

# Building a FAIR image data ecosystem for microscopy communities

Isabel Kemmer<sup>1</sup> (0000-0002-8799-4671), Antje Keppler<sup>1</sup> (0000-0003-4358-2269), Beatriz Serrano-Solano<sup>1</sup> (0000-0002-5862-6132), Arina Rybina<sup>1</sup> (0000-0002-5609-5710), Buğra Özdemir<sup>1</sup> (0000-0001-9823-0581), Johanna Bischof<sup>1</sup> (0000-0002-5668-841X), Ayoub El Ghadraoui<sup>1</sup> (0000-0002-5494-036X), John E. Eriksson<sup>2</sup> (0000-0002-1570-7725), Aastha Mathur<sup>1\*</sup> (0000-0001-9734-9767)

<sup>1</sup> Euro-Biolmaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany

<sup>2</sup> Euro-Biolmaging ERIC Statutory Seat, P.O. Box 123, Tykistökatu 6, FI-20521, Turku, Finland

\* Corresponding: [aastha.mathur@eurobioimaging.eu](mailto:aastha.mathur@eurobioimaging.eu)

## Abstract

Bioimaging has now entered the era of big data with faster than ever development of complex microscopy technologies leading to increasingly complex datasets. This enormous increase in data size and informational complexity within those datasets has brought with it several difficulties in terms of common and harmonized data handling, analysis and management practices, which are currently hampering the full potential of image data being realized. Here we outline a wide range of efforts and solutions currently being developed by the microscopy community to address these challenges on the path towards FAIR bioimage data. We also highlight how different actors in the microscopy ecosystem are working together, creating synergies that develop new approaches, and how research infrastructures, such as Euro-Biolmaging, are fostering these interactions to shape the field.

**Keywords:** bioimaging, FAIR, community, metadata, data management

## Author contributions

*Conceptualization:* Isabel Kemmer, Antje Keppler, Aastha Mathur

*Writing - original draft preparation:* Isabel Kemmer, Antje Keppler, Aastha Mathur

*Writing - review and editing:* Isabel Kemmer, Antje Keppler, Beatriz Serrano-Solano, Arina Rybina, Buğra Özdemir, Johanna Bischof, Ayoub El Ghadraoui, Aastha Mathur

*Supervision:* Antje Keppler, John E. Eriksson

## Statements and Declarations

### *Competing interests*

The authors have no competing interests to declare that are relevant to the content of this article.

### *Funding*

I.K. receives funding from EU Horizon 2020: grant agreement no. 101046203 (BY-COVID). B.S. receives funding from the EU Horizon Europe research and innovation programme: grant agreement no. 101057970 (AI4life). A.R. receives funding from EU Horizon 2020: grant agreement no. 824087 (EOSC-life) and EU Horizon Europe research and innovation programme: grant agreement no. 101046133 (ISIDORE). B.Ö. receives funding from EU Horizon 2020: grant agreement no. 101017536 (EOSC Future).

### *Acknowledgements*

The authors acknowledge all members of the Euro-Biolmaging Hub for discussions and support.

### *Data availability*

Data sharing is not applicable to this article as no datasets were generated or analyzed.

## Introduction: The hidden potential and current challenges of bioimaging data

In recent years, the field of bioimaging has evolved from its qualitative origins, mainly focusing on the visualization of individual biological processes and structures, to a highly quantitative discipline, producing large and complex, sometimes multimodal, datasets requiring sophisticated and robust computational and statistical methods for analysis. At the same time, imaging is becoming a key enabling technology in the biological and biomedical sciences and beyond, as evidenced by an incredibly wide range of imaging modalities and techniques that are constantly being developed (Ouyang & Zimmer, 2017). Bioimaging methods now span length scales from single molecules to cells, all the way to entire multicellular organisms, as well as time scales ranging from fractions of milliseconds to days. Additionally, multiplexing allows the simultaneous visualization of a multitude of sample characteristics in parallel (Ellenberg et al., 2018). The underlying technical and, more importantly, computational advancements are now enabling scientists to understand and map spatiotemporal biological processes in unprecedented quantity and detail.

While this diversity demonstrates the immense potential of bioimaging, it also results in larger and more complex datasets that present many new challenges. Thus, we are now witnessing the transformation of bioimaging into the era of big data, where data requires rigorous standards for metadata, data management and dissemination procedures than ever before (Driscoll & Zaritsky, 2021). Given the rising cost and time involved in increasingly intricate, including correlative and multimodal imaging experiments and analysis, it is imperative to maximize the reliable use and potential reuse to increase the value output of microscopy data. While imaging has the inherent potential to become increasingly impactful to study processes on molecular, cellular and organism level *in vivo* and *in vitro*, its full potential can only be realized and effectively exploited when these challenges are adequately addressed.

This goal can be achieved by adhering to the FAIR principles which provide guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability of data so that it can be shared in a way that enhances and promotes reuse by both humans and machines (Wilkinson et al., 2016). In short, according to the FAIR principles, (i) each data element should have a unique identifier associated with searchable metadata to be findable; (ii) data should be accessible and retrievable through well-defined and standardized access protocols; (iii) data and metadata should be described using widely-adopted and domain-relevant vocabularies and ontologies to be interoperable; and (iv) data should be described with rich metadata and cross-references to other data sources and publications as well as state a clear license for reuse.

Since their formal statement in 2016, the FAIR principles have become a cornerstone for making research output more impactful and usable, with growing adoption by scientific communities. Research communities are at different stages along the spectrum of FAIR data: Some domains have been implementing similar principles long before the advent of 'FAIR'. Examples of communities with well-established FAIR standards include structural biology, genomics and astronomy, among others (Berman et al., 2020; Byrd et al., 2020; Morris, 2018; Pepe et al., 2014). While some research communities are just starting their FAIR journeys, others, such as bioimaging, have been experiencing a shift in mindsets towards FAIR over the past years. On this path, the imaging community has made significant efforts, but also faces considerable difficulties due to a large diversity of methods, research questions and produced data which are currently being addressed with various solutions that will be outlined in this

commentary. Although the FAIR principles equally apply to both sensitive and Open data, within the scope of this work we will elucidate current state and undertakings in the field of Open microscopy image data, acknowledging the potential for interoperability between microscopy and medical imaging fields. Overall, we believe that realizing the FAIR vision will require both concrete support and incentives from an ecosystem of diverse stakeholders: researchers, imaging facilities, universities, policy makers, funders, journals, communities, and instrument manufacturers, among others, working together towards the same goal. By linking communities across scientific areas and disciplines and fostering exchange of knowledge and know-how on an international level, research infrastructures like Euro-BiolImaging (see Box 1) are a central player within that landscape. In this way, bioimaging will move to a place where data and software sharing according to the FAIR principles is a top priority, laying the foundation for accelerated innovation and cutting-edge research.

