perspectives on a research question. FAIR data management is an essential contribution to successfully applying advanced image analysis approaches, including artificial-intelligence-based procedures. This applies in particular to complex and large imaging data, where new and unbiased algorithms might reveal unexpected new patterns. Thus, FAIR image data management paves the way to a new understanding of spatiotemporal analysis, which can be described as integrative image data science.

For core facilities, increasing their capacity and capability for professional RDM in bioimaging alleviates the workload that today partially stems from the lack of RDM routines. But importantly, it comprises an additional layer underlining the value and impact of core facility work for rigor and trust in scientific discovery. Core facilities have a strategically important position at the interface between relevant stakeholders to make FAIR bioimage data state-of-the-art in research routines.

## AUTHOR CONTRIBUTIONS

CS, TB, JD, SK, EFM and SWP drafted the concept, wrote and edited the manuscript or sections of it. SK and JD drafted the figures. TW and RN contributed to the concept and reviewed the manuscript.

## ACKNOWLEDGEMENTS


We thank the I3D:bio use case partners for collaborating on implementing bioimaging RDM systems within the project, the members of the RDM4Mic group, German BioImaging – Society for Microscopy and Image Analysis, and international partners. We also thank the members and supporters of the NFDI4BIOIMAGE consortium. In particular, we thank Josh Moore, Janina Hanne, and Michele Bortolomeazzi for discussing the article's scope and content. We express our gratitude to the reviewers for feedback and comments to improve the manuscript and to the Scientific Editor, Kurt Anderson, for support.

Open access funding enabled and organized by Projekt DEAL.


## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## ORCID

Christian Schmidt  <https://orcid.org/0000-0001-9568-895X>

Tom Boissonnet  <https://orcid.org/0000-0002-3328-9467>

Karen Bernhardt  <https://orcid.org/0000-0002-7716-587X>

Elisa Ferrando-May  <https://orcid.org/0000-0002-5567-8690>

Tobias Wernet  <https://orcid.org/0009-0004-2093-7885>

Roland Nitschke  <https://orcid.org/0000-0002-9397-8475>

Susanne Kunis  <https://orcid.org/0000-0001-6523-7496>

Stefanie Weidtkamp-Peters  <https://orcid.org/0000-0001-7734-3771>

## REFERENCES

- Ferrando-May, E., Hartmann, H., Reymann, J., Ansari, N., Utz, N., Fried, H.-U., Kukut, C., Peychl, J., Liebig, C., Terjung, S., Laketa, V., Sporbert, A., Weidtkamp-Peters, S., Schauss, A., Zusratter, W., & Avilov, S. (2016). Advanced light microscopy core facilities: Balancing service, science and career. *Microscopy Research and Technique*, 79(6), 463–479. <https://doi.org/10.1002/jemt.22648>
- Meder, D., Morales, M., Pepperkok, R., Schlapbach, R., Tiran, A., & Van Minnebruggen, G. (2016). Institutional core facilities: Prerequisite for breakthroughs in the life sciences: Core facilities play an increasingly important role in biomedical research by providing scientists access to sophisticated technology and expertise. *Embo Reports*, 17(8), 1088–1093. <https://doi.org/10.15252/embr.201642857>
- Renaud, O., Aulner, N., Salles, A., Halidi, N., Brunstein, M., Mallet, A., Aumayr, K., Terjung, S., Levy, D., Lippens, S., Verbavatz, J.-M., Heuser, T., Santarella-Mellwig, R., Tinevez, J.-Y., Woller, T., Botzki, A., Cawthorne, C., & Munck, S., The Core4Life Consortium. (2024). Staying on track – Keeping things running in a high-end scientific imaging core facility. *Journal of Microscopy*, 294(3), 276–294. <https://doi.org/10.1111/jmi.13304>
- Kivinen, K., Van Luenen, H., Alcalay, M., Bock, C., Dodzian, J., Hoskova, K., Hoyle, D., Hradil, O., Christensen, S. K., Korn, B., Kosteas, T., Morales, M., Skowronek, K., Theodorou, V., Van Minnebruggen, G., Salamero, J., & Premvardhan, L. (2022). Acknowledging and citing core facilities: Key contributions to data lifecycle should be recognised in the scientific literature. *Embo Reports*, 23(9), e55734. <https://doi.org/10.15252/embr.202255734>
- Kos-Braun, I. C., Gerlach, B., & Pitzer, C. (2020). A survey of research quality in core facilities. *Elife*, 9, e62212. <https://doi.org/10.7554/eLife.62212>
- DeMaggio, S. (2002). Chapter 14 – Running and setting up a confocal microscope core facility. In B. Matsumoto (Ed.), *Methods in cell biology* (pp. 475–485). Academic Press. [https://doi.org/10.1016/S0091-679X\(02\)70015-0](https://doi.org/10.1016/S0091-679X(02)70015-0)
- Schmidt, C., Hanne, J., Moore, J., Meesters, C., Ferrando-May, E., Weidtkamp-Peters, S., & Members of the NFDI4BIOIMAGE Initiative. (2022). Research data management for bioimaging: The 2021 NFDI4BIOIMAGE community survey. *F1000Research*, 11, 638. <https://doi.org/10.12688/f1000research.121714.2>
- Imreh, G., Hu, J., & Le Guyader, S. (2023). Improving light microscopy training routines with evidence-based education. *Journal of Microscopy*, 294(3), 295–307. <https://doi.org/10.1111/jmi.13216>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Jacobsen, A., De Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Immig, M., Jeffery, K. G., ... Schultes, E. (2020). FAIR principles: Interpretations and implementation considerations. *Data Intelligence*, 2(1–2), 10–29. [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024)
- Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., Macdonald, D., Moore, W. J., Neves, C., Patterson, A., Porter, M., Tarkowska, A., Loranger, B., Avondo, J., Lagerstedt, I., Lianas, L., Leo, S., Hands, K., Hay, R. T., ... Swedlow, J. R. (2012). OMERO: flexible, model-driven data management for experimental biology. *Nature Methods*, 9(3), 245–253. <https://doi.org/10.1038/nmeth.1896>
- Moore, J., Allan, C., Burel, J. M., Loranger, B., MacDonald, D., Monk, J., & Swedlow, J. R. (2008). Open tools for storage and management of quantitative image data. *Methods in Cell Biology*, 85, 555–570. [https://doi.org/10.1016/S0091-679X\(08\)85024-8](https://doi.org/10.1016/S0091-679X(08)85024-8)
- Schmidt, C., Bortolomeazzi, M., Boissonnet, T., Dohle, J., Wernet, T., Hanne, J., Nitschke, R., Kunis, S., Bernhardt, K., Weidtkamp-Peters, S., & Ferrando-May, E. (2023). Bring-



