nature methods

Perspective

https://doi.org/10.1038/s41592-024-02585-z

Enabling global image data sharing in the life sciences

Received: 12 February 2024

Accepted: 3 December 2024

Published online: 28 March 2025

Check for updates

Peter Bajcsy [1], Sreenivas Bhattiprolu[2], Katy Börner [3], Beth A. Cimini [4], Lucy Collinson [5], Jan Ellenberg [6], Reto Fiolka [7], Maryellen Giger [8], Wojtek Goscinski[9], Matthew Hartley [10], Nathan Hotaling [11], Rick Horwitz[12], Florian Jug[13], Isabel Kemmer [14], Anna Kreshuk [6], Emma Lundberg [15,16], Aastha Mathur [14], Kedar Narayan[17,18], Shuichi Onami [19], Anne L. Plant [1], Fred Prior [20], Jason R. Swedlow [21], Adam Taylor [22] & Antje Keppler [14] ✉

Despite the importance of imaging in biological and medical research, a large body of informative and precious image data never sees the light of day. To ensure scientific rigor as well as the reuse of data for scientific discovery, image data need to be made FAIR (findable, accessible, interoperable and reusable). Image data experts are working together globally to agree on common data formats, metadata, ontologies and supporting tools toward image data FAIRification. With this Perspective, we call on public funders to join these efforts to support their national scientists. What researchers most urgently need are openly accessible resources for image data storage that are operated under long-term commitments by their funders. Although existing resources in Australia, Japan and Europe are already collaborating to enable global image data sharing, these efforts will fall short unless more countries invest in operating and federating their own open data resources. This will allow us to harvest the enormous potential of existing image data, preventing substantial loss of unrealized value from past investments in imaging acquisition infrastructure.

Image data are not only intriguing to look at but they are also the fastest growing data resource in the life sciences. Biological and medical images are complex and deep in information content based on the nature of many imaging modalities, making the data conducive to reuse. With simultaneous rapid advances in and access to computation power, the scientific community is only beginning to tap into this invaluable resource. Acquiring and managing image data according to the FAIR principles[1] is therefore being requested by major public funders such as the European Union (https://digital-strategy.ec.europa.eu/en/policies/data-act) and the National Institutes of Health (https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html). Beyond supporting their open science policy objectives, many of the most exciting opportunities for deriving value from sharing and reusing imaging data at scale rely on the ability to curate and centrally access those data. This aggregation based on common standards not only enables very large-scale artificial intelligence (AI) and machine learning model training but also the creation of cross-modality and cross-domain benchmarks and reference datasets. These gains are not possible when FAIR data are siloed and fragmented in many different places, often without public access or consistent organization. The key to easy data access lies in consolidating data in a limited number of locations that are connected with each other and that follow compatible standards for data and metadata storage – that is, they are interoperable.

Australia, Japan and the European Union have established nationally funded, coordinated research infrastructures for life science and biomedical imaging (collectively 'BioImaging'). These BioImaging research infrastructures are the output of concerted long-term strategic planning efforts from key stakeholders that include academic,

A full list of affiliations appears at the end of the paper ✉e-mail: antje.keppler@eurobioimaging.eu