---

### **Box 1: Realizing the FAIR vision through European Research Infrastructures**

Research Infrastructures<sup>1</sup> refer to facilities, services and resources of high-quality standards and of national and international significance that are open to the scientific community. Due to their long-lived nature, they are particularly conducive to knowledge and technology transfer and foster links and collaborations with all their stakeholders, including academia, industry and national and international authorities. As one of these, Euro-BiolImaging ERIC<sup>2</sup> is a European Research Infrastructure that democratizes access to world-class imaging services ranging from nano to macroscopic scales (Pfander et al., 2022). As of March 2023, it provides open access to 105 different state-of-the-art biological and biomedical imaging technologies in 173 imaging facilities organized in 35 Nodes hosted by national research institutions and universities across Europe. 16 European countries and the international organization EMBL (European Molecular Biology Laboratory) have committed to jointly operate this pan-European research infrastructure. This tremendous support at the European level makes the infrastructure a unique resource for researchers, even on the global landscape. All scientists, regardless of country, affiliation, research area, or expertise, can benefit from these high-quality services provided by leading imaging facilities including a network of experts. In addition, researchers have access to highly specialized training, much-needed data management and image analysis services to help them to extract meaningful information from bioimaging data. In this way, Euro-BiolImaging promotes collaboration, innovation, sharing of resources and expertise as well as Open Science and Open Access in the field of life science imaging, with the goal of advancing scientific knowledge and driving technological innovation.

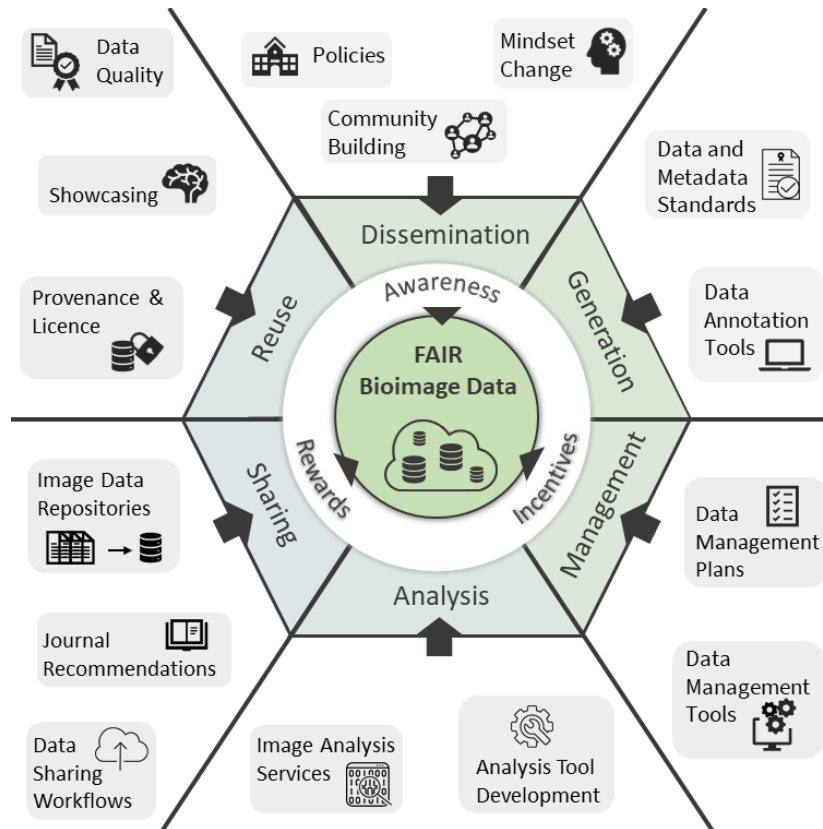
---

<sup>1</sup>[https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures\\_en](https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures_en)

<sup>2</sup> <https://www.eurobioimaging.eu/>

## Stepping stones on the road to FAIR

In contrast to more homogeneous disciplines that have a tradition of implementing FAIR principles and data sharing, bioimaging is a much more heterogeneous field with more variable methodologies and protocols for preparing samples, acquiring, and analyzing data. Based on experiences from other domains we can identify several obstacles that currently impede FAIR bioimage data, and devise a multifaceted approach (Figure 1) to promote data sharing and reuse at scale (Bagheri et al., 2022). Although it may not be possible to fully adopt currently available strategies and pipelines, the bioimaging community can still draw inspiration from them to shape the path of microscopy towards FAIR data.



**Figure 1:** Six facets of FAIR bioimage data (data generation, data management, data analysis, data sharing, data reuse and FAIR dissemination) alongside the approaches and tools currently under active development by the bioimaging community. Further improving all of these areas still requires three overarching elements (awareness, incentives, and rewards) to move toward the ultimate goal of FAIR bioimage data.

### **Facet 1: Image Data Generation and Metadata**

Annotating data with descriptive, high-quality metadata is a prerequisite for almost all aspects of FAIR data. Thus, imaging datasets become fully valuable and usable only when accompanied by rich metadata (Huisman et al., 2019). Metadata is a set of data that provides information about the original data and should include all information needed to enable discovery, interpretation and re-use of that data including, for example, data type, origin, organization, relationships to other data and more. However, there are currently still significant gaps in how complete and accurate microscopy experiments are documented and reported upon publication (Marqués et al., 2020). One reason for this reporting gap is the lack of clear

recommendations and implemented standards on what metadata should be reported alongside an imaging experiment. Frequently, journals request that the description of an experiment must allow for replication of the results, which can lead to a rather vague interpretation by a large proportion of authors and data owners.