- ing FAIR Bioimage Data Management into Practice: The Information Infrastructure for BioImage Data (I3D:bio) Project – bottom-up Community Support for Microscopy Data Sharing and Preservation. In V. Heuveline, N. Bisheh, P. Kling (Eds.), *E-Science-Tage 2023: Empower your research – Preserve your data* (pp. 289–294). heiBOOKS. <https://doi.org/10.11588/heibooks.1288.c18090>
14. Schmidt, C., & Ferrando-May, E. (2021) NFDI4BIOIMAGE – An initiative for a national research data infrastructure for microscopy data. In *E-Science-Tage 2021: Share your Research Data*. Heidelberg: heiBOOKS. <https://doi.org/10.11588/heidok.00029489>
  15. Zaritsky, A. (2018). Sharing and reusing cell image data. *Molecular Biology of the Cell*, 29(11), 1274–1280. <https://doi.org/10.1091/mbc.E17-10-0606>
  16. Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134–140. <https://doi.org/10.2218/ijdc.v3i1.48>
  17. Griffin, P., Khadake, J., LeMay, K. S., Lewis, S. E., Orchard, S., Pask, A., Pope, B., Roessner, U., Russell, K., Seemann, T., Treloar, A., Tyagi, S., Christiansen, J. H., Dayalan, S., Gladman, S., Hangartner, S. B., Hayden, H. L., Ho, W. W. H., Keeble-Gagnère, G., ... Schneider, M. V. (2018). Best practice data life cycle approaches for the life sciences [version 2; peer review: 2 approved]. *FI000Research*, 6, 1618. <https://doi.org/10.12688/fi000research.12344.2>
  18. Data sharing is the future. *Nature Methods* 2023, 20(4), 471–471. <https://doi.org/10.1038/s41592-023-01865-4>
  19. Wilson, S. L., Way, G. P., Bittremieux, W., Armache, J.-P., Haendel, M. A., & Hoffman, M. M. (2021). Sharing biological data: Why, when, and how. *FEBS Letters*, 595(7), 847–863. <https://doi.org/10.1002/1873-3468.14067>
  20. Bagheri, N., Carpenter, A. E., Lundberg, E., Plant, A. L., & Horwitz, R. (2022). The new era of quantitative cell imaging – Challenges and opportunities. *Molecular Cell*, 82(2), 241–247. <https://doi.org/10.1016/j.molcel.2021.12.024>
  21. Swedlow, J. R., Kankaanpää, P., Sarkans, U., Goscinski, W., Galloway, G., Malacrida, L., Sullivan, R. P., Härtel, S., Brown, C. M., Wood, C., Keppler, A., Paina, F., Loos, B., Zullino, S., Longo, D. L., Aime, S., & Onami, S. (2021). A global view of standards for open image data formats and repositories. *Nature Methods*, 18(12), 1440–1446. <https://doi.org/10.1038/s41592-021-01113-7>
  22. Ouyang, W., & Zimmer, C. (2017). The imaging tsunami: Computational opportunities and challenges. *Current Opinion in Systems Biology*, 4, 105–113. <https://doi.org/10.1016/j.coisb.2017.07.011>
  23. Pearson, H. (2007). The good, the bad and the ugly. *Nature*, 447(7141), 138–140. <https://doi.org/10.1038/447138a>
  24. Lee, R. M., Eisenman, L. R., Khuon, S., Aaron, J. S., & Chew, T.-L. (2024). Believing is seeing – The deceptive influence of bias in quantitative microscopy. *Journal of Cell Science*, 137(1), jcs261567. <https://doi.org/10.1242/jcs.261567>
  25. Jost, A. P.-T., & Waters, J. C. (2019). Designing a rigorous microscopy experiment: Validating methods and avoiding bias. *Journal of Cell Biology*, 218(5), 1452–1466. <https://doi.org/10.1083/jcb.201812109>
  26. Pertusa, J. F., & Morante-Redolat, J. M. (2019). Main steps in image processing and quantification: The analysis workflow. *Methods in Molecular Biology*, 2040, 3–21. [https://doi.org/10.1007/978-1-4939-9686-5\\_1](https://doi.org/10.1007/978-1-4939-9686-5_1)
  27. D'Antuono, R. (2022). Basic digital image acquisition, design, processing, analysis, management, and presentation. In V. Nechiporuk-Zloy (Ed.), *Principles of light microscopy: From basic to advanced* (pp. 77–104). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-031-04477-9\\_4](https://doi.org/10.1007/978-3-031-04477-9_4)
  28. Wallace, C. T., St Croix, C. M., & Watkins, S. C. (2015). Data management and archiving in a large microscopy-and-imaging, multi-user facility: Problems and solutions. *Molecular Reproduction and Development*, 82(9), 630–634. <https://doi.org/10.1002/mrd.22538>
  29. Michener, W. K. (2015). Ten simple rules for creating a good data management plan. *PLoS Computational Biology*, 11(10), e1004525. <https://doi.org/10.1371/journal.pcbi.1004525>
  30. Hahnel, M., Smith, G., Schoenenberger, G., Scaplehorn, N., & Day, L. (2023). The state of open data 2023. Digital Science. <https://doi.org/10.6084/m9.figshare.24428194.v1>
  31. Science Europe Practical Guide to the International Alignment of Research Data Management – Extended Edition, 2021. Zenodo, <https://doi.org/10.5281/zenodo.4915861>
  32. Hartley, M., Anita, L., Babalola, K., Russell, C., Yoldaş, A. K., & Zulueta-Coarasa, T. (2023). Pictures at an exhibition: How to share your imaging data. *Journal of Microscopy*, Online ahead of print. <https://doi.org/10.1111/jmi.13221>
  33. Hartley, M., Kleywegt, G. J., Patwardhan, A., Sarkans, U., Swedlow, J. R., & Brazma, A. (2022). The BioImage Archive – Building a home for life-sciences microscopy data. *Journal of Molecular Biology*, 434(11), 167505. <https://doi.org/10.1016/j.jmb.2022.167505>
  34. Ellenberg, J., Swedlow, J. R., Barlow, M., Cook, C. E., Sarkans, U., Patwardhan, A., Brazma, A., & Birney, E. (2018). A call for public archives for biological image data. *Nature Methods*, 15(11), 849–854. <https://doi.org/10.1038/s41592-018-0195-8>
  35. Kunis, S. (2022). *Structuring of data and metadata in bioimaging: Concepts and technical Solutions in the Context of Linked Data*. Zenodo. <https://doi.org/10.5281/zenodo.7018929>
  36. Montero Llopis, P., Senft, R. A., Ross-Elliott, T. J., Stephansky, R., Keeley, D. P., Koshar, P., Marqués, G., Gao, Y. S., Carlson, B. R., Pengo, T., Sanders, M. A., Cameron, L. A., & Itano, M. S. (2021). Best practices and tools for reporting reproducible fluorescence microscopy methods. *Nature Methods*, 18(12), 1463–1476. <https://doi.org/10.1038/s41592-021-01156-w>
  37. Sarkans, U., Chiu, W., Collinson, L., Darrow, M. C., Ellenberg, J., Grunwald, D., Hériché, J.-K., Iudin, A., Martins, G. G., Meehan, T., Narayan, K., Patwardhan, A., Russell, M. R. G., Saibil, H. R., Strambio-De-Castillia, C., Swedlow, J. R., Fischer, C., Uhlmann, V., Verkade, P., ... Brazma, A. (2021). REMBI: Recommended Metadata for Biological Images-enabling reuse of microscopy data in biology. *Nature Methods*, 18, 1418–1422. <https://doi.org/10.1038/s41592-021-01166-8>
  38. Hammer, M., Huisman, M., Rigano, A., Boehm, U., Chambers, J. J., Gaudreault, N., North, A. J., Pimentel, J. A., Sudar, D., Bajcsy, P., Brown, C. M., Corbett, A. D., Faklaris, O., Lacoste, J., Laude, A., Nelson, G., Nitschke, R., Farzam, F., Smith, C. S., ... Strambio-De-Castillia, C. (2021). Towards community-driven metadata standards for light microscopy: Tiered specifications extending the OME model. *Nature Methods*, 18(12), 1427–1440. <https://doi.org/10.1038/s41592-021-01327-9>
  39. Schapiro, D., Yapp, C., Sokolov, A., Reynolds, S. M., Chen, Y.-A., Sudar, D., Xie, Y., Muhlich, J., Arias-Camison, R., Arena, S.,