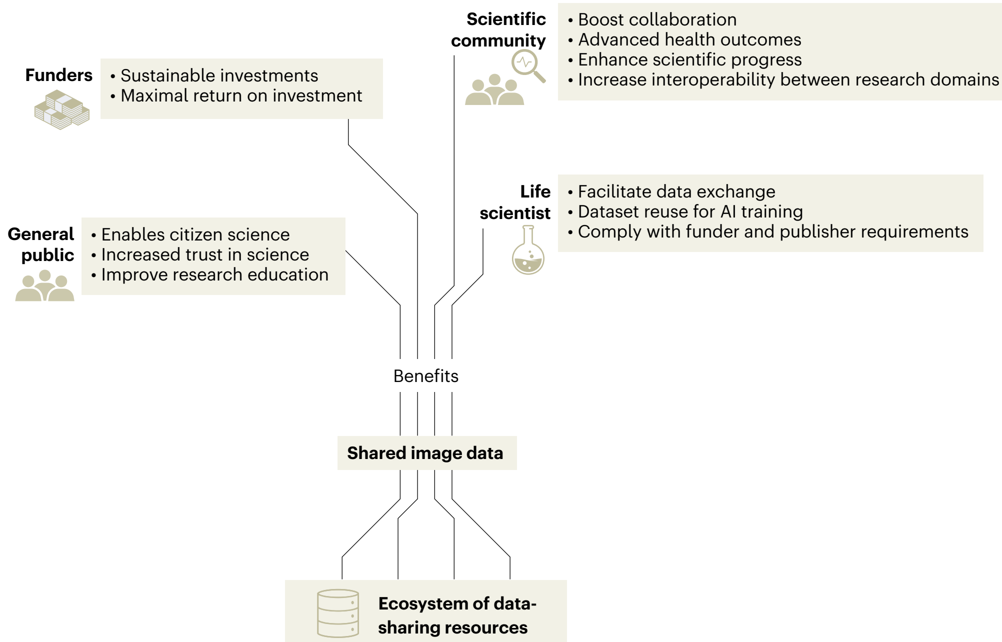

Fig. 1 | Data sharing is too impactful not to do. An outline of the tremendous benefits of globally shared image data. Data sharing drives open science, empowers life scientists and enables reuse of valuable FAIR image datasets for creating new knowledge.

funding and political as well as industrial partners. They are now in operation and deliver technology and community support for quality image data acquisition, management, analysis and publication for their stakeholders. They are supported by dedicated image data repositories, such as the BioImage Archive[2], the Image Data Resource[3] and the Electron Microscopy Public Image Archive[4] in Europe; the Systems Science of Biological Dynamics database[5] in Japan; and the Australian Imaging Service repository in Australia (https://australian-imaging-service.github.io/) that together hold around 6 PB of image data. Publicly funded projects and community initiatives are also actively building and supporting image data formats (https://ngff.openmicroscopy.org/) and applications for managing, analyzing and sharing data using the most advanced technologies, including AI (https://ai4life.eurobioimaging.eu/). Collaboration between academic and commercial organizations is maturing and beginning to provide a powerful ecosystem for the development and eventual scaling of new technologies and products.

However, these regionally limited efforts are only a starting point of a future global ecosystem for sharing image data. We therefore urge funders to invest in open-access infrastructure supporting image data acquisition, management and sharing, which then can be coordinated around the world. This is a necessary step toward realizing the substantial value that is currently lying dormant in underused image data and instead can catalyze the exploitation of this invaluable resource, which was largely generated through previous public investments.

Image data sharing: cui bono?

Everyone can benefit tremendously if open-access image data resources were available to them (Fig. 1). Life scientists, as producers of large image datasets, can strengthen their research output by opening up the data that support their publications and exchanging the data more easily with existing and future researchers. AI experts, data analysts and software developers can use the available image datasets for training, testing and developing their tools. Trainers and educators can expand their portfolio of training material and expedite the sharing of knowledge and new tools. Researchers can follow their data management plan and upload their final data for publication in these resources, meaningfully fulfilling the requirements of their funders and publishers. Funders can sustain their research investments and enable the reuse of existing research data for the creation of new knowledge. Finally, the image data become accessible to the general public, which can lead to increased trust in and support of public research and education as well as new and unanticipated discoveries.

As a key technology in the biological and biomedical sciences and beyond, imaging is enabling scientists to map and understand the complex spatiotemporal patterns of biological processes in greater detail than ever before. Whereas other fields such as genomics, proteomics and so on have benefited heavily from investing into extensive data-sharing mechanisms and practices, recent advances in computational infrastructure have only now given the imaging community the same level of aspiration to make image data an accessible, shared global resource.