Providing concrete metadata recommendations is not an easy task, especially considering the wide variety of microscopy methods, it is not feasible to define concrete requirements for all existing imaging modalities and their sub-variants. In addition, the amount and type of information needed for reuse depend on the further scientific application and the needs of other broad or targeted user communities. Therefore, the bioimaging community has come together to converge on an initially flexible, high-level framework of metadata reporting guidelines that can be progressively tightened and specified by consensus of different sub-domains. These "REcommended Metadata for Biological Imaging" (REMBI) guidelines (Sarkans et al., 2021) provide metadata aspects in eight categories ranging from the imaging data itself to metadata describing the study and the sample used that should accompany any imaging dataset. Additional detailed metadata models exist, such as OME (Goldberg et al., 2005) and 4DN-BINA-OME (Hammer et al., 2021), but they require dedicated and sometimes extensive effort to meet the requirements and therefore lack widespread adoption. While adherence to these standards may not automatically lead to a fine-grained understanding of the imaging experiment or ensure full reproducibility, it is certainly a step in the right direction.

Euro-BiolImaging advocates for the adoption of clear metadata standards and supports consistent metadata reporting in general as well as by repositories and journals. Some journals already provide their own metadata reporting checklist for imaging studies to try to reduce the metadata reporting gap. Others are working with community initiatives for quality assessment and quality control such as QUAREP-LiMi (Boehm et al., 2021) to propose a set of minimum microscopy metadata keywords that should always be reported at publication. Ultimately, journals are an important lever to incentivize the field to move towards more consistent metadata reporting, thereby increasing the credibility and usability of microscopy data. However, care should be taken not to create too many different requirements, as this will result in considerable heterogeneity in reporting standards and thus reluctance on the part of the researcher to collect metadata. Standards should therefore be consolidated, unified, and agreed upon by the community. In this way, scientists can consider them and already comply with them before generating data, thus improving the way research is conducted long before publication.

It is also important to bear in mind that the burden and workload of recording metadata should not fall solely on the shoulders of individual researchers. The desire for comprehensive and rich metadata must therefore be carefully weighed against the additional effort and time it requires in practice. Hence investing in tools for automatic and standardized harvesting of metadata from imaging instruments, preferably at the time of image acquisition itself, is beneficial for the community. Several open-source tools for coupling image acquisition with metadata capture are currently under active development in the imaging community. These include the Micro-Meta App (Rigano et al., 2021), a tool to collect microscope metadata according to the community-developed metadata standard 4DN-BINA-OME (Hammer et al., 2021); MDEmic (Kunis et al., 2021), a metadata annotation tool; MethodsJ2 (Ryan et al., 2021), a tool to create comprehensive methods sections for publications.

Alongside developing these accessible tools and pipelines it is essential to establish over time a community-wide network of experts to advise and practically consult researchers on the

implementation of these solutions. Euro-Biolmaging supports imaging facilities by offering guidance on how to properly - and preferably automatically - record and report metadata using appropriate tools. On the other hand, microscope manufacturers also play a crucial role in the dissemination of image acquisition tools and software and work towards the development of better and more intuitive and automated metadata acquisition. Research infrastructures are in close contact and exchange with the industry partners, for example through the Euro-Biolmaging Industry Board<sup>3</sup>, and aim to align the efforts of the academic communities with those of the manufacturers.

Experience in other domains, such as the omics field, has taught us that some degree of standardization is essential for data to be truly interoperable and reusable (Chervitz et al., 2011). Therefore, it is not only important to capture metadata, but also to ensure that different metadata schemas are interoperable, i.e., they need to speak the same metadata language or at least be translatable. In this way, metadata ideally identifies relevant (biological) entities by means of controlled vocabularies and ontologies, for example using FBBI or EDAM-Biolmaging (Kalaš et al., 2019). We recommend their use and provide targeted guidance for research conducted using Euro-Biolmaging access as well as in the collaborating bioimaging repositories.

The COVID-19 pandemic revealed a critical lack of metadata interoperability, especially between distant research domains. To address this, Euro-Biolmaging is actively shaping Horizon Europe (HE)-funded projects such as BY-COVID<sup>4</sup> and EOSC4Cancer<sup>5</sup>, which aim to make COVID and cancer-related data accessible to everyone and to harmonize metadata standards so that data are interoperable across disciplines. With the goal of accelerating and improving the preparedness for future pandemics and combating acute diseases of concern, these cross-disciplinary initiatives bring together key Research Infrastructures providing relevant services. Together they aim to develop and integrate common operational principles and strategies, from data collection to harmonization of and open access to different types of data. The existential need for data stewardship to advise and assist on metadata types and standards, as well as the resources to provide data stewardship, is becoming increasingly evident and recognized as one important outcome of these projects. In this respect, Euro-Biolmaging is at the forefront of these efforts by currently appointing a data steward to serve as a bridge between bioimaging researchers, communities and repositories.

## ***Facet 2: Image Data Storage and Management***

As bioimaging technologies and associated datasets continue to evolve, the amount of data that is generated has grown exponentially. Today, individual experiments can easily produce terabytes of data. As a result, data management capabilities and strategies have typically failed to keep pace with the growth of image data size. It is therefore imperative for all imaging facilities and researchers to consider how data will be stored, documented and moved during and after the project's lifetime, before the data is even generated (Schmidt et al., 2022; Wallace et al., 2015).

---

<sup>3</sup> <https://www.eurobioimaging-industryboard.com/>

<sup>4</sup> <https://by-covid.org/>

<sup>5</sup> <https://eosc4cancer.eu/>

As data volumes explode, conventional file-system-based storage solutions are quickly reaching their limits, partly because traditional microscopy data formats were not built for big data, let alone the cloud. As image data storage inevitably moves to cloud-based options, it is necessary to adapt image data formats to this mechanism to allow for open data access and sharing. The imaging community, organized by the Open Microscopy Environment (OME, (Swedlow et al., 2003)) is now addressing this need for a standardized file format by supporting and advancing Next Generation File Formats (NGFFs), specifically OME-Zarr, which is cloud-native and allows for easy streaming of chunks of very large datasets for interactive visualization and analysis (Moore et al., 2021, 2023).