- Taylor, A. J., Nikolov, M., Tyler, M., Lin, J.-R., Burlingame, E. A., Abravanel, D. L., Achilefu, S., Ademuyiwa, F. O., Dubinett, S. M., ... Sorger, P. K. (2022). MITI minimum information guidelines for highly multiplexed tissue images. *Nature Methods*, 19(3), 262–267. <https://doi.org/10.1038/s41592-022-01415-4>
40. Hosseini, R., Vlasveld, M., Willemse, J., Van De Water, B., Le Dévédec, S. E., & Wolstencroft, K. J. (2023). FAIR high content screening in bioimaging. *Scientific Data*, 10(1), 462. <https://doi.org/10.1038/s41597-023-02367-w>
41. Ropelewski, A. J., Rizzo, M. A., Swedlow, J. R., Huisken, J., Osten, P., Khanjani, N., Weiss, K., Bakalov, V., Engle, M., Gridley, L., Krzyzanowski, M., Madden, T., Maiese, D., Mandal, M., Waterfield, J., Williams, D., Hamilton, C. M., & Huggins, W. (2022). Standard metadata for 3D microscopy. *Scientific Data*, 9(1), 449. <https://doi.org/10.1038/s41597-022-01562-5>
42. Bourget, M.-H., Kamensky, L., Ghosh, S. S., Mazzamuto, G., Lazari, A., Markiewicz, C. J., Oostenveld, R., Niso, G., Halchenko, Y. O., Lipp, I., Takerkart, S., Toussaint, P.-J., Khan, A. R., Nilsson, G., Castelli, F. M., & Cohen-Adad, J. (2022). Microscopy-BIDS: An extension to the brain imaging data structure for microscopy data. *Frontiers in Neuroscience*, 16, 871228. <https://doi.org/10.3389/fnins.2022.871228>
43. Kunis, S., Hänsch, S., Schmidt, C., Wong, F., Strambio-De-Castillia, C., & Weidtkamp-Peters, S. (2021). MDEmic: A meta-data annotation tool to facilitate management of FAIR image data in the bioimaging community. *Nature Methods*, 18(12), 1416–1417. <https://doi.org/10.1038/s41592-021-01288-z>
44. Rigano, A., Ehmsen, S., Öztürk, S. U., Ryan, J., Balashov, A., Hammer, M., Kirli, K., Boehm, U., Brown, C. M., Bellve, K., Chambers, J. J., Cosolo, A., Coleman, R. A., Faklaris, O., Fogarty, K. E., Guilbert, T., Hamacher, A. B., Itano, M. S., Keeley, D. P., ... Strambio-De-Castillia, C. (2021). Micro-Meta App: An interactive tool for collecting microscopy metadata based on community specifications. *Nature Methods*, 18(12), 1489–1495. <https://doi.org/10.1038/s41592-021-01315-z>
45. Ryan, J., Pengo, T., Rigano, A., Llopis, P. M., Itano, M. S., Cameron, L. A., Marqués, G., Strambio-De-Castillia, C., Sanders, M. A., & Brown, C. M. (2021). MethodsJ2: A software tool to capture metadata and generate comprehensive microscopy methods text. *Nature Methods*, 18(12), 1414–1416. <https://doi.org/10.1038/s41592-021-01290-5>
46. Perrier, L., Blondal, E., & Macdonald, H. (2020). The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PLoS ONE*, 15(2), e0229182. <https://doi.org/10.1371/journal.pone.0229182>
47. Linkert, M., Rueden, C. T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Lorange, B., Moore, J., Neves, C., Macdonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K. W., & Swedlow, J. R. (2010). Metadata matters: access to image data in the real world. *Journal of Cell Biology*, 189(5), 777–782. <https://doi.org/10.1083/jcb.201004104>
48. Moore, J., Allan, C., Besson, S., Burel, J.-M., Diel, E., Gault, D., Kozłowski, K., Lindner, D., Linkert, M., Manz, T., Moore, W., Pape, C., Tischer, C., & Swedlow, J. R. (2021). OME-NGFF: A next-generation file format for expanding bioimaging data-access strategies. *Nature Methods*, 18(12), 1496–1498. <https://doi.org/10.1038/s41592-021-01326-w>