We are convinced that sharing FAIR image data from all domains will substantially increase interoperability among the different research fields in the life sciences. Most importantly, it will allow a comprehensive view of biological processes at different levels of organization, from molecular to cellular to whole organism, and thereby will advance both basic biological as well as translational research. The integration of microscopy and medical imaging can provide a better understanding of disease mechanisms, particularly in the context of complex diseases that involve multiple organ systems and cellular processes. Preclinical and clinical imaging techniques, such as magnetic resonance imaging, provide valuable data on anatomical structures and physiological functions. Integrating these data with light microscopy images can provide a more accurate and detailed understanding of the underlying biology. Such image data integration will support and grow the field of bioimage informatics, trigger advances in drug discovery and development and help to identify potential therapeutic targets as well as to evaluate drug efficacy and safety. If we foster image data sharing across domains, then we will not only democratize science but also enable new discoveries by linking different types of image datasets across the entire spectrum of life.

What are the technical challenges?

There are several technical challenges that make sharing image data more difficult than sharing genomic data[6] or protein structures[7], which

https://doi.org/10.1038/s41592-024-02585-z

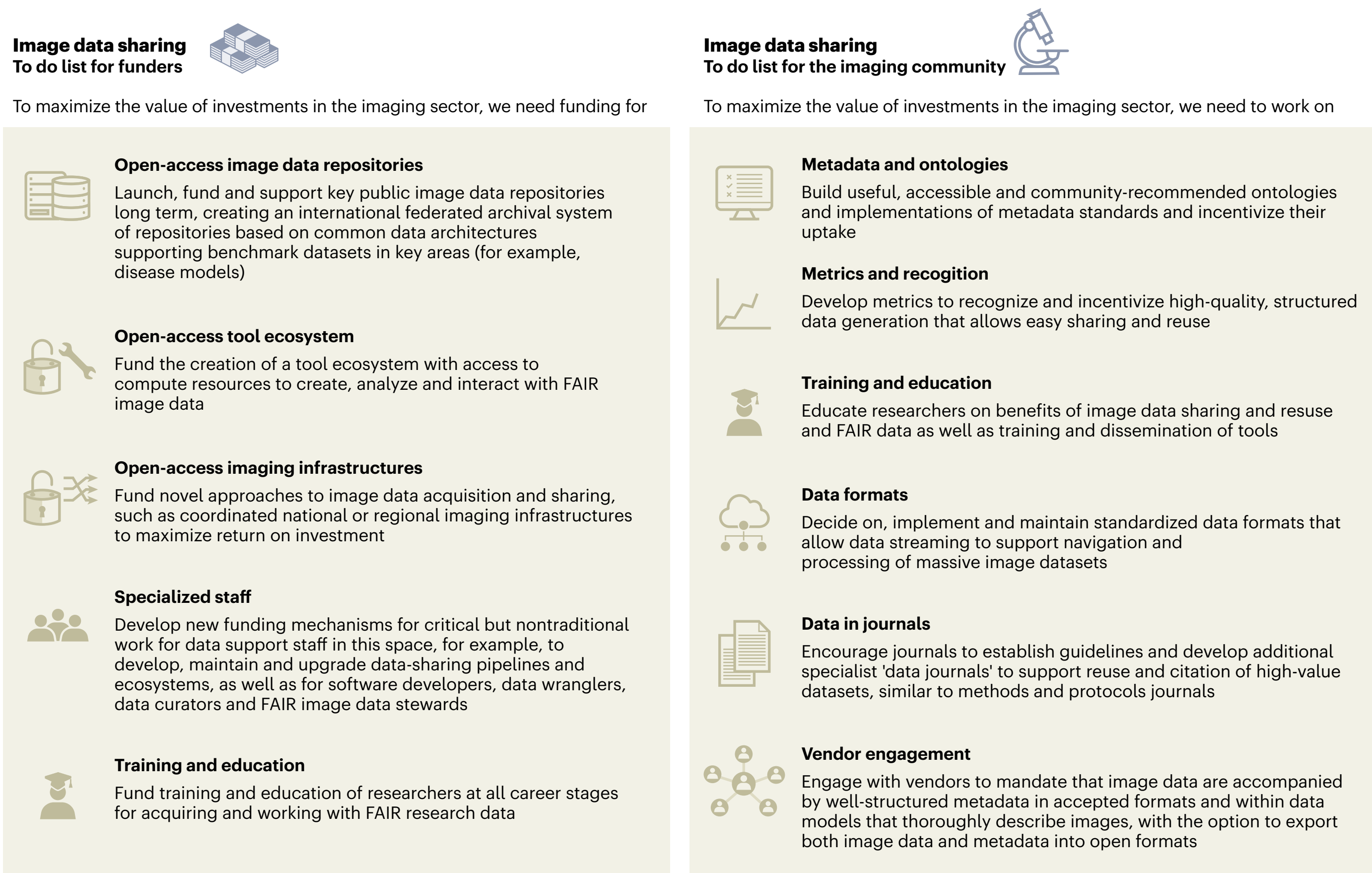

Fig. 2 | Actionable steps towards a global image data ecosystem. How public decision makers, funders and the scientific community together can enable global image data sharing.

are currently the best examples of open shared data and data repositories. One is that the data may be very large with individual files of sizes from 100 GB to 10 TB and beyond[8] and require either high-speed internet connections to be uploaded and downloaded or physical transportation on external drives. These practices lead to duplication of data, making data management and sharing expensive and complex for large datasets over time. For supporting interconnectivity between research institutions and data repositories, a robust technical infrastructure, including large data storage capacity and powerful computing resources, is therefore required, as many research institutions still do not have the necessary infrastructure to support image data sharing.

