OME2024 NGFF Challenge Results

Josh Moore, Research Software Engineer

creative commons
10.5281/zenodo.14234608

Next-generation file formats (NGFF)

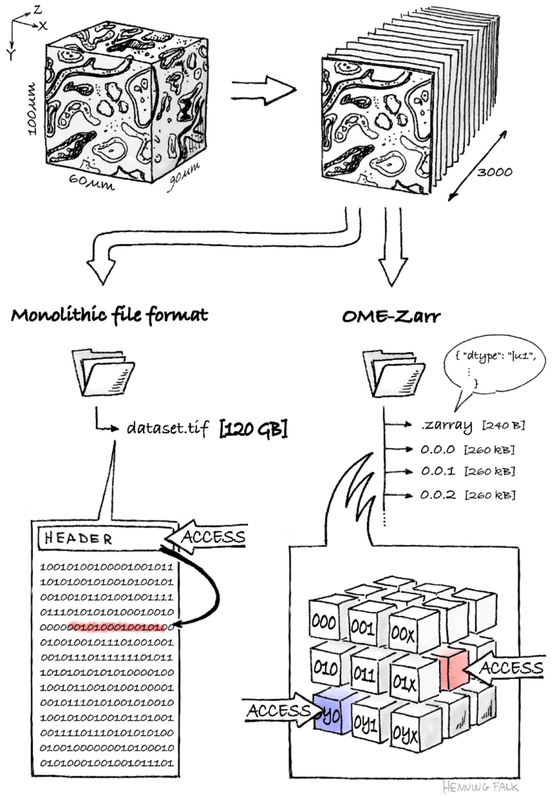


Image “Monolithic vs. chunked” by Henning Falk
https://github.com/zarr-developers/zarr-illustrations-falk-2022
©2022 NumFOCUS (CC BY 4.0 license)

Def. *(n)* directory-like layout of data rather than a single file, header information is available in formats like JSON or YAML.

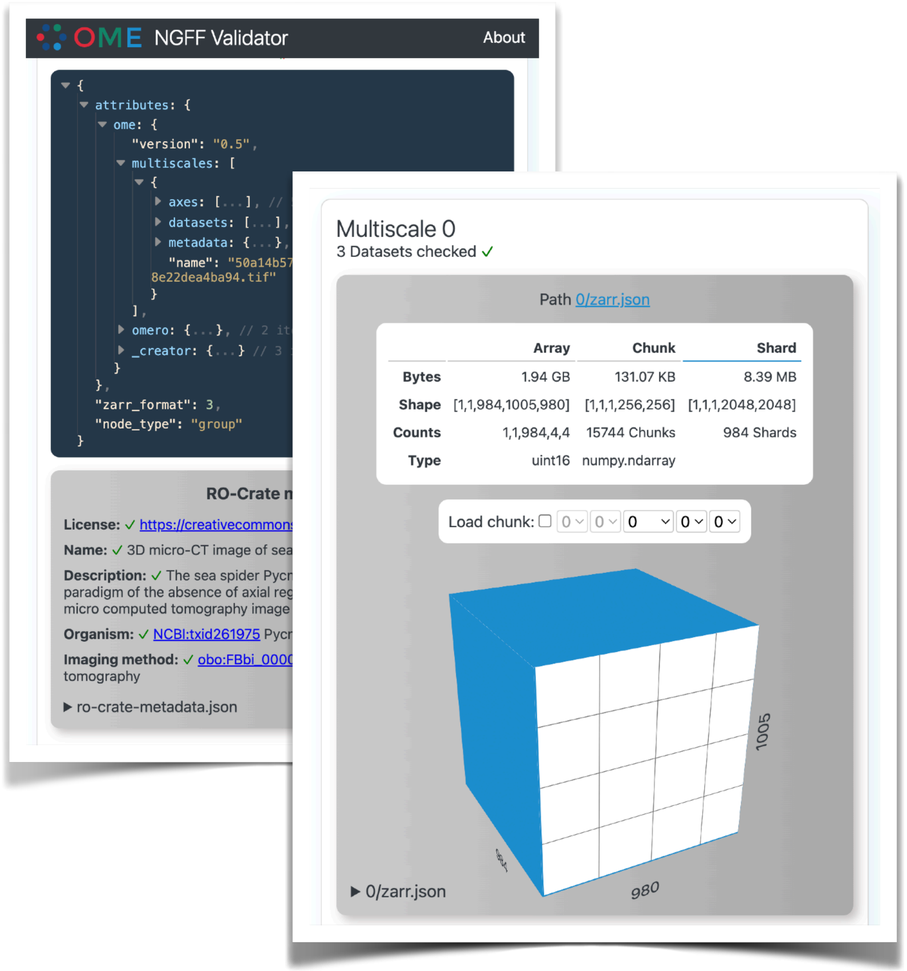
- See Zarr and N5.