49. Heddlestone, J. M., Aaron, J. S., Khuon, S., & Chew, T.-L. (2021). A guide to accurate reporting in digital image acquisition – Can anyone replicate your microscopy data? *Journal of Cell Science*, 134(6), jcs254144. <https://doi.org/10.1242/jcs.254144>
50. Petrak, L. J., & Waters, J. C. (2014). A practical guide to microscope care and maintenance. *Methods in Cell Biology*, 123, 55–76. <https://doi.org/10.1016/B978-0-12-420138-5.00004-5>
51. Faklaris, O., Bancel-Vallée, L., Dauphin, A., Monterroso, B., Frère, P., Geny, D., Manoliu, T., De Rossi, S., Cordelières, F. P., Schapman, D., Nitschke, R., Cau, J., & Guilbert, T. (2022). Quality assessment in light microscopy for routine use through simple tools and robust metrics. *Journal of Cell Biology*, 221(11), e202107093. <https://doi.org/10.1083/jcb.202107093>
52. Boehm, U., Nelson, G., Brown, C. M., Bagley, S., Bajcsy, P., Bischof, J., Dauphin, A., Dobbie, I. M., Eriksson, J. E., Faklaris, O., Fernandez-Rodriguez, J., Ferrand, A., Gelman, L., Gheisari, A., Hartmann, H., Kukut, C., Laude, A., Mitkovski, M., Munck, S., ... Nitschke, R. (2021). QUAREP-LiMi: A community endeavor to advance quality assessment and reproducibility in light microscopy. *Nature Methods*, 18(12), 1423–1426. <https://doi.org/10.1038/s41592-021-01162-y>
53. Marx, V. (2022). Imaging standards to ease reproducibility and the everyday. *Nature Methods*, 19(7), 784–788. <https://doi.org/10.1038/s41592-022-01544-w>
54. Carvalho, M. T., Fried, H., Jarvis, S., Mitchell, C. A., Nitschke, R., van der Oord, K., Rühl, A., Schwartz, S., Spitaler, M., Zantl, R., & Seitz, A. (2023). Repeatability of the XY positioning of motorized xy-tables of light microscopes. *protocols.io*. <https://doi.org/10.17504/protocols.io.kxygxz2rdv8j/v1>
55. Gaudreault, N., Bagley, S., Bammann, R. R., Barachati, F., Barna, L., Boczonadi, V., Boehm, U., Bosch, M., Brideau, C., Carvalho, M. T., Colarusso, P., Cole, R., Darwish-Miranda, N., Duwé, S., Eismann, F., Faklaris, O., Felscher, A., Gonnert, ... Gelman, L. (2023). Illumination Power, stability, and linearity measurements for confocal and widefield microscopes V.2. *protocols.io*. <https://doi.org/10.17504/protocols.io.5jyl853ndl2w/v2>
56. Deagle, R. C., Wee, T.-L., & Brown, C. M. (2017). Reproducibility in light microscopy: Maintenance, standards and SOPs. *International Journal of Biochemistry & Cell Biology*, 89, 120–124. <https://doi.org/10.1016/j.biocel.2017.06.008>
57. Nechiporuk-Zloy, V. (2022). *Principles of light microscopy: From basic to advanced*. Springer Cham. <https://doi.org/10.1007/978-3-031-04477-9>
58. Shaw, S. L., & Hinchcliffe, E. H. (2013). 65,000 shades of grey: Use of digital image files in light microscopy. *Methods in Cell Biology*, 114, 317–336. <https://doi.org/10.1016/B978-0-12-407761-4.00013-0>
59. Larobina, M., & Murino, L. (2014). Medical image file formats. *Journal of Digital Imaging*, 27(2), 200–206. <https://doi.org/10.1007/s10278-013-9657-9>
60. Andreev, A., & Koo, D. E. S. (2020). Practical guide to storage of large amounts of microscopy data. *Microscopy Today*, 28(4), 42–45. <https://doi.org/10.1017/S1551929520001091>
61. Moore, J., Basurto-Lozada, D., Besson, S., Bogovic, J., Bragantini, J., Brown, E. M., Burel, J.-M., Casas Moreno, X., De Medeiros, G., Diel, E. E., Gault, D., Ghosh, S. S., Gold, I., Halchenko, Y. O., Hartley, M., Horsfall, D., Keller, M. S., Kittisopikul, M., Kovacs, G., ... Swedlow, J. R. (2023). OME-Zarr: A cloud-optimized bioimaging file format with international community