Another key challenge is the need for careful management and organization to ensure that image data are accessible to and, ideally, reusable by other researchers. Although medical images typically are easily shared, as most are acquired in DICOM format[9], biological imaging data sharing is still limited because of the lack of standardization in image data formats, ontologies and metadata. This makes it difficult for researchers to compare and analyze data from different sources. Ensuring that the data are properly labeled, annotated and stored in a standardized format can require substantial time and resources[10]; however, it is an essential yet often underappreciated process. In addition, proper curation and harmonization of imaging data and associated metadata are needed to ensure machine actionability in future use of datasets at scale.

Finally, depending on the context, intellectual property issues, questions of ownership and data privacy concerns may arise. Because of this, researchers may be reluctant to share their image data, and, in case of medical imaging data, they must take steps to ensure that the data are de-identified and cannot be traced back to individual patients. Medical image datasets must remain in their country of origin, and the researchers together with the legislative authorities need to find common legal and ethical solutions. Also, there may be cultural barriers, such as a lack of trust or a reluctance to share image data due to concerns about competition or ownership.

Which solutions are currently developed?

Despite these challenges, community initiatives around the world are actively working on solutions to share image data globally[11]. However, those developments are not sustainable without funding to support open-access image data storage. We must move forward with the input from all stakeholders involved in the process of image data production, sharing and analysis, including scientists acquiring data, operators of large-scale data infrastructures and data experts. Educational opportunities addressing the importance of quality-managed and harmonized data are increasing, but they need to be scaled and embedded in each researcher's training curriculum from the beginning. It is essential to recognize and reward the efforts of everyone involved in producing and curating FAIR data, including specialized data support staff, such as FAIR data stewards, data curators, research software engineers and so on. Many current funding plans to promote data sharing fail to adequately include these roles, resulting in poor implementation and limited uptake of developed solutions. Although establishment of these positions is on the rise, wider recognition and sustained funding is essential to support and secure these invaluable positions. Efforts similar to those driven by Global BioImaging for imaging scientists working in core facilities[12] are needed to define career progression and to develop measures to support professional development and job security of these important positions.