Pros:

- natively supports **parallel writes**
- **server-less** access to contents
- much more **web-like** access
- can avoid costly downloads, e.g. when on S3

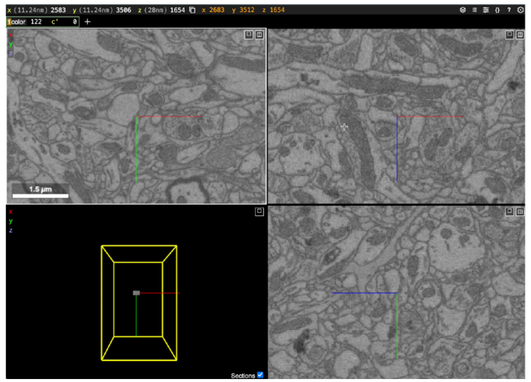
Cons:

- many more files on disk
- no easy way to lock on update
- reduced desktop support

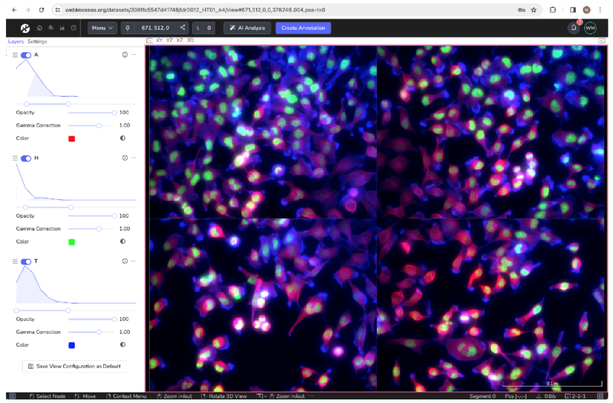


https://bit.ly/40hnF50

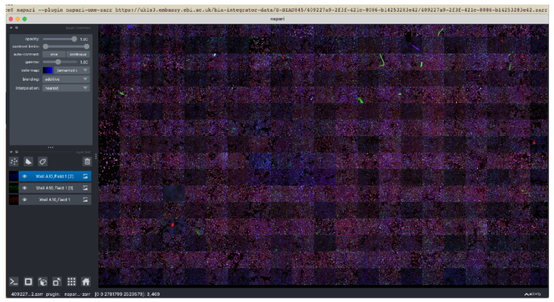
“Standardized” access



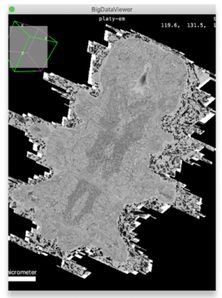
Neuroglancer
https://bit.ly/4hh9P8A



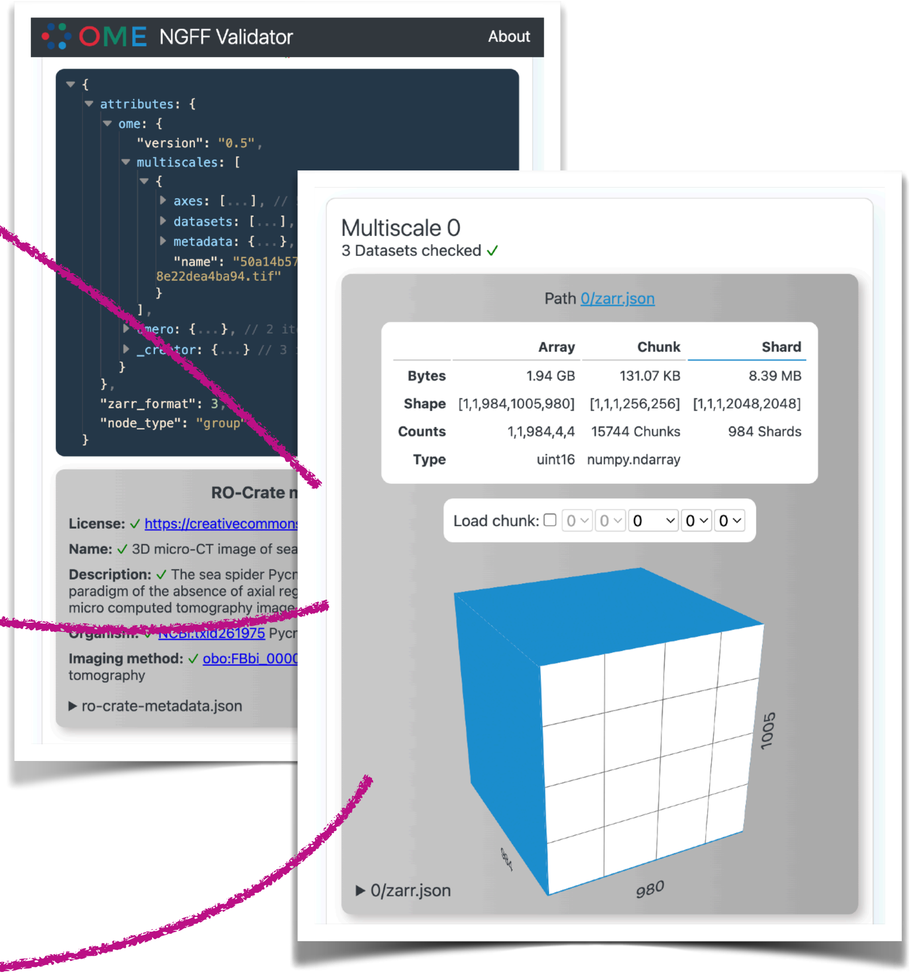
Webknossos
https://wklink.org/6422



Napari

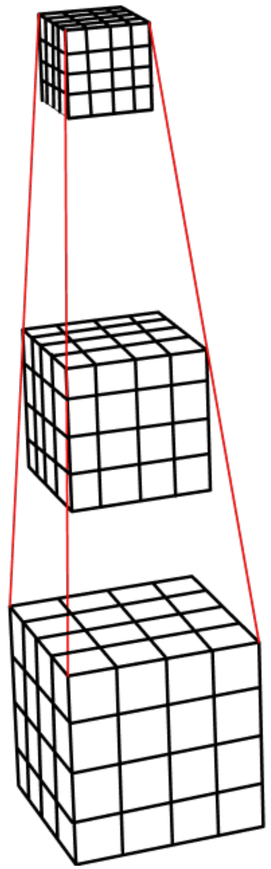


Fiji/BDV/MoBIE



https://bit.ly/40hnF50

Becoming OME-Zarr



```
example/
├── s0    # Full-sized
├── s1                          ...
└── s2    # Smallest
```

```
example/
├── 0    # Full-sized
├── 1
└── 2    # Smallest
```