- support. *Histochemistry and Cell Biology*, 160(3), 223–251. <https://doi.org/10.1007/s00418-023-02209-1>
62. Brooks, M. (2023). How open-source software could finally get the world's microscopes speaking the same language. *Nature*, 622(7981), 206–208. <https://doi.org/10.1038/d41586-023-03064-9>
  63. Huawei Technologies Co., Ltd. (2023). Storage basics in cloud computing. In *Cloud computing technology* (pp. 197–250). Springer Nature Singapore. [https://doi.org/10.1007/978-981-19-3026-3\\_5](https://doi.org/10.1007/978-981-19-3026-3_5)
  64. Jonas, E., Pu, Q., Venkataraman, S., Stoica, I., & Recht, B. (2017). Occupy the cloud: Distributed computing for the 99%. In *Proceedings of the 2017 Symposium on Cloud Computing*. Santa Clara, California: Association for Computing Machinery, pp. 445–451. <https://doi.org/10.1145/3127479.3128601>
  65. Suchodoletz, D., Hahn, U., Wiebelt, B., Glogowski, K., & Seifert, M. (2019). Storage infrastructures to support advanced scientific workflows. In M. Janczyk, D. v. Suchodoletz, & B. Wiebelt (Eds.), *Proceedings of the 5th bwHPC Symposium*. Universitätsbibliothek Tübingen. <https://doi.org/10.15496/publikation-29058>
  66. Abernathy, R. P., Augspurger, T., Anderson, B., Blackmon-Luca, C. C., Crone, T. J., Gentemann, C. L., Hamman, J. J., Henderson, N., Lepore, C., McCaie, T. A., Robinson, N. H., & Signell, R. P. (2021). Cloud-native repositories for big scientific data. *Computing in Science & Engineering*, 23(2), 26–35. <https://doi.org/10.1109/MCSE.2021.3059437>
  67. Kuhn Cuellar, L., Friedrich, A., Gabernet, G., De La Garza, L., Fillinger, S., Seyboldt, A., Koch, T., Zur Oven-Krockhaus, S., Wanke, F., Richter, S., Thaiss, W. M., Horger, M., Malek, N., Harter, K., Bitzer, M., & Nahnsen, S. (2022). A data management infrastructure for the integration of imaging and omics data in life sciences. *BMC Bioinformatics [Electronic Resource]*, 23(1), 61. <https://doi.org/10.1186/s12859-022-04584-3>
  68. Mohr, C., Friedrich, A., Wojnar, D., Kenar, E., Polatkan, A. C., Codrea, M. C., Czemm, S., Kohlbacher, O., & Nahnsen, S. (2018). qPortal: A platform for data-driven biomedical research. *PLoS ONE*, 13(1), e0191603. <https://doi.org/10.1371/journal.pone.0191603>
  69. Sheffield, N. C., Bonazzi, V. R., Bourne, P. E., Burdett, T., Clark, T., Grossman, R. L., Spjuth, O., & Yates, A. D. (2022). From biomedical cloud platforms to microservices: Next steps in FAIR data and analysis. *Scientific Data*, 9(1), 553. <https://doi.org/10.1038/s41597-022-01619-5>
  70. Marée, R., Rollus, L., Stévens, B., Hoyoux, R., Louppe, G., Vandaele, R., Begon, J.-M., Kainz, P., Geurts, P., & Wehenkel, L. (2016). Collaborative analysis of multi-gigapixel imaging data using Cytomine. *Bioinformatics*, 32(9), 1395–1401. <https://doi.org/10.1093/bioinformatics/btw013>
  71. Saalfeld, S., Cardona, A., Hartenstein, V., & Tomancák, P. (2009). CATMAID: Collaborative annotation toolkit for massive amounts of image data. *Bioinformatics*, 25(15), 1984–1986. <https://doi.org/10.1093/bioinformatics/btp266>
  72. Kvilekval, K., Fedorov, D., Obara, B., Singh, A., & Manjunath, B. S. (2010). Bisque: A platform for bioimage analysis and management. *Bioinformatics*, 26(4), 544–552. <https://doi.org/10.1093/bioinformatics/btp699>
  73. Shi, B., Bourne, J., & Harris, K. M. (2011). SynapticDB, effective web-based management and sharing of data from serial section electron microscopy. *Neuroinformatics*, 9(1), 39–57. <https://doi.org/10.1007/s12021-010-9088-4>
  74. Langenkämper, D., Zurowietz, M., Schoening, T., & Nattkemper, T. W. (2017). BIIGLE 2.0 – Browsing and annotating large marine image collections. *Frontiers in Marine Science*, 4, . <https://doi.org/10.3389/fmars.2017.00083>
  75. Marcus, D. S., Olsen, T. R., Ramaratnam, M., & Buckner, R. L. (2007). The Extensible Neuroimaging Archive Toolkit: An informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, 5(1), 11–34. <https://doi.org/10.1385/ni.5.1.11>
  76. Swedlow, J. R., Goldberg, I., Brauner, E., & Sorger, P. K. (2003). Informatics and quantitative analysis in biological imaging. *Science*, 300(5616), 100–102. <https://doi.org/10.1126/science.1082602>
  77. Swedlow, J. R., Lewis, S. E., & Goldberg, I. G. (2006). Modelling data across labs, genomes, space and time. *Nature Cell Biology*, 8(11), 1190–1194. <https://doi.org/10.1038/ncb1496>
  78. Wolstencroft, K., Owen, S., Krebs, O., Nguyen, Q., Stanford, N. J., Golebiewski, M., Weidemann, A., Bittkowski, M., An, L., Shockley, D., Snoep, J. L., Mueller, W., & Goble, C. (2015). SEEK: A systems biology data and model management platform. *BMC Systems Biology*, 9, 33. <https://doi.org/10.1186/s12918-015-0174-y>
  79. Prigent, S., Valades-Cruz, C. A., Leconte, L., Maury, L., Salamero, J., & Kervrann, C. (2022). BioImageIT: Open-source framework for integration of image data management with analysis. *Nature Methods*, 19(11), 1328–1330. <https://doi.org/10.1038/s41592-022-01642-9>
  80. Kunis, S., Bernhardt, K., & Hensel, M. (2023). Setting up a data management infrastructure for bioimaging. *Biological Chemistry*, 404(5), 433–439. <https://doi.org/10.1515/hsz-2022-0304>
  81. Zobel, T., Weicher, S., & Wendt, J. (2022). OMERO for microscopy research data management. Wiley Analytical Science.
  82. Rocca-Serra, P., Brandizi, M., Maguire, E., Sklyar, N., Taylor, C., Begley, K., Field, D., Harris, S., Hide, W., Hofmann, O., Neumann, S., Sterk, P., Tong, W., & Sansone, S.-A. (2010). ISA software suite: Supporting standards-compliant experimental annotation and enabling curation at the community level. *Bioinformatics*, 26(18), 2354–2356. <https://doi.org/10.1093/bioinformatics/btq415>
  83. Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bougueleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., ... Hide, W. (2012). Toward interoperable bioscience data. *Nature Genetics*, 44(2), 121–126. <https://doi.org/10.1038/ng.1054>
  84. Halchenko, Y., Meyer, K., Poldrack, B., Solanky, D., Wagner, A., Gors, J., Macfarlane, D., Pustina, D., Sochat, V., Ghosh, S., Mönch, C., Markiewicz, C., Waite, L., Shlyakhter, I., De La Vega, A., Hayashi, S., Häusler, C., Poline, J.-B., Kadelka, T., ... Hanke, M. (2021). DataLad: Distributed system for joint management of code, data, and their relationship. *The Journal of Open Source Software*, 6(63), 3262. <https://doi.org/10.21105/joss.03262>
  85. Miura, K., & Nørrelykke, S. F. (2021). Reproducible image handling and analysis. *Embo Journal*, 40(3), e105889. <https://doi.org/10.15252/embj.2020105889>
  86. Martin, C., & Blatt, M. (2013). Manipulation and misconduct in the handling of image data. *Plant Cell*, 25(9), 3147–3148. <https://doi.org/10.1105/tpc.113.250980>