In the long term, global sharing of image data can only be achieved with a globally federated model. This requires that added value data resources dedicated to particular use cases or domains are stored and made available through a small number of centralized data archives or other open-access servers. These archives must provide both direct data hosting as well as indexing and search capabilities. This model must be built on shared core metadata models for interoperability based on international guidelines, such as Recommended Metadata for Biological Images[13], which can be extended for domain-specific applications such as Minimum Information about highly multiplexed Tissue Imaging[14]. In addition to metadata formats, effective use of data requires standardized image formats. In light of the exponential growth in the volume of new image data, it is critical that the imaging community collectively agrees on standard image formats with immediate effect. To guarantee the widest possible adoption, it is essential to define these formats in discussion with vendors, a process that has long been carried out by tireless community efforts driven by initiatives such as the Open Microscopy Environment (OME)[15], which is now working toward future-proof next-generation file formats designed for large-scale, cloud-enabled use, such as OME-Zarr[16,17]. Allowing data to be easily located, accessed through consistent mechanisms in standardized formats and coupled with rich metadata unlocks valuable large-scale applications. These include training universal image classifiers to support automated quality-control pipelines for researchers and instrument manufacturers, automated segmentation of common cell structures for use as biomarkers of disease or to better understand mechanisms of action and prediction of physiological or biomolecular properties of cells to speed up biomanufacturing and reduce assay variability and consumable use, and so on. By adopting these approaches, we can transform petabytes to exabytes of immensely valuable scientific data, representing enormous funder investments, into a global resource that will be maximally useful for the global community.

How to go forward?

Although enabling global image data sharing requires a concerted effort and endurance from researchers, institutions, publishers and industry to establish internationally recognized best practices, it foremost requires a long-term commitment by public funders and their investments in essential technical infrastructure and data resources. Funders need to recognize that they currently lose substantial value from their previous investments into imaging acquisition infrastructure unless they provide the essential resources for storing and making the resulting image data accessible. It is essential that the imaging community and funders join forces to establish a sustainable and profitable image data ecosystem ensuring optimal output for all stakeholders (Fig. 2). We are looking forward to an exciting and fruitful dialog with national funders on the suggested action items presented here.

References

1. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
2. Hartley, M. et al. The BioImage Archive — building a home for life-sciences microscopy data. *J. Mol. Biol.* **434**, 167505 (2022).
3. Williams, E. et al. Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods* **14**, 775–781 (2017).
4. Iudin, A. et al. EMPIAR: the Electron Microscopy Public Image Archive. *Nucleic Acids Res.* **51**, D1503–D1511 (2023).
5. Tohsato, Y., Ho, K. H. L., Kyoda, K. & Onami, S. SSBD: a database of quantitative data of spatiotemporal dynamics of biological phenomena. *Bioinformatics* **32**, 3471–3479 (2016).
6. Arita, M., Karsch-Mizrachi, I. & Cochrane, G. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* **49**, D121–D124 (2021).
7. wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **47**, D520–D528 (2019).
8. Shapson-Coe, A. et al. A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. *Science* **384**, eadk4858 (2024).
9. Larobina, M. Thirty years of the DICOM standard. *Tomography* **9**, 1829–1838 (2023).
10. Swedlow, J. R. et al. A global view of standards for open image data formats and repositories. *Nat. Methods* **18**, 1440–1446 (2021).
11. Kemmer, I. et al. Building a FAIR image data ecosystem for microscopy communities. *Histochem. Cell Biol.* **160**, 199–209 (2023).
12. Wright, G. D. et al. Recognising the importance and impact of imaging scientists: global guidelines for establishing career paths within core facilities. *J. Microsc.* **294**, 397–410 (2024).
13. Sarkans, U. et al. REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology. *Nat. Methods* **18**, 1418–1422 (2021).
14. Schapiro, D. et al. MITI minimum information guidelines for highly multiplexed tissue images. *Nat. Methods* **19**, 262–267 (2022).
15. Linkert, M. et al. Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**, 777–782 (2010).
16. Moore, J. et al. OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies. *Nat. Methods* **18**, 1496–1498 (2021).
17. Moore, J. et al. OME-Zarr: a cloud-optimized bioimaging file format with international community support. *Histochem. Cell Biol.* **160**, 223–251 (2023).
18. Bajcsy, P. et al. Enabling global image data sharing in the life sciences. Preprint at https://doi.org/10.48550/arXiv.2401.13023 (2024).
19. Bialy, N. et al. Harmonizing the generation and pre-publication stewardship of FAIR image data. Preprint at https://doi.org/10.48550/arXiv.2401.13022 (2024).