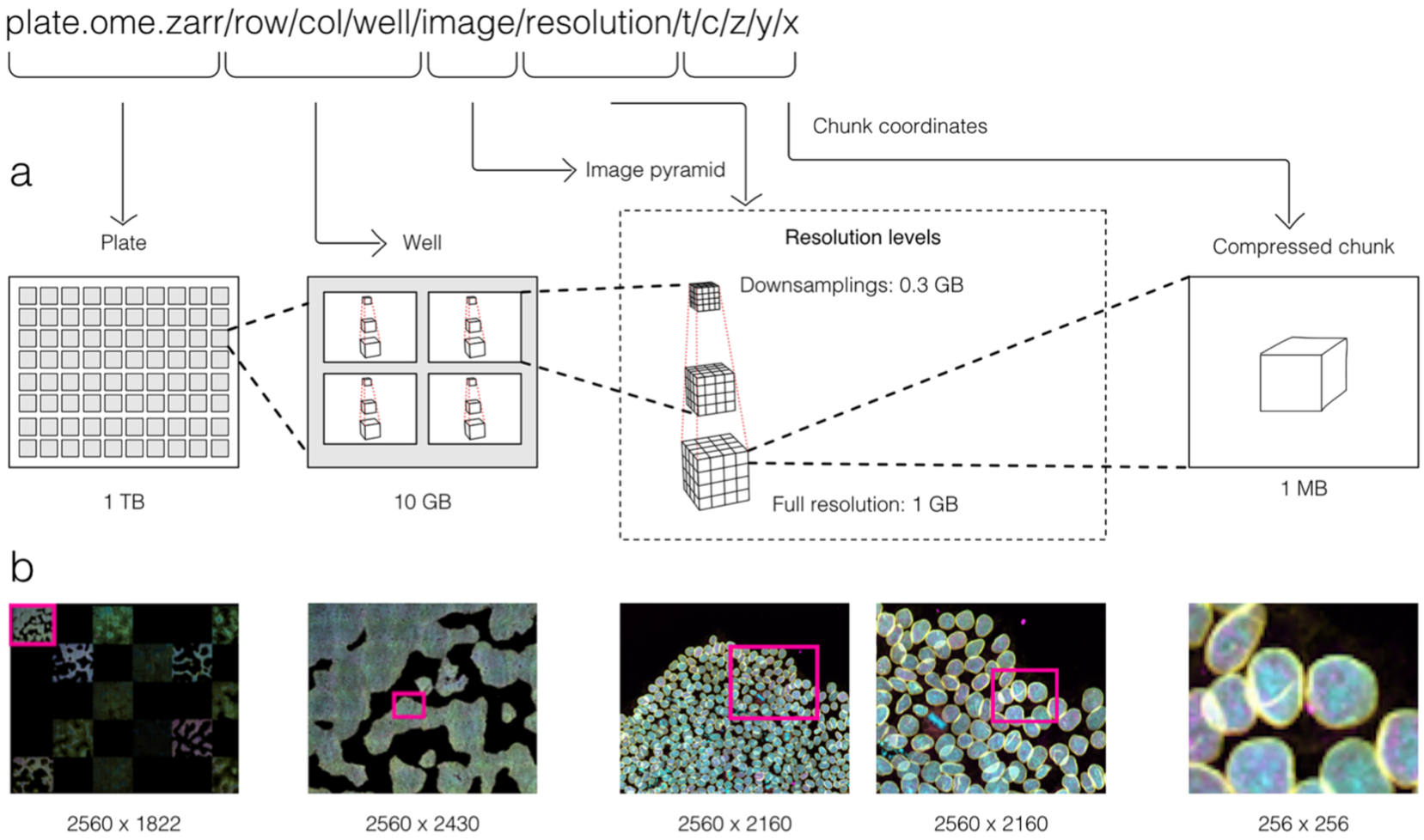
```
example/
├── 0    # Smallest
├── 1
└── 2    # Full-sized
```

```
{ “multiscales”: [
    { “datasets” : [
        {"path": “0"}, {"path": “1"}, {"path": "2"}
```