87. Chen, J., Viana, M. P., & Rafelski, S. M. (2023). When seeing is not believing: Application-appropriate validation matters for quantitative bioimage analysis. *Nature Methods*, 20(7), 968–970. <https://doi.org/10.1038/s41592-023-01881-4>
88. Crome, D. W. (2013). Digital images are data: and should be treated as such. *Methods in Molecular Biology*, 931, 1–27. [https://doi.org/10.1007/978-1-62703-056-4\\_1](https://doi.org/10.1007/978-1-62703-056-4_1)
89. Crome, D. W. (2010). Avoiding twisted pixels: Ethical guidelines for the appropriate use and manipulation of scientific digital images. *Science and Engineering Ethics*, 16(4), 639–667. <https://doi.org/10.1007/s11948-010-9201-y>
90. Peng, H. (2008). Bioimage informatics: A new area of engineering biology. *Bioinformatics*, 24(17), 1827–1836. <https://doi.org/10.1093/bioinformatics/btn346>
91. Chen, W., Xia, X., Huang, Y., Chen, X., & Han, J.-D. J. (2016). Bioimaging for quantitative phenotype analysis. *Methods (San Diego, California)*, 102, 20–25. <https://doi.org/10.1016/j.ymeth.2016.01.017>
92. Soltwedel, J. R., & Haase, R. (2024). Challenges and opportunities for bioimage analysis core-facilities. *Journal of Microscopy*, 294(3), 338–349. <https://doi.org/10.1111/jmi.13192>
93. Cimini, B. A. (2024). Creating and troubleshooting microscopy analysis workflows: Common challenges and common solutions. *Journal of Microscopy*, 1–9. <https://doi.org/10.1111/jmi.13288>
94. Jamali, N., Dobson, E. T. A., Eliceiri, K. W., Carpenter, A. E., & Cimini, B. A. (2022). 2020 BioImage Analysis Survey: Community experiences and needs for the future. *Biological Imaging*, 1, e4. <https://doi.org/10.1017/S2633903%D721000039>
95. Chessel, A. (2017). An Overview of data science uses in bioimage informatics. *Methods (San Diego, California)*, 115, 110–118. <https://doi.org/10.1016/j.ymeth.2016.12.014>
96. Levet, F., Carpenter, A. E., Eliceiri, K. W., Kreshuk, A., Bankhead, P., & Haase, R. (2021). Developing open-source software for bioimage analysis: opportunities and challenges. *FI000Res*, 10, 302. <https://doi.org/10.12688/fi000research.52531.1>
97. Deschamps, J., Dalle Nogare, D., & Jug, F. (2023). Better research software tools to elevate the rate of scientific discovery or why we need to invest in research software engineering. *Frontiers in Bioinformatics*, 3, 1255159. <https://doi.org/10.3389/fbinf.2023.1255159>
98. Rubens, U., Mormont, R., Paavolainen, L., Bäcker, V., Pavie, B., Scholz, L. A., Michiels, G., Maška, M., Ůnay, D., Ball, G., Hoyoux, R., Vandaele, R., Golani, O., Stanciu, S. G., Sladoje, N., Paul-Gilloteaux, P., Marée, R., & Tosi, S. (2020). BIAFLOWS: A collaborative framework to reproducibly deploy and benchmark bioimage analysis workflows. *Patterns (N Y)*, 1(3), 100040. <https://doi.org/10.1016/j.patter.2020.100040>
99. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676–682. <https://doi.org/10.1038/nmeth.2019>
100. Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development*, 82(7–8), 518–529. <https://doi.org/10.1002/mrd.22489>
101. Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., Taylor, J., & Nekrutenko, A. (2020). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research*, 48(W1), W395–W402. <https://doi.org/10.1093/nar/gkaa434>
102. Afgan, E., Nekrutenko, A., Grünig, B. A., Blankenberg, D., Goecks, J., Schatz, M. C., Ostrovsky, A. E., Mahmoud, A., Lonie, A. J., Syme, A., Fouilloux, A., Bretaudeau, A., Nekrutenko, A., Kumar, A., Eschenlauer, A. C., Desanto, A. D., Guerler, A., Serrano-Solano, B., Batut, B., ... Briggs, P. J. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50(W1), W345–W351. <https://doi.org/10.1093/nar/gkac247>
103. Soiland-Reyes, S., Sefton, P., Crosas, M., Castro, L. J., Coppens, F., Fernández, J. M., Garijo, D., Grünig, B., La Rosa, M., Leo, S., Ó Carragáin, E., Portier, M., Trisovic, A., Ro-Crate Community, Groth, P., & Goble, C. (2022). Packaging research artefacts with RO-Crate. *Data Science*, 5, 97–138. <https://doi.org/10.3233/DS-210053>
104. David, R., Rybina, A., Burel, J.-M., Heriche, J.-K., Audergon, P., Boiten, J.-W., Coppens, F., Crockett, S., Exter, K., Fahrner, S., Fratelli, M., Goble, C., Gormanns, P., Grantner, T., Grünig, B., Gurwitz, K. T., Hancock, J. M., Harmse, H., Holub, P., ... Gribbon, P. (2023). Be sustainable": EOSC-Life recommendations for implementation of FAIR principles in life science data handling. *Embo Journal*, 42(23), e115008. <https://doi.org/10.15252/embj.2023115008>
105. Senft, R. A., Diaz-Rohrer, B., Colarusso, P., Swift, L., Jamali, N., Jambor, H., Pengo, T., Brideau, C., Llopis, P. M., Uhlmann, V., Kirk, J., Gonzales, K. A., Bankhead, P., Evans, E. L., Eliceiri, K. W., & Cimini, B. A. (2023). A biologist's guide to planning and performing quantitative bioimaging experiments. *PLoS Biology*, 21(6), e3002167. <https://doi.org/10.1371/journal.pbio.3002167>
106. Kent, W. (1983). A simple guide to five normal forms in relational database theory. *Communications of the ACM*, 26(2), 120–125. <https://doi.org/10.1145/358024.358054>
107. Scherr, T., Seiffarth, J., Wollenhaupt, B., Neumann, O., Schilling, M. P., Kohlheyer, D., Scharr, H., Nöh, K., & Mikut, R. (2022). microscopeSEG: A deep learning software tool with OMERO data management for efficient and accurate cell segmentation. *PLoS ONE*, 17(11), e0277601. <https://doi.org/10.1371/journal.pone.0277601>
108. Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., Mcart, D. G., Dunne, P. D., Mcquaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., & Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1), 16878. <https://doi.org/10.1038/s41598-017-17204-5>
109. Kluyver, T., Benjain, R.-K., Fernando, P., Brian, G., Matthias, B., Jonathan, F., Kyle, K., Jessica, H., Jason, G., Sylvain, C., Paul, I., Damián, A., Safia, A., Carol, W., & Jupyter Development Team (2016). Jupyter Notebooks – A publishing format for reproducible computational workflows. In L. Fernando & B. Schmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press.
110. Ratamero, E. M., Burel, J.-M., Moehl, C., Gay, G., Govek, K., Sobolewski, P., & Moore, W. ome/omero-cli-transfer: 1.0.0rc1 - Figures, plugin interface, -binaries option, XML schemas. Zenodo. <https://doi.org/10.5281/zenodo.10593027>