Ensuring proper implementation of data management practices early in a project can significantly reduce the time for required data management and frustration that may arise later on. Having such a Data Management Plan (DMP) for research data is increasingly requested by funders, such as the European Commission<sup>6</sup>, universities as well as research institutes. While the RDMkit<sup>7</sup> compiles common guidelines to assist life scientists in their efforts to manage their research data, universities and facilities also ideally provide opportunities for training and consultation to scientists and personnel at various career stages. However, care must be taken to ensure that the Data Management requirements do not become yet another administrative burden without value. To be of actual impact for the researcher, a DMP is thus properly tailored to the actual research project. Similar to the collection of metadata, the responsibility and effort of data management should not solely rest on the imaging scientists. Therefore, facilities and universities ideally provide and clearly communicate standardized data management procedures and recommendations. In addition, institutions would implement tools for data management plan creation such as the Data Stewardship Wizard (Pergl et al., 2019) or Argos<sup>8</sup> that provide smart and tailored data management prompts for specific research projects. By currently offering data stewardship services, Euro-BiolImaging assists its facilities with providing data management templates, and also individual researchers in adapting and customizing those templates to their particular project. In this way, life sciences Research Infrastructures are commonly shaping the current progress of data management recommendations. They achieve this through close contact with the needs of scientists as well as with funders and EU-funded consortia and cluster projects such as EOSC-Life<sup>9</sup> as exemplified by the ongoing efforts of HE-funded projects such as ISIDORE<sup>10</sup> to create umbrella DMPs (David et al., 2023). These extend into many research domains and encompasses research infrastructures beyond bioimaging which is particularly important to facilitate and manage cross-disciplinary projects.

Sensibly, data management strategies should not only consider the actual task of storing and moving the data, but also how the appropriate metadata is stored and maintained in conjunction with the actual data. By providing researchers with access to image data management platforms, such as OME (Allan et al., 2012), universities, institutions and facilities can increase the adoption and use of these tools. For example, the establishment of institutional OME servers is being catalyzed by the support and sharing of experiences

---

<sup>6</sup>[https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)

<sup>7</sup> <https://rdmkit.elixir-europe.org/>

<sup>8</sup> <https://argos.openaire.eu/home>

<sup>9</sup> <https://www.eosc-life.eu/>

<sup>10</sup> <https://isidore-project.eu/>

through research infrastructures. This will result in FAIRer image data with minimal effort on the part of the researchers.

### ***Facet 3: Image analysis***

Not only has the size of image data grown with the advances in resolution, speed, automation and multimodality, but it has also become more complex and richer in information content. The analysis of such image data is hence more demanding, in terms of understanding the complex biological subject matter, the technological details of the microscopy method, as well as technical expertise and resources to efficiently deal with intricate image data in large volumes (Jamali et al., 2022). The visual nature of microscopy data has meant that in the past, analysis methods have been largely manually supported, making it hard to scale and report in a standardized manner. In fact, studies have shown that image analysis methods are vastly underreported in peer-reviewed studies (Marqués et al., 2020), thus hampering scientific rigor, reproducibility and reusability. However, over time, there has been a growing emphasis on standardizing the reporting of image analysis methods (Schmied et al., 2023), even when the analysis is performed in a manual or semi-automated manner. In addition, we see a general shift towards more automated, modular and thus reusable methodologies.

To take advantage of high performance and cloud computing it is necessary to support the current evolution of image analysis methods in making them more scalable, reproducible and FAIR (Goble et al., 2020; Miura & Nørrelykke, 2021). Following standardization, similar to other fields like genomics, image analysis methods are seeing increasing use of workflow management systems that allow users to chain different tools to create reproducible analysis workflows. They also promote shareability of analysis procedures by automatically storing the workflow in a machine-readable and executable format. At the same time, they enhance image data analysis in the big data sphere by facilitating parallelization and execution on high-performance computing resources and sometimes offer data storage and execution environments, hence enabling completely cloud-based analysis solutions. These platforms include generalized workflow management systems (Paul-Gilloteaux et al., 2021), such as Galaxy (Jalili et al., 2020), Nextflow (Di Tommaso et al., 2017), and Snakemake (Mölder et al., 2021) as well as platforms such as KNIME (Berthold et al., 2009) offering image data specific options for management, deployment and analysis that integrate different image processing softwares. Moreover, generated workflows can be registered and thus disseminated in workflow registries such as the WorkflowHub (Goble et al., 2021). By adopting and extending common workflow management systems, Euro-BiolImaging empowers life scientists to benefit from the European Open Science Cloud (EOSC<sup>11</sup>), an open multidisciplinary environment for hosting and processing research data to support EU science. In particular, we are building parallelised data conversion and analysis tools and workflows<sup>12</sup> that assist researchers to remotely work with image data in the OME-Zarr format as part of the EOSC Future<sup>13</sup> project, which aims to integrate existing data, resources, and services across scientific disciplines to provide a set of persistent and interoperable resources. Such efforts bridge the gap between image file formats and data repositories, thus further accelerating FAIR image data sharing.

---

<sup>11</sup> <https://eosc-portal.eu/>

<sup>12</sup> <https://github.com/Euro-BiolImaging>

<sup>13</sup> <https://eoscfuture.eu/>



The evolution in image analysis methodologies is largely driven by expert bioimage analysts (Miura, 2016) that sit between the blurring boundaries of life and computer sciences domains. While larger research institutes and laboratories benefit from these dedicated personnel to meet their image analysis needs, small facilities and individual labs often lack the resources to host and fund such positions. To meet these growing analysis needs of researchers, Euro-BioImaging provides Image Data Analysis as a service, providing users with open access to expertise of bioimage analysts working at internationally recognised facilities and institutes at different Euro-BioImaging Nodes. By providing the possibility to decouple image data acquisition from analysis, researchers using this standalone service can leverage the expertise of specialized analysts who may be only available at distinct locations. On one hand, outsourcing the development of advanced and intricate analysis pipelines to experienced analysts allows researchers to make the most out of their image data thus increasing the overall output of a research project within its lifespan. On the other hand, it gives bioimage analysts the opportunity to work on external projects, increasing the impact radius of their expertise and their visibility beyond home institutions. This enhances the recognition of skilled personnel for both image analysis and image acquisition in facilities (Kivinen et al., 2022; Schlaeppli et al., 2022). Pioneering work in this direction has been done by the NEUBIAS community (Cimini et al., 2020) which brings together bioimage analysts from facilities distributed across Europe to define, streamline and acknowledge the importance of image analysis in biological research. Through the NEUBIAS academy, they are actively working on providing training materials and courses, aimed at educating a new generation of bioimage analysts.