https://github.com/zarr-developers/zarr-specs/issues/50

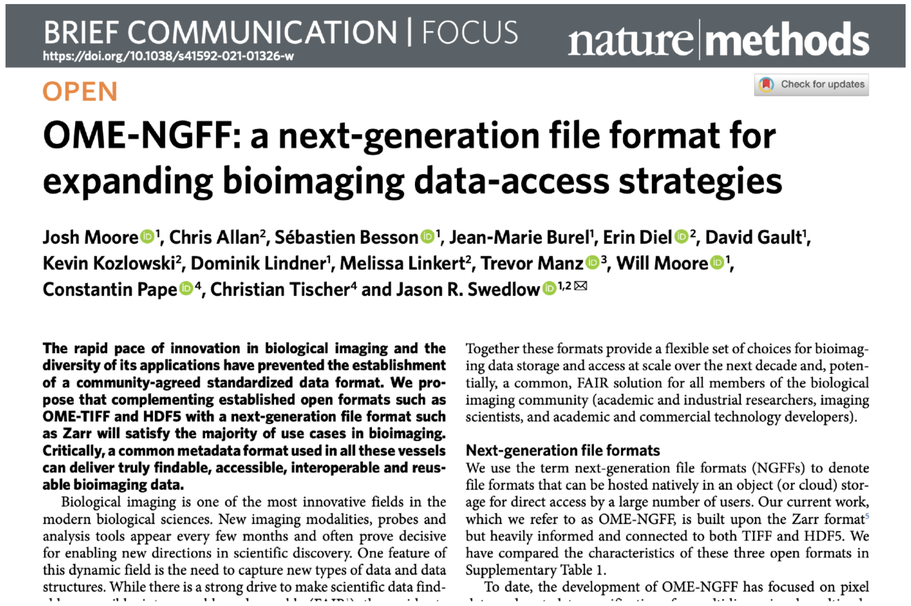
https://ngff.openmicroscopy.org/0.4/#multiscale-md

Multi-resolution, 5D+ image data



Moore et. al: Histochemistry and Cell Biology (2023)

For more information…



BRIEF COMMUNICATION | FOCUS

https://doi.org/10.1038/s41592-021-01326-w

nature methods

OPEN

Check for updates

OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies

Josh Moore[1], Chris Allan[2], Sébastien Besson[1], Jean-Marie Burel[1], Erin Diel[2], David Gault[1], Kevin Kozlowski[2], Dominik Lindner[1], Melissa Linkert[2], Trevor Manz[3], Will Moore[1], Constantin Pape[4], Christian Tischer[4] and Jason R. Swedlow[1,2] ✉

The rapid pace of innovation in biological imaging and the diversity of its applications have prevented the establishment of a community-agreed standardized data format. We propose that complementing established open formats such as OME-TIFF and HDF5 with a next-generation file format such as Zarr will satisfy the majority of use cases in bioimaging. Critically, a common metadata format used in all these vessels can deliver truly findable, accessible, interoperable and reusable bioimaging data.

Biological imaging is one of the most innovative fields in the modern biological sciences. New imaging modalities, probes and analysis tools appear every few months and often prove decisive for enabling new directions in scientific discovery. One feature of this dynamic field is the need to capture new types of data and data structures. While there is a strong drive to make scientific data find-

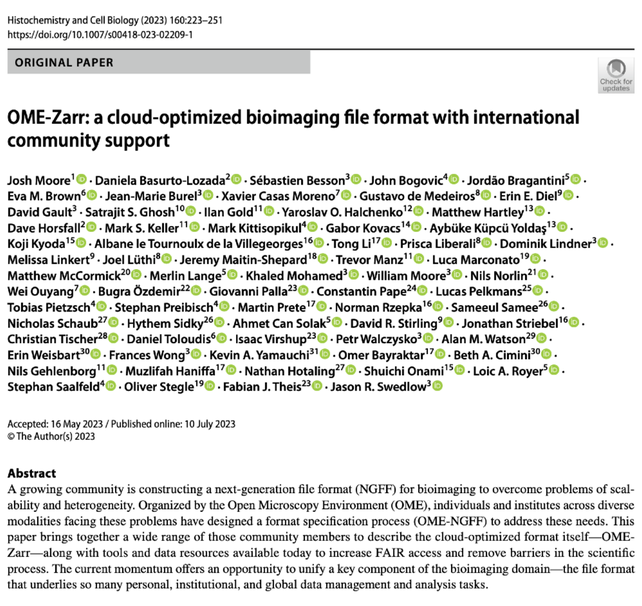
Together these formats provide a flexible set of choices for bioimaging data storage and access at scale over the next decade and, potentially, a common, FAIR solution for all members of the biological imaging community (academic and industrial researchers, imaging scientists, and academic and commercial technology developers).

Next-generation file formats

We use the term next-generation file formats (NGFFs) to denote file formats that can be hosted natively in an object (or cloud) storage for direct access by a large number of users. Our current work, which we refer to as OME-NGFF, is built upon the Zarr format[5] but heavily informed and connected to both TIFF and HDF5. We have compared the characteristics of these three open formats in Supplementary Table 1.

To date, the development of OME-NGFF has focused on pixel

https://doi.org/10.1038/s41592-021-01326-w



Histochemistry and Cell Biology (2023) 160:223–251
https://doi.org/10.1007/s00418-023-02209-1

ORIGINAL PAPER

Check for updates

OME-Zarr: a cloud-optimized bioimaging file format with international community support