111. Ratamero, E. M., Govek, K., Langerak, J. M., Sanchez, F. C., & Mellert, D. J. (2023). Easing OMERO adoption with ezomero. *bioRxiv*, pp. 2023.06.29.546930. <https://doi.org/10.1101/2023.06.29.546930>
112. Pouchin, P., Zoghliami, R., Valarcher, R., Delannoy, M., Carvalho, M., Belle, C., Mongy, M., Desset, S., & Brau, F. (2022). Easing batch image processing from OMERO: A new toolbox for ImageJ. *F1000Res*, 11, 392. <https://doi.org/10.12688/f1000research.110385.2>
113. Goldberg, I. G., Allan, C., Burel, J.-M., Creager, D., Falconi, A., Hochheiser, H., Johnston, J., Mellen, J., Sorger, P. K., & Swedlow, J. R. (2005). The Open Microscopy Environment (OME) Data Model and XML file: Open tools for informatics and quantitative analysis in biological imaging. *Genome Biology*, 6(5), R47. <https://doi.org/10.1186/gb-2005-6-5-r47>
114. Li, S., Besson, S., Blackburn, C., Carroll, M., Ferguson, R. K., Flynn, H., Gillen, K., Leigh, R., Lindner, D., Linkert, M., Moore, W. J., Ramalingam, B., Rozbicki, E., Rustici, G., Tarkowska, A., Walczysko, P., Williams, E., Allan, C., Burel, J.-M., ... Swedlow, J. R. (2016). Metadata management for high content screening in OMERO. *Methods (San Diego, California)*, 96, 27–32. <https://doi.org/10.1016/j.jymeth.2015.10.006>
115. Berman, H. M. (2000). The protein data bank. *Nucleic Acids Research*, 28(1), 235–242. <https://doi.org/10.1093/nar/28.1.235>
116. Hingamp, P., Van Den Broek, A. E., Stoesser, G., & Baker, W. (1999). The EMBL Nucleotide Sequence Database. Contributing and accessing data. *Molecular Biotechnology*, 12(3), 255–267. <https://doi.org/10.1385/MB:12:3:255>
117. Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, 41(Database issue), D36–D42. <https://doi.org/10.1093/nar/gks1195>
118. Williams, E., Moore, J., Li, S. W., Rustici, G., Tarkowska, A., Chessel, A., Leo, S., Antal, B., Ferguson, R. K., Sarkans, U., Brazma, A., Carazo Salas, R. E., & Swedlow, J. R. (2017). The image data resource: A bioimage data integration and publication platform. *Nature Methods*, 14(8), 775–781. <https://doi.org/10.1038/nmeth.4326>
119. Tohsato, Y., Ho, K. H. L., Kyoda, K., & Onami, S. (2016). SSBD: A database of quantitative data of spatiotemporal dynamics of biological phenomena. *Bioinformatics*, 32(22), 3471–3479. <https://doi.org/10.1093/bioinformatics/btw417>
120. Bajcsy, P., Bhattiprolu, S., Boerner, K., Cimini, B. A., Collinson, L., Ellenberg, J., Fiolka, R., Giger, M., Goscinski, W., Hartley, M., Hotaling, N., Horwitz, R., Jug, F., Kreshuk, A., Lundberg, E., Mathur, A., Narayan, K., Onami, S., Plant, A. L., ... Keppler, A. (2024). Enabling global image data sharing in the life sciences. *arXiv:2401.13023*. <https://doi.org/10.48550/arXiv.2401.13023>
121. Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: a mixed-methods study. *Journal of Clinical Epidemiology*, 150, 33–41. <https://doi.org/10.1016/j.jclinepi.2022.05.019>
122. Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J.-S., Moyers, B. T., Renaut, S., Rennison, D. J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *Faseb Journal*, 27(4), 1304–1308. <https://doi.org/10.1096/fj.12-218164>
123. Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), 192. <https://doi.org/10.1038/s41597-021-00981-0>
124. Bik, E. M., Casadevall, A., & Fang, F. C. (2016). The prevalence of inappropriate image duplication in biomedical research publications. *MBio*, 7(3), e00809–e00816. <https://doi.org/10.1128/mBio.00809-16>
125. Bik, E. M., Fang, F. C., Kullas, A. L., Davis, R. J., & Casadevall, A. (2018). Analysis and correction of inappropriate image duplication: The molecular and cellular biology experience. *Molecular and Cellular Biology*, 38(20), e00309–e00318. <https://doi.org/10.1128/MCB.00309-18>
126. Marques, G., Pengo, T., & Sanders, M. A. (2020). Imaging methods are vastly underreported in biomedical research. *Elife*, 9, e55133. <https://doi.org/10.7554/eLife.55133>
127. Colavizza, G., Hrynaskiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS ONE*, 15(4), e0230416. <https://doi.org/10.1371/journal.pone.0230416>
128. Duke, C. S., & Porter, J. H. (2013). The ethics of data sharing and reuse in biology. *Bioscience*, 63(6), 483–489. <https://doi.org/10.1525/bio.2013.63.6.10>
129. Margnoi, T., & Tsiavos, P. (2018). *Toolkit for researchers on legal issues*. Zenodo. <https://doi.org/10.5281/zenodo.2574619>
130. Guidance for Developing a Research Data Management (RDM) Policy. (2017). In LEARN Toolkit of Best Practice for Research Data Management. UCL Discovery. <https://doi.org/10.14324/000.learn.27>
131. Tsoukala, V., Angelaki, M., Kalaitzi, V., Wessels, B., Price, L., Taylor, M. J., Smallwood, R., Linde, P., Sondervan, J., Reilly, S., Noorman, M., Wyatt, Bigagli, S., Finn, L., Sveinsdottir, R., & Wadhwa, T. K. (2016). RECODE: Policy recommendations for open access to research data. Zenodo. <https://doi.org/10.5281/zenodo.50863>
132. Van den Veerle, E., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon, D., Frank, M., Jimmy, W., & Louise, C. (2016). Survey of Wellcome researchers and their attitudes to open research. Wellcome Trust. Journal Contribution. <https://doi.org/10.6084/m9.figshare.4055448.v1>
133. Campbell, H. A., Micheli-Campbell, M. A., & Udyawer, V. (2019). Early career researchers embrace data sharing. *Trends in Ecology & Evolution*, 34(2), 95–98. <https://doi.org/10.1016/j.tree.2018.11.010>
134. Schmied, C., Nelson, M. S., Avilov, S., Bakker, G.-J., Bertocchi, C., Bischof, J., Boehm, U., Brocher, J., Carvalho, M. T., Chiritescu, C., Christopher, J., Cimini, B. A., Conde-Sousa, E., Ebner, M., Ecker, R., Eliceiri, K., Fernandez-Rodriguez, J., Gaudreault, N., Gelman, L., ... Jambor, H. K. (2023). Community-developed checklists for publishing images and image analyses. *Nature Methods*, ArXiv, arXiv:2302.07005v2. <https://doi.org/10.1038/s41592-023-01987-9>
135. Taatjes, D. J., Koji, T., Schrader, M., & Roth, J. (2023). Editorial: Histochemistry and Cell Biology implements new submission guidelines for image presentation and image analysis. *Histochemistry and Cell Biology*, 160(6), 495–497. <https://doi.org/10.1007/s00418-023-02247-9>
136. Jambor, H. K. (2023). A community-driven approach to enhancing the quality and interpretability of microscopy images. *Journal of Cell Science*, 136(24). <https://doi.org/10.1242/jcs.261837>