#### ***Facet 4: Sharing of image data***

Besides properly structured data and metadata, another hallmark of FAIR is sharing of research outputs beyond paper publication to the extent possible (Wilson et al., 2021). FAIR-advanced domains have established well-curated and deeply-integrated repositories for certain data types (e.g. PDB<sup>14</sup>, Genbank<sup>15</sup>, Uniprot<sup>16</sup>) and often the creation of these repositories leads to, or contributed at least to the success of, the development of new tools and even research areas (Feng et al., 2020; Swedlow et al., 2021). Over the last decade, the imaging community is witnessing the creation and more widespread adoption of bioimaging repositories and archives in an effort to establish a mature and functional bioimage data ecosystem (Ellenberg et al., 2018). There are three key components to such an ecosystem: (1) primary archives, which allow for rapid and direct deposition of data associated with publications from a broad domain with a limited amount of metadata; (2) added-value databases that house curated data with rich metadata that are highly integrated with other resources, and (3) topic-specific resources or portals, bringing together data resources from different modalities and scientific domains, hosting, for example, organ-specific or biomedical data.

A central, primary archive for bioimage data is currently provided by the BioImage Archive (BIA (Hartley et al., 2022)), which hosts data from all imaging modalities where a more specialized resource does not exist. To ensure minimal metadata quality, the BIA employs minimal

---

<sup>14</sup> <https://www.rcsb.org/>

<sup>15</sup> <https://www.ncbi.nlm.nih.gov/genbank/>

<sup>16</sup> <https://www.uniprot.org/>

metadata reporting according to REMBI. In contrast, the Image Data Resource (IDR (Williams et al., 2017)) is an added-value database for several light microscopic imaging modalities with highly-curated metadata, including high-content screening data, aiming to link the imaging data with other databases, such as those for genetic and chemical information, as well as cell and tissue phenotypes. Public archiving of raw 2D electron microscopy data underlying 3D cryo-EM protein structures and 3D volume EM is performed by the Electron Microscopy Public Image ARchive (EMPIAR (Iudin et al., 2023)). The Systems Science of Biological Dynamics repository and database (SSBD (Tohsato et al., 2016)) are a pair of primary archive and added-value databases for quantitative data of spatiotemporal dynamics of biological objects mostly obtained from microscopy. According to the FAIR principles, data should be shared as openly as possible, but as closed as necessary. This principle is especially important for sensitive and biomedical data that cannot be fully openly shared and deposited. Thus, restricted-access and topic-specific repositories for bioimaging data include for example the cancer imaging archive (Clark et al., 2013).

Given this variety of image repositories, it is important that the resources are well connected and coordinated, thus interoperable, to provide some level of consistency and thereby work towards creating a complete landscape of repositories to cover each type and modality of image data. Moreover, these data platforms will be most utilized when their deposition requirements and infrastructures are aligned with the current state of data production, not just data following tomorrow's quality standards. Consequently, we believe it is crucial to initiate and maintain direct contact between depositors and archives, and are currently establishing this contact through the efforts of data stewards. This mutual exchange of real-world problems will meaningfully shape the deposition pipelines, thereby lowering barriers, time and energy required to easily share bioimaging data. In this way, data generation and deposition will evolve in parallel, leading to the production of increasingly FAIR data over time.

However, even the best and most thorough deposition ecosystem would be worthless if no data were deposited. Therefore, it is crucial to raise awareness of the FAIR image repositories, for example by engaging directly with researchers and institutions facilitated by data stewards and research infrastructures as a common platform. Yet, for widespread adoption of bioimage data sharing, scientific journals have the long leverage to change standard practices by encouraging data sharing and providing targeted suggestions for repositories. As a step in the direction of FAIR science, they are increasingly requiring data availability statements to be included in the publication as well as directing authors to resources like FAIRsharing (Sansone et al., 2019), a catalog of databases, standards, and policies, that can help to discover suitable repositories. Infrastructures and communities are in active communication with journal representatives to define clearer incentives and recommendations for image-sharing platforms and procedures. In addition, permissive data availability should be more rewarded thereby highlighting and promoting well-documented and reproducible studies. Many ecosystem partners invest considerable effort to raise awareness of the additional benefits of data sharing such as increased recognition of scientific effort beyond papers (Wilson et al., 2021) including increased citations (Colavizza et al., 2020) and reduced burden of long-term data storage (Ellenberg et al., 2018). The popularity and usage of image repositories will increase as a result of these efforts. Our vision is to move to a place where data sharing alongside publication is as second nature for bioimage data as it is in other fields.

### ***Facet 5: Establishing trust in data to foster reuse***

The ultimate goal of the FAIR principles is to increase the value of scientific data by promoting its reuse in subsequent studies, thus getting more out of the effort expended to generate the data in the first place. In theory, if data is available in archives, data reuse should be possible. However, we currently see very little reuse of bioimaging data; in fact, there is a large discrepancy between bioimaging researchers who believe they could benefit from data reuse and those who actually incorporate image reuse into their research (Schmidt et al., 2022). Reasons for this discrepancy may be a lack of awareness of other datasets, so there is a need to spotlight well-documented datasets with rich information content. But even if the data can be located and retrieved, reuse requires that the data is of adequate quality, because data of questionable quality can be challenging to interpret and should not be reused (Miura & Nørrelykke, 2021). Indeed, a recent study in the fields of ecology/evolution found that nearly two-thirds of the data in public archives in these domains were "unusable" (Roche et al., 2015). Given this, it's not surprising that scientists are generally skeptical towards other researchers' data (Chan et al., 2021), which manifests itself as a reluctance to reuse deposited data. Thus, as it did in other disciplines like structural biology, the establishment of common data and metadata standards will also increase confidence in the deposited data and foster reuse. Community initiatives dedicated to data quality such as QUAREP-LiMi for Light Microscopy and the volumeEM community<sup>17</sup>, play a crucial role in the development and adoption of such standards while also underpinning them with the required broad support and credibility. Bioimaging data archives and databases are currently engaging with the community to work towards implementing metadata reporting standards as described above while also exploring mechanisms to evaluate and declare the quality of data reporting, thus highlighting well-documented datasets. In the future, it will also be valuable to implement and display internal quality controls of the image data itself and the associated imaging experiment, as is the case with structural biology data.