Josh Moore[1] · Daniela Basurto-Lozada[2] · Sébastien Besson[3] · John Bogovic[4] · Jordão Bragantini[5] · Eva M. Brown[6] · Jean-Marie Burel[3] · Xavier Casas Moreno[7] · Gustavo de Medeiros[8] · Erin E. Diel[9] · David Gault[3] · Satrajit S. Ghosh[10] · Ilan Gold[11] · Yaroslav O. Halchenko[12] · Matthew Hartley[13] · Dave Horsfall[2] · Mark S. Keller[11] · Mark Kittisopikul[4] · Gabor Kovacs[14] · Aybüke Küpcü Yoldaş[13] · Koji Kyoda[15] · Albane le Tournoulx de la Villegeorges[16] · Tong Li[17] · Prisca Liberali[8] · Dominik Lindner[3] · Melissa Linkert[9] · Joel Lüthi[8] · Jeremy Maitin-Shepard[18] · Trevor Manz[11] · Luca Marconato[19] · Matthew McCormick[20] · Merlin Lange[5] · Khaled Mohamed[3] · William Moore[3] · Nils Norlin[21] · Wei Ouyang[7] · Bugra Özdemir[22] · Giovanni Palla[23] · Constantin Pape[24] · Lucas Pelkmans[25] · Tobias Pietzsch[4] · Stephan Preibisch[4] · Martin Prete[17] · Norman Rzepka[16] · Sameeul Samee[26] · Nicholas Schaub[27] · Hythem Sidky[26] · Ahmet Can Solak[5] · David R. Stirling[9] · Jonathan Striebel[16] · Christian Tischer[28] · Daniel Toloudis[6] · Isaac Virshup[23] · Petr Walczysko[3] · Alan M. Watson[29] · Erin Weisbart[30] · Frances Wong[3] · Kevin A. Yamauchi[31] · Omer Bayraktar[17] · Beth A. Cimini[26] · Nils Gehlenborg[11] · Muzlifah Haniffa[17] · Nathan Hotaling[27] · Shuichi Onami[15] · Loic A. Royer[5] · Stephan Saalfeld[4] · Oliver Stegle[19] · Fabian J. Theis[23] · Jason R. Swedlow[3]

Accepted: 16 May 2023 / Published online: 10 July 2023
© The Author(s) 2023

Abstract
A growing community is constructing a next-generation file format (NGFF) for bioimaging to overcome problems of scalability and heterogeneity. Organized by the Open Microscopy Environment (OME), individuals and institutes across diverse modalities facing these problems have designed a format specification process (OME-NGFF) to address these needs. This paper brings together a wide range of those community members to describe the cloud-optimized format itself—OME-Zarr—along with tools and data resources available today to increase FAIR access and remove barriers in the scientific process. The current momentum offers an opportunity to unify a key component of the bioimaging domain—the file format that underlies so many personal, institutional, and global data management and analysis tasks.

https://doi.org/10.1007/s00418-023-02209-1

Next?

Illustration by Henning Falk,
licensed under CC-BY 4.0

NFDI 4 BIOIMAGE

National Research Data Initiative (2023-2028; *Germany*)