137. Möller, K., Amber, V., & Kemmer, I. (2022). In vivo SPIM and confocal imaging of neuronal efferocytosis by microglia. EMBL-EBI: BioImage Archive.
138. Nöth, J., et al. (2023). Analysis of vascular disruption in zebrafish embryos as an endpoint to predict developmental toxicity – High Content Screening Raw Data (OBI/WIK strain). EMBL-EBI: BioImage Archive.
139. Deutsche Forschungsgemeinschaft. (2019). Guidelines for Safeguarding Good Research Practice. Code of Conduct. <https://doi.org/10.5281/zenodo.3923602>
140. Whitlock, M. C., Mcpeek, M. A., Rausher, M. D., Rieseberg, L., & Moore, A. J. (2010). Data archiving. *American Naturalist*, 175(2), 145–146. <https://doi.org/10.1086/650340>
141. Pernet, C., Svarer, C., Blair, R., Van Horn, J. D., & Poldrack, R. A. (2023). On the long-term archiving of research data. *Neuroinformatics*, 21(2), 243–246. <https://doi.org/10.1007/s12021-023-09621-x>
142. Calhoun, S. P., Akin, D., Zimmerman, B., & Neeman, H. (2019). Large scale research data archiving: Training for an inconvenient technology. *Journal of Computational Science*, 36, 100523. <https://doi.org/10.1016/j.jocs.2016.07.005>
143. Rubens, U., Hoyoux, R., Vanosmael, L., Ouras, M., Tasset, M., Hamilton, C., Longuespée, R., & Marée, R. (2018). Cytomine: Toward an open and collaborative software platform for digital pathology bridged to molecular investigations. *PROTEOMICS – Clinical Applications*, 13(1). <https://doi.org/10.1002/prca.201800057>
144. Moore, J., & Kunis, S. (2023). NFDI4BIOIMAGE: Perspective for a national bioimaging standard. In SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences. CEUR Workshop Proceedings.
145. Hoh, J. H., Heinz, W. F., & Werbin, J. L. (2013). Spatial information dynamics during early zebrafish development. *Developmental Biology*, 377(1), 126–137. <https://doi.org/10.1016/j.ydbio.2013.02.005>
146. Thurley, K., Wu, L. F., & Altschuler, S. J. (2018). Modeling cell-to-cell communication networks using response-time distributions. *Cell Systems*, 6(3), 355–367 e5. <https://doi.org/10.1016/j.cels.2018.01.016>
147. Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E., & Scheres, S. H. W. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife*, 7, e42166. <https://doi.org/10.7554/eLife.42166>
148. Driscoll, M. K., & Zaritsky, A. (2021). Data science in cell imaging. *Journal of Cell Science*, 134(7), jcs254292. <https://doi.org/10.1242/jcs.254292>
149. Prior, F., Almeida, J., Kathiravelu, P., Kurc, T., Smith, K., Fitzgerald, T. J., & Saltz, J. (2020). Open access image repositories: High-quality data to enable machine learning research. *Clinical Radiology*, 75(1), 7–12. <https://doi.org/10.1016/j.crad.2019.04.002>
150. Conrad, R., & Narayan, K. (2023). Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset. *Cell Systems*, 14(1), 58–71 e5. <https://doi.org/10.1016/j.cels.2022.12.006>
151. Pachitariu, M., & Stringer, C. (2022). Cellpose 2.0: How to train your own model. *Nature Methods*, 19(12), 1634–1641. <https://doi.org/10.1038/s41592-022-01663-4>
152. Zullino, S., Paglialonga, A., Dastrù, W., Longo, D. L., & Aime, S. (2022). XNAT-PIC: Extending XNAT to preclinical imaging centers. *Journal of Digital Imaging*, 35(4), 860–875. <https://doi.org/10.1007/s10278-022-00612-z>
153. Schmidt, C., Bortolomeazzi, M., Boissonnet, T., Fortmann-Grote, C., Dohle, J., Zentis, P., Kandpal, N., Kunis, S., Zobel, T., Weidtkamp-Peters, S., & Ferrando-May, E. (2023). I3D:bio's OMERO training material: Re-usable, adjustable, multi-purpose slides for local user training. Zenodo. <https://doi.org/10.5281/zenodo.8323588>
154. Silkotch, C., Garcia-Milian, R., & Hersey, D. (2023). Partnering with health sciences libraries to address challenges in bioimaging data management and sharing. *Histochemistry and Cell Biology*, 160(3), 193–198. <https://doi.org/10.1007/s00418-023-02198-1>
155. Musen, M. A., O'connor, M. J., Schultes, E., Martínez-Romero, M., Hardi, J., & Graybeal, J. (2022). Modeling community standards for metadata as templates makes data FAIR. *Scientific Data*, 9(1), 696. <https://doi.org/10.1038/s41597-022-01815-3>
156. Bialy, N., Alber, F., Andrews, B., Angelo, M., Beliveau, B., Bintu, L., Boettiger, A., Boehm, U., Brown, C. M., Maina, M. B., Chambers, J. J., Cimini, B. A., Eliceiri, K., Errington, R., Faklaris, O., Gaudreault, N., Germain, R. N., Goscinski, W., Grunwald, D., ... Strambio-De-Castillia, C. (2024). Harmonizing the generation and pre-publication stewardship of FAIR image data. arXiv preprint arXiv:2401.13022.
157. Wendelborn, C., Anger, M., & Schickhardt, C. (2023). What is data stewardship? Towards a comprehensive understanding. *Journal of Biomedical Informatics*, 140, 104337. <https://doi.org/10.1016/j.jbi.2023.104337>
158. Rothfrit, L. (2021). Data stewardship als boundary-work. *Bausteine Forschungsdatenmanagement*, (3), 106–118. <https://doi.org/10.17192/bfdm.2021.3.8344>
159. Navale, V., & McAuliffe, M. (2018). Long-term preservation of biomedical research data. *FI000Res*, 7, 1353. <https://doi.org/10.12688/fi000research.16015.1>
160. Nöth, J., Busch, W., Tal, T., Lai, C., Ambekar, A., Kießling, T. R., & Scholz, S. (2024). Analysis of vascular disruption in zebrafish embryos as an endpoint to predict developmental toxicity. *Archives of Toxicology*, 98(2), 537–549. <https://doi.org/10.1007/s00204-023-03633-x>
161. Jannasch, A., Welzel, C., Pablik, J., Von Hauff, E., Galli, R., Rix, J., Schauer, A., Dittfeld, C., & Tugtekin, S.-M. (2024). In vivo application of a glutaraldehyde-free, UVA/riboflavin cross-linked bovine pericardium confirms suitability for cardiovascular substitutes. *APL Materials*, 12(1), 011104. <https://doi.org/10.1063/5.0182672>

**How to cite this article:** Schmidt, C., Boissonnet, T., Dohle, J., Bernhardt, K., Ferrando-May, E., Wernet, T., Nitschke, R., Kunis, S., & Weidtkamp-Peters, S. (2024). A practical guide to bioimaging research data management in core facilities. *Journal of Microscopy*, 294, 350–371. <https://doi.org/10.1111/jmi.13317>