Additionally, reuse of image data will be enhanced by promoting the findability of image data from other life science domains by providing concrete and relevant cross-links between publicly available image data and other data types through curation. Several of these interdisciplinary communication approaches, such as 3DBionotes-WS<sup>18</sup>, a web service focusing on annotating structural models with relevant biochemical and biomedical information, and the Covid-19 disease map (Ostaszewski et al., 2021), are now working towards the integration of bioimage data.

Another important factor in facilitating reuse and increasing the credibility of data is that the license for reuse must be properly defined and the provenance of the data must be clear (Carbon et al., 2019; Huisman et al., 2019). Only when people are aware that they can use the data safely, and know exactly where it came from and how it was created, can they use it properly. Especially for sensitive data, it is important to clarify how, to what extent, for what purpose, and in what context others are allowed to reuse the data (Navale & McAuliffe, 2018). Euro-BioImaging, also in the context of European projects like BY-COVID and EUCAIM<sup>19</sup>, is involved in defining guidelines for both best practices and the ethics of sharing and reuse of these types of data. Furthermore, to encourage and normalize the increasingly widespread reuse of bioimaging data, it is imperative to credit data owners when their datasets are reused,

---

<sup>17</sup> <https://www.volumeem.org>

<sup>18</sup> <https://3dbionotes.cnb.csic.es/ws>

<sup>19</sup> <https://digital-strategy.ec.europa.eu/en/policies/cancer-imaging>

so that they get attributed for their scientific work beyond traditional publications (Pierce et al., 2019). This approach directs towards a future where the most important scientific output of researchers is not just the individual contributions to papers, but also the totality of datasets, pipelines, software, and protocols of value to the scientific community that they have produced.

In summary, progress is needed on all of the above-described facets of FAIR data to collectively increase the volume and quality of openly available bioimages, in order to build trust and thereby encourage reuse. Currently, however, the perceived value of archived datasets is still low, so we still lack a community that pioneers data reuse and favors reuse over new data generation. Recently, the Artificial Intelligence (AI) community has begun to emphasize the benefits of sharing bioimage data since Deep Learning algorithms need to be trained on large, diverse, and well-annotated datasets (Ouyang et al., 2022). This reinforces the need to prioritize data sharing and standardized metadata for the development of advanced AI algorithms for bioimage analysis. In this regard, Euro-BioImaging is coordinating the European project AI4Life<sup>20</sup>, which aims to bridge the gap between computer and life sciences communities by offering infrastructure and services to support life scientists in adopting Machine Learning solutions for bioimage analysis. Ongoing discussions within AI4Life focus on developing annotation standards and improving the submission, presentation, and retrieval of AI-ready images and annotations. To fully leverage the potential of AI, it is crucial to ensure that the metadata associated with both the AI models and the data used to train them are of high quality. Standardized metadata is therefore essential for effective data exploitation and reuse, and a priority for the scientific community.

### ***Facet 6: Dissemination of FAIR practices***

The mostly technical facets of FAIR data described above cannot be considered as separate entities, but must be treated as interdependent and interrelated. Progress in one area advances the others, and conversely, lack of progress in one area hinders advances in others. Therefore, careful consideration of their intersections is necessary to promote progress in all facets simultaneously. Additionally, the imaging community is very diverse and highly engaged in this topic, leading to the formation of a number of different community initiatives dedicated to various aspects of the image data challenge. These initiatives are all individually highly valuable, but to ensure efficiency of the mostly volunteer work involved in these, facilitating their alignment with each other and with science policy is critical. Thus, coordination, integration, and communication at the regional, national and global levels are equally necessary to successfully integrate FAIR image data initiatives into a landscape of independent community initiatives and science policy interests.

This is a key priority and responsibility of large-scale research infrastructures like Euro-BioImaging which have the necessary connections within the landscape and have acquired an in-depth understanding of science policy developments. With the creation of its Image Data Expert Group, Euro-BioImaging has become a source of information and common discussion platform on current developments of these initiatives for anyone engaged with bioimaging data. Great emphasis is placed on the progress of our participation in several Horizon Europe funded projects and their practical results relevant to the European image data communities. These projects are currently providing critical resources for the technical facets of the FAIR bioimage data as well as community support. In the future, further relevant EU projects will increase the

---

<sup>20</sup> <https://ai4life.eurobioimaging.eu/>

presence of bioimaging data in the European Health Data Space of digital infrastructures dealing with the management and re-use of clinical data.

The overall expected outcome of these European-level projects is to facilitate the participation and contribution of the European image data community to the EOSC, a key strategic priority of the European Commission contributing to the European Digital Decade. European researchers but also innovators, businesses and citizens can publish, discover and reuse data, tools and services for research, innovation and education within the federated and open multidisciplinary environment of EOSC. By proactively engaging with the EOSC, Euro-BioImaging's image data strategy and services contribute to and align with EOSC developments thus allowing researchers to take full advantage of cloud services. This creates and fosters the impact of bioimaging in FAIR big data research on the European landscape and beyond.

But successful plans for image data sharing cannot be restricted to one continent. To this end, the Global BioImaging<sup>21</sup> network was established as an exchange platform for staff and managers working in national and regional imaging infrastructures and communities representing currently 27 countries around the world<sup>22</sup>. Here, international working groups such as the one for Image Data Management convene to ensure that open discussion on data formats and standardization takes place internationally, and communities can align their efforts also globally (Swedlow et al., 2021).

## Conclusions

As the volume and potential of biological images grows, everyone from individual researchers and institutions to funding agencies and the general public has expectations regarding image and metadata quality, accessibility, and potential for analyzing the resource. Following the growing adoption of Open Science policy and the FAIR principles for scientific data, the need for image data services is growing rapidly. Together with other national and international imaging initiatives, Euro-BioImaging has set out to tackle these challenges and is helping the research community to collectively develop much-needed solutions for managing, sharing and analyzing image data. Each member of the FAIR imaging ecosystem influences the area in which they are most engaged and involved, but the entire ecosystem has to work together to impact the advancement of the field as a whole. Once the combined efforts described by community members in this Special Edition are successfully and widely adopted, we are anticipating a drastic shift in the way images are used, shared and reused in biology. In the future, new research studies will build on re-analysis of existing image data and cross-references between already existing datasets rather than generating new data. In this way, image data will become a global resource that everybody can use beyond national boundaries, creating research opportunities in many places which are still lacking imaging technologies for data acquisition. As a whole, more readily available image data will also be a catalyst for greater public interest and trust in science. These efforts echo the unprecedented opportunities of today's imaging technologies in the life sciences, and the need for a predictive

---

<sup>21</sup> <https://globalbioimaging.org/>

<sup>22</sup> Advanced Bioimaging Support (Japan), BioImaging North America, Canada BioImaging, Euro-BioImaging, India BioImaging, Latin America BioImaging, Laboratorio Nacional de Microscopía Avanzada (Mexico), Microscopy Australia, National Imaging Facility (Australia), SingaScope (Singapore) and South Africa BioImaging

understanding of biology, with pandemics and the effects of climate change on our ecosystems increasingly affecting our everyday life.