DFG Deutsche Forschungsgemeinschaft German Research Foundation

Realizing the FAIR principles

even for the largest images

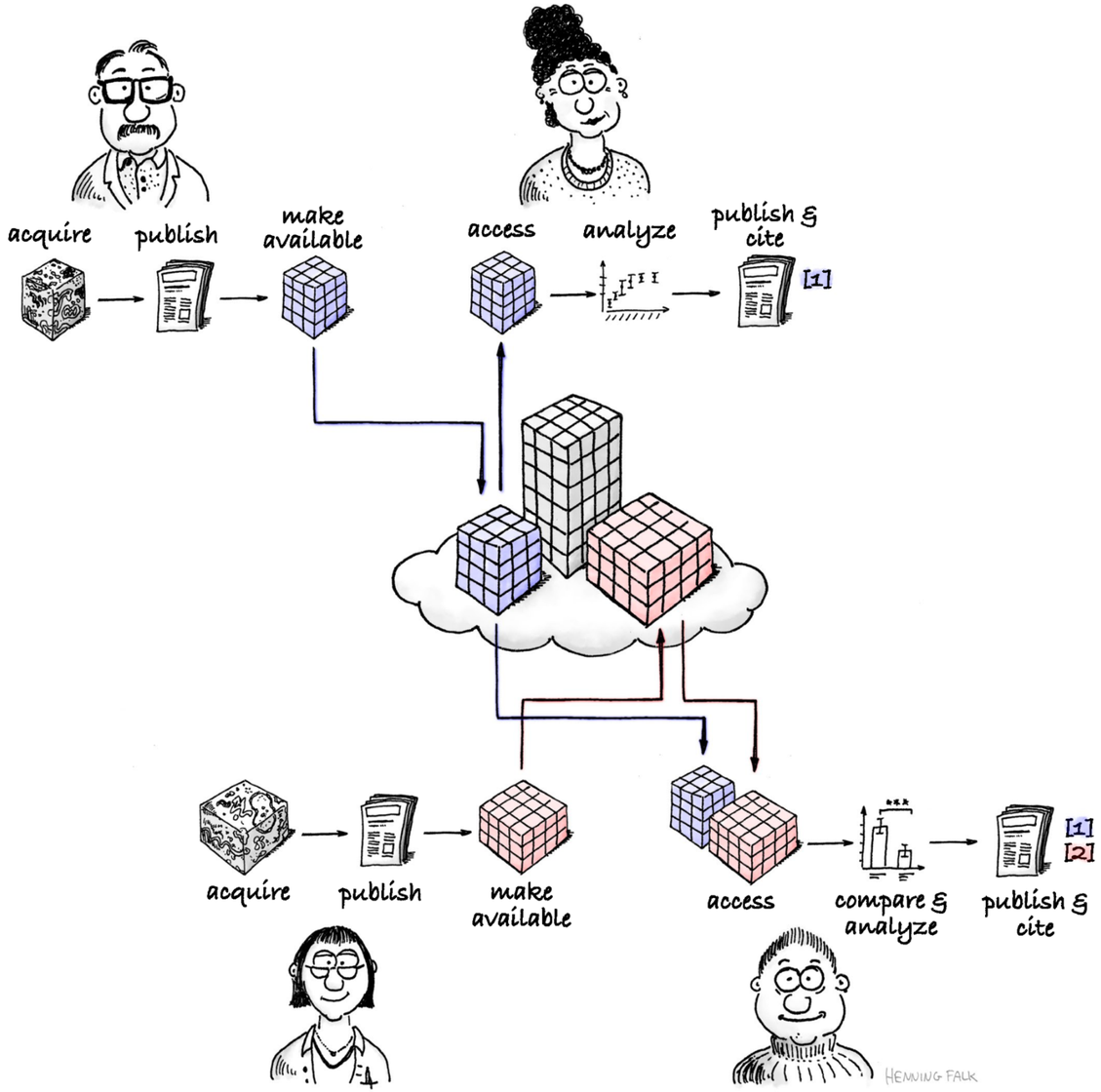
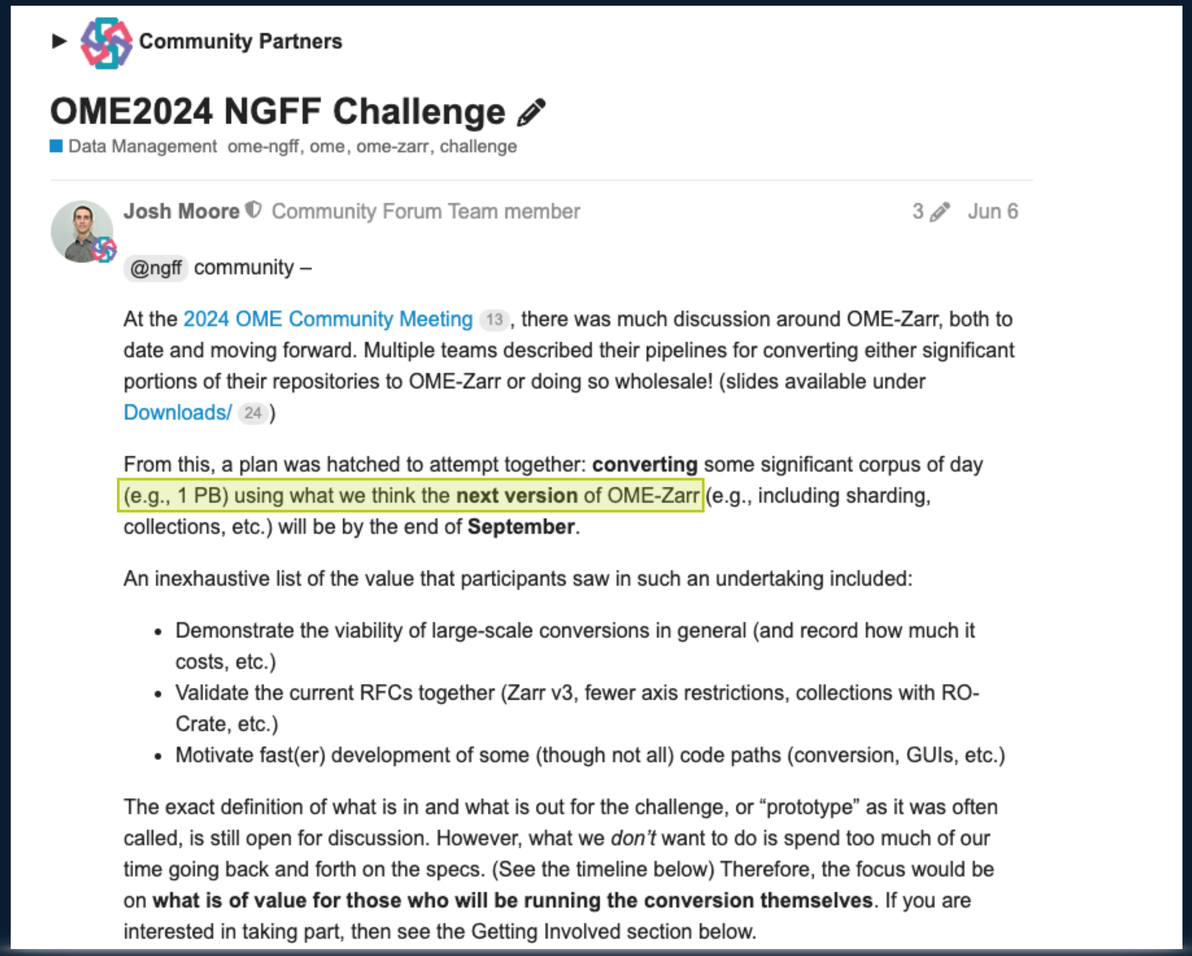


Image “FAIR re-use” by Henning Falk ©2022 NumFOCUS (CC BY 4.0 license)
https://github.com/zarr-developers/zarr-illustrations-falk-2022

ome2024-ngff-challenge



▶ Community Partners

OME2024 NGFF Challenge

Data Management ome-ngff, ome, ome-zarr, challenge

Josh Moore Community Forum Team member 3 Jun 6

@ngff community –

At the 2024 OME Community Meeting 13, there was much discussion around OME-Zarr, both to date and moving forward. Multiple teams described their pipelines for converting either significant portions of their repositories to OME-Zarr or doing so wholesale! (slides available under Downloads/ 24)

From this, a plan was hatched to attempt together: **converting** some significant corpus of day (e.g., 1 PB) using what we think the **next version** of OME-Zarr (e.g., including sharding, collections, etc.) will be by the end of **September**.