Acknowledgements

K.N. is funded by federal funds from the National Cancer Institute, National Institutes of Health, under contract 75N91019D00024. R.F. is funded by the National Cancer Institute (grants R33 CA235254 and U54CA268072) and the National Institute of General Medical Sciences (grant R35 GM133522). I.K. receives funding from EU Horizon 2020 (grant agreement 101046203 (BY-COVID)) and from the European Union under grant agreement 101130986 (EVOLVE). This Perspective is based on the more extensive white paper 'Enabling global image data sharing in the life sciences' (ref. 18), which is published with a closely related companion entitled 'Harmonizing the generation and pre-publication stewardship of FAIR image data', which can be found at ref. 19. Disclaimer: commercial products are identified in this document to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the products identified are necessarily the best available for the purpose. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US government.

Author contributions

This Perspective is authored and reviewed by members of an ad hoc international working group who are listed alphabetically as coauthors of the Perspective. I.K. prepared the figures.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Antje Keppler.

Peer review information *Nature Methods* thanks Kelly Rogers and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Rita Strack, in collaboration with the *Nature Methods* team.

Reprints and permissions information is available at www.nature.com/reprints.

[1]National Institute of Standards and Technology, Gaithersburg, MD, USA. [2]ZEISS Microscopy Customer Center, Dublin, OH, USA. [3]Intelligent Systems Engineering, Indiana University, Bloomington, IN, USA. [4]Imaging Platform, Broad Institute, Cambridge, MA, USA. [5]Electron Microscopy Science Technology Platform, Francis Crick Institute, London, UK. [6]European Molecular Biology Laboratory, Heidelberg, Germany. [7]Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas, TX, USA. [8]Department of Radiology and Committee on Medical Physics, University of Chicago, Chicago, IL, USA. [9]National Imaging Facility, Brisbane, Queensland, Australia. [10]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK. [11]National Center for Advancing Translational Science, National Institutes of Health, Rockville, MD, USA. [12]Allen Institute for Cell Science, Seattle, WA, USA. [13]Fondazione Human Technopole, Milan, Italy. [14]Euro-BioImaging Bio-Hub, European Molecular Biology Laboratory, Heidelberg, Germany. [15]Stanford University, Stanford, CA, USA. [16]SciLifeLab, KTH Royal Institute of Technology, Stockholm, Sweden. [17]Center for Molecular Microscopy, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [18]Cancer Research Technology Program, Frederick National Laboratory for Cancer Research, Frederick, MD, USA. [19]RIKEN Center for Biosystems Dynamics Research, Kobe, Japan. [20]Department of Biomedical Informatics, Department of Radiology, University of Arkansas for Medical Sciences, Little Rock, AR, USA. [21]Divisions of Computational Biology and Molecular, Cell and Developmental Biology, University of Dundee, Dundee, UK. [22]Sage Bionetworks, Seattle, WA, USA. ✉e-mail: antje.keppler@eurobioimaging.eu