## References

- 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>
- Berman, H. M., Vallat, B., & Lawson, C. L. (2020). The data universe of structural biology. *IUCrJ*, *7*(4), 630–638. <https://doi.org/10.1107/S205225252000562X>
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., & Wiswedel, B. (2009). KNIME - the Konstanz information miner: Version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter*, *11*(1), 26–31. <https://doi.org/10.1145/1656274.1656280>
- 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., Kukat, 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>
- Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X., & Greene, C. S. (2020). Responsible, practical genomic data sharing that accelerates research. *Nature Reviews Genetics*, *21*(10), 615–629. <https://doi.org/10.1038/s41576-020-0257-5>
- Carbon, S., Champieux, R., McMurry, J. A., Winfree, L., Wyatt, L. R., & Haendel, M. A. (2019). An analysis and metric of reusable data licensing practices for biomedical resources. *PLOS ONE*, *14*(3), e0213090. <https://doi.org/10.1371/journal.pone.0213090>
- Chan, V., Gherardini, P. F., Krummel, M. F., & Fragiadakis, G. K. (2021). A “data sharing trust” model for rapid, collaborative science. *Cell*, *184*(3), 566–570. <https://doi.org/10.1016/j.cell.2021.01.006>
- Chervitz, S. A., Deutsch, E. W., Field, D., Parkinson, H., Quackenbush, J., Rocca-Serra, P., Sansone, S.-A., Stoeckert, C. J., Taylor, C. F., Taylor, R., & Ball, C. A. (2011). Data Standards for Omics Data: The Basis of Data Sharing and Reuse. In B. Mayer (Hrsg.), *Bioinformatics for Omics Data* (Bd. 719, S. 31–69). Humana Press. [https://doi.org/10.1007/978-1-61779-027-0\\_2](https://doi.org/10.1007/978-1-61779-027-0_2)
- Cimini, B. A., Nørrelykke, S. F., Louveaux, M., Sladoje, N., Paul-Gilloteaux, P., Colombelli, J., & Miura, K. (2020). The NEUBIAS Gateway: A hub for bioimage analysis methods and materials. *F1000Research*, *9*, 613. <https://doi.org/10.12688/f1000research.24759.1>
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *Journal of Digital Imaging*, *26*(6), 1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
- Colavizza, G., Hrynaszkiewicz, 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>
- David, R., Richard, Audrey S., Connellan, Claire, Lauer, Katharina B., Chiusano, Maria Luisa, Goble,

- Carole, Houde, Martin, Kemmer, Isabel, Keppler, Antje, Lieutaud, Philippe, Ohmann, Christian, Panagiotopoulou, Maria, Raza Khan, Sara, Rybina, Arina, Soiland-Reyes, Stian, Wit, Charlotte, Wittner, Rudolf, Andrade Buono, Rafael, Marsh, Sarah Arnaud, ... Ewbank, Jonathan. (2023). *Umbrella Data Management Plans to integrate FAIR data: Lessons from the ISIDORE and BY-COVID consortia for pandemic preparedness*.  
<https://doi.org/10.5281/ZENODO.7517068>
- Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4), 316–319. <https://doi.org/10.1038/nbt.3820>
- 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>
- 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>
- Feng, Z., Verdigué, N., Di Costanzo, L., Goodsell, D. S., Westbrook, J. D., Burley, S. K., & Zardecki, C. (2020). Impact of the Protein Data Bank Across Scientific Disciplines. *Data Science Journal*, 19(1), 25. <https://doi.org/10.5334/dsj-2020-025>
- Goble, C., Cohen-Boulakia, S., Soiland-Reyes, S., Garijo, D., Gil, Y., Crusoe, M. R., Peters, K., & Schober, D. (2020). FAIR Computational Workflows. *Data Intelligence*, 2(1–2), 108–121. [https://doi.org/10.1162/dint\\_a\\_00033](https://doi.org/10.1162/dint_a_00033)
- Goble, C., Soiland-Reyes, S., Bacall, F., Owen, S., Williams, A., Eguinoa, I., Driesbeke, B., Leo, S., Pireddu, L., Rodríguez-Navas, L., Fernández, J. M., Capella-Gutierrez, S., Ménager, H., Grüning, B., Serrano-Solano, B., Ewels, P., & Coppens, F. (2021). *Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory*. <https://doi.org/10.5281/ZENODO.4605654>
- 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>
- 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>
- 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>
- Huisman, M., Hammer, 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., Grunwald, D., & Strambio-De-Castillia, C. (2019). *A perspective on Microscopy Metadata: Data provenance and quality control*. <https://doi.org/10.48550/ARXIV.1910.11370>
- Iudin, A., Korir, P. K., Somasundharam, S., Weyand, S., Cattavittello, C., Fonseca, N., Salih, O., Kleywegt, G. J., & Patwardhan, A. (2023). EMPIAR: The Electron Microscopy Public Image Archive. *Nucleic Acids Research*, 51(D1), D1503–D1511. <https://doi.org/10.1093/nar/gkac1062>