An inexhaustive list of the value that participants saw in such an undertaking included:

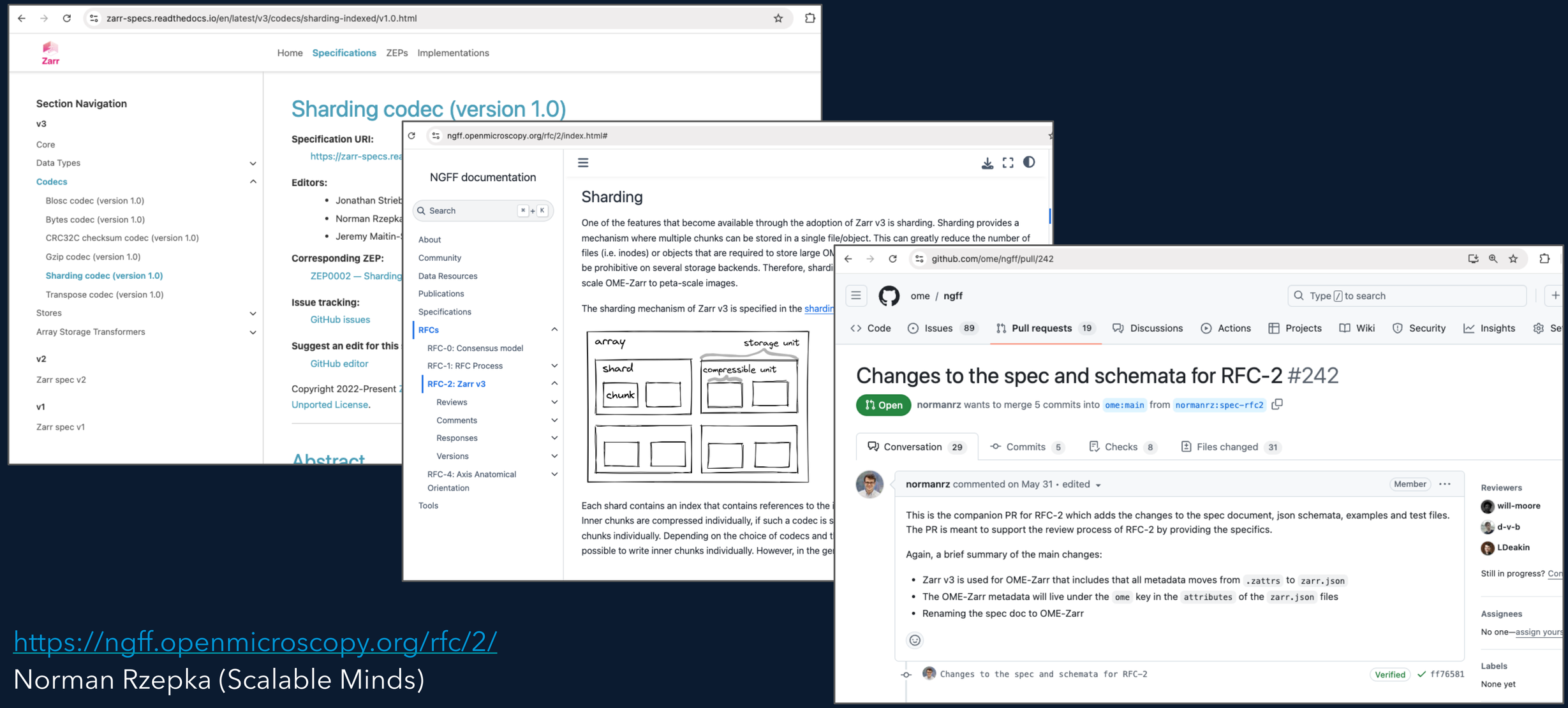
- Demonstrate the viability of large-scale conversions in general (and record how much it costs, etc.)
- Validate the current RFCs together (Zarr v3, fewer axis restrictions, collections with RO-Crate, etc.)
- Motivate fast(er) development of some (though not all) code paths (conversion, GUIs, etc.)

The exact definition of what is in and what is out for the challenge, or "prototype" as it was often called, is still open for discussion. However, what we *don't* want to do is spend too much of our time going back and forth on the specs. (See the timeline below) Therefore, the focus would be on **what is of value for those who will be running the conversion themselves**. If you are interested in taking part, then see the Getting Involved section below.