- 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>
- 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/S2633903X21000039>
- Kalaš, M., Plantard, L., Nataša Sladoje, Lindblad, J., Kirschmann, M. A., Jones, M., Chessel, A., Scholz, L. A., Rössler, F., Dufour, A., Bogovic, J., Zhang, C., Waithe, D., Sampaio, P., Paavolainen, L., Hörl, D., Munck, S., Golani, O., Moore, J., ... And Welcoming New Contributors. (2019). *EDAM-bioimaging: The ontology of bioimage informatics operations, topics, data, and formats (2019 update)*. <https://doi.org/10.7490/F1000RESEARCH.1116432.1>
- Kivinen, K., van Luenen, H. G. A. M., 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). <https://doi.org/10.15252/embr.202255734>
- Kunis, S., Hänsch, S., Schmidt, C., Wong, F., Strambio-De-Castillia, C., & Weidtkamp-Peters, S. (2021). MDEmic: A metadata 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>
- Marqués, 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>
- Miura, K. (Hrsg.). (2016). *Bioimage data analysis*. Wiley-VCH.
- Miura, K., & Nørrelykke, S. F. (2021). Reproducible image handling and analysis. *The EMBO Journal*, *40*(3). <https://doi.org/10.15252/emboj.2020105889>
- Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. <https://doi.org/10.12688/f1000research.29032.2>
- 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>
- Moore, J., Basurto-Lozada, D., Besson, S., Bogovic, J., Brown, E. M., Burel, J.-M., 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., Küpcü Yoldaş, A., de la Villegeorges, A. le T., ... Swedlow, J. R. (2023). *OME-Zarr: A cloud-optimized bioimaging file format with international community support* [Preprint]. *Bioinformatics*. <https://doi.org/10.1101/2023.02.17.528834>
- Morris, C. (2018). The Life Cycle of Structural Biology Data. *Data Science Journal*, *17*, 26. <https://doi.org/10.5334/dsj-2018-026>
- Navale, V., & McAuliffe, M. (2018). Long-term preservation of biomedical research data. *F1000Research*, *7*, 1353. <https://doi.org/10.12688/f1000research.16015.1>
- Ostaszewski, M., Niarakis, A., Mazein, A., Kuperstein, I., Phair, R., Orta-Resendiz, A., Singh, V., Aghamiri, S. S., Acencio, M. L., Glaab, E., Ruepp, A., Fobo, G., Montrone, C., Brauner, B., Frishman, G., Monraz Gómez, L. C., Somers, J., Hoch, M., Kumar Gupta, S., ... the COVID-19



- Disease Map Community. (2021). COVID-19 Disease Map, a computational knowledge repository of virus-host interaction mechanisms. *Molecular Systems Biology*, *17*(12). <https://doi.org/10.15252/msb.202110851>
- Ouyang, W., Beuttenmueller, F., Gómez-de-Mariscal, E., Pape, C., Burke, T., Garcia-López-de-Haro, C., Russell, C., Moya-Sans, L., de-la-Torre-Gutiérrez, C., Schmidt, D., Kutra, D., Novikov, M., Weigert, M., Schmidt, U., Bankhead, P., Jacquemet, G., Sage, D., Henriques, R., Muñoz-Barrutia, A., ... Kreshuk, A. (2022). *Biolmage Model Zoo: A Community-Driven Resource for Accessible Deep Learning in Biolmage Analysis* [Preprint]. Bioinformatics. <https://doi.org/10.1101/2022.06.07.495102>
- 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>
- Paul-Gilloteaux, P., Tosi, S., Hériché, J.-K., Gaignard, A., Ménager, H., Marée, R., Baecker, V., Klemm, A., Kalaš, M., Zhang, C., Miura, K., & Colombelli, J. (2021). Bioimage analysis workflows: Community resources to navigate through a complex ecosystem. *F1000Research*, *10*, 320. <https://doi.org/10.12688/f1000research.52569.1>
- Pepe, A., Goodman, A., Muench, A., Crosas, M., & Erdmann, C. (2014). How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. *PLoS ONE*, *9*(8), e104798. <https://doi.org/10.1371/journal.pone.0104798>
- Pergl, R., Hooft, R., Suchánek, M., Knaisl, V., & Slifka, J. (2019). “Data Stewardship Wizard”: A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning. *Data Science Journal*, *18*, 59. <https://doi.org/10.5334/dsj-2019-059>
- Pfander, C., Bischof, J., Childress-Poli, M., Keppler, A., Viale, A., Aime, S., & Eriksson, J. E. (2022). Euro-Biolmaging – Interdisciplinary research infrastructure bringing together communities and imaging facilities to support excellent research. *IScience*, *25*(2), 103800. <https://doi.org/10.1016/j.isci.2022.103800>
- Pierce, H. H., Dev, A., Statham, E., & Bierer, B. E. (2019). Credit data generators for data reuse. *Nature*, *570*(7759), 30–32. <https://doi.org/10.1038/d41586-019-01715-4>
- 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>
- Roche, D. G., Kruuk, L. E. B., Lanfear, R., & Binning, S. A. (2015). Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLOS Biology*, *13*(11), e1002295. <https://doi.org/10.1371/journal.pbio.1002295>
- 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>
- Sansone, S.-A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A. L., Thurston, M., & the FAIRsharing Community. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology*, *37*(4), 358–367. <https://doi.org/10.1038/s41587-019-0080-8>
- 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., Tischer, C., Uhlmann, V., Verkade, P., ... Brazma, A. (2021). REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology. *Nature Methods*, *18*(12), 1418–1422. <https://doi.org/10.1038/s41592-021-01166-8>
- Schlaeppli, A., Adams, W., Haase, R., Huisken, J., MacDonald, R. B., Eliceiri, K. W., & Kugler, E. C. (2022). Meeting in the Middle: Towards Successful Multidisciplinary Bioimage Analysis Collaboration. *Frontiers in Bioinformatics*, *2*, 889755. <https://doi.org/10.3389/fbinf.2022.889755>
- 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>
- Schmied, C., Nelson, M., Avilov, S., Bakker, G.-J., Bertocchi, C., Bischof, J., Boehm, U., Brocher, J., Carvalho, M., Chiritescu, C., Christopher, J., Cimini, B., 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 analysis*. <https://doi.org/10.48550/ARXIV.2302.07005>
- 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>
- 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>
- 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>
- 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: MICROSCOPY AND DATA MANAGEMENT. *Molecular Reproduction and Development*, *82*(9), 630–634. <https://doi.org/10.1002/mrd.22538>
- 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*(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- 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). Image Data Resource: A bioimage data integration and publication platform. *Nature Methods*, *14*(8), 775–781. <https://doi.org/10.1038/nmeth.4326>
- Wilson, S. L., Way, G. P., Bittremieux, W., Armache, J., 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>