https://forum.image.sc/t/ome2024-ngff-challenge/97363

- 1PB
- "bring your own bucket"
- proposed next version
- drive development

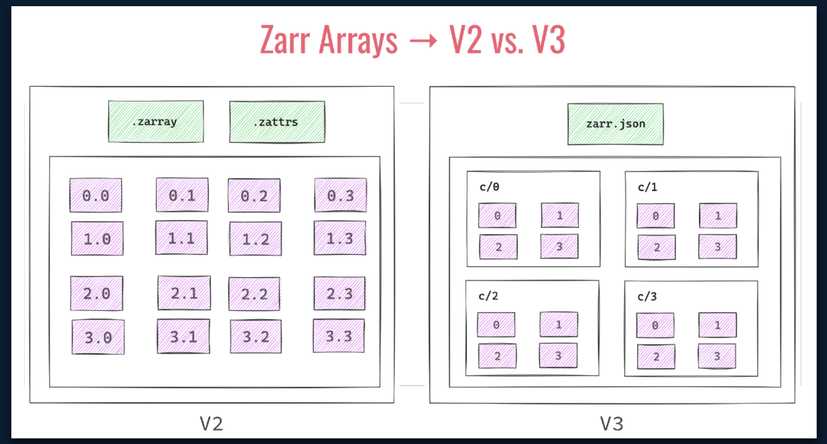
Specification (1/2) Adopting Zarr v3

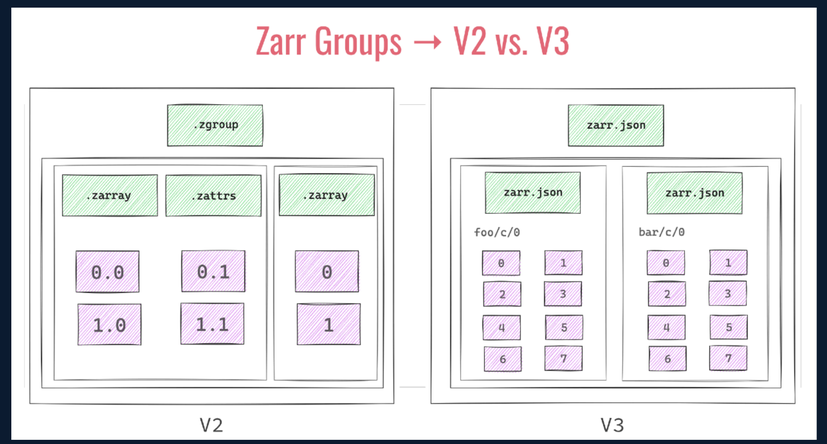


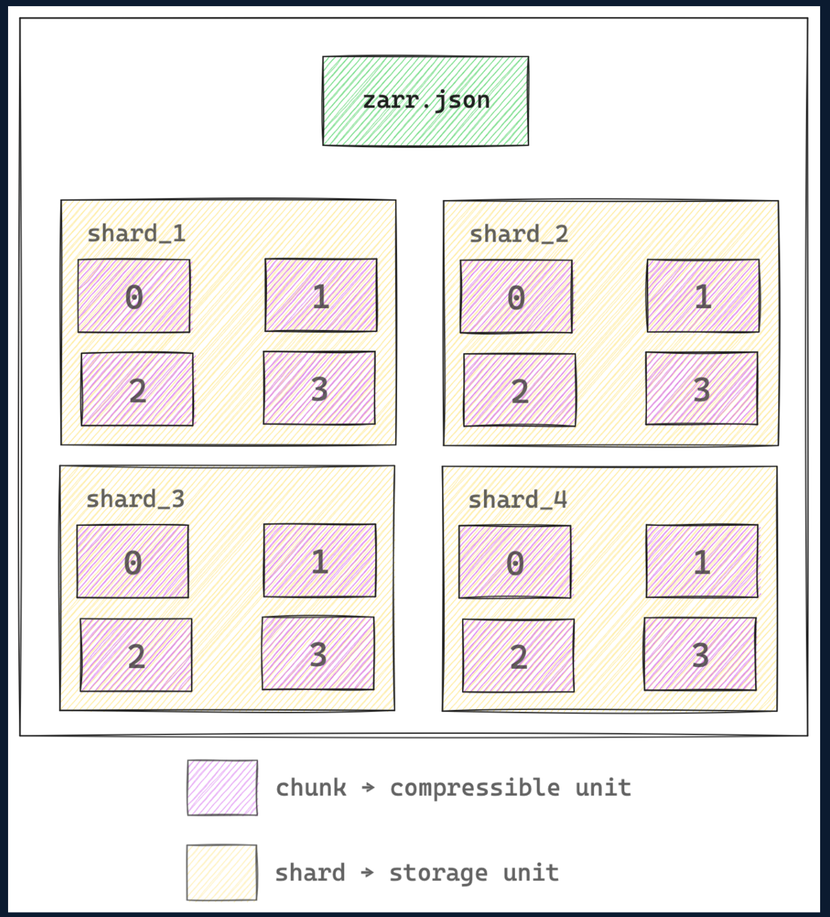
https://ngff.openmicroscopy.org/rfc/2/

Norman Rzepka (Scalable Minds)

Specification (1/2) Adopting Zarr v3







https://bit.ly/explore_zarr_v3

Sanket Verma (Zarr/NumFOCUS)

Specification (2/2) RO-Crate metadata

image.zarr / ro-crate-metadata.json

RO-Crate

```
@type: Dataset
name
description
license
resultOf
```

```
@type: image_acquisition
fbbi_id: {"@id": "obo:FBbi_00000369"}
specimen
```

```
@type: specimen
biosample
```

```
@type: biosample
organism_classification: {"@id": "NCBI:txid10090"}
```

Francois Sherwood (EBI)

🪄Hocus Pocus

```
$ pip install ome2024-ngff-challenge
$ ome2024-ngff-challenge resave input.zarr output.zarr
```

Svelte.js viewer

samples1.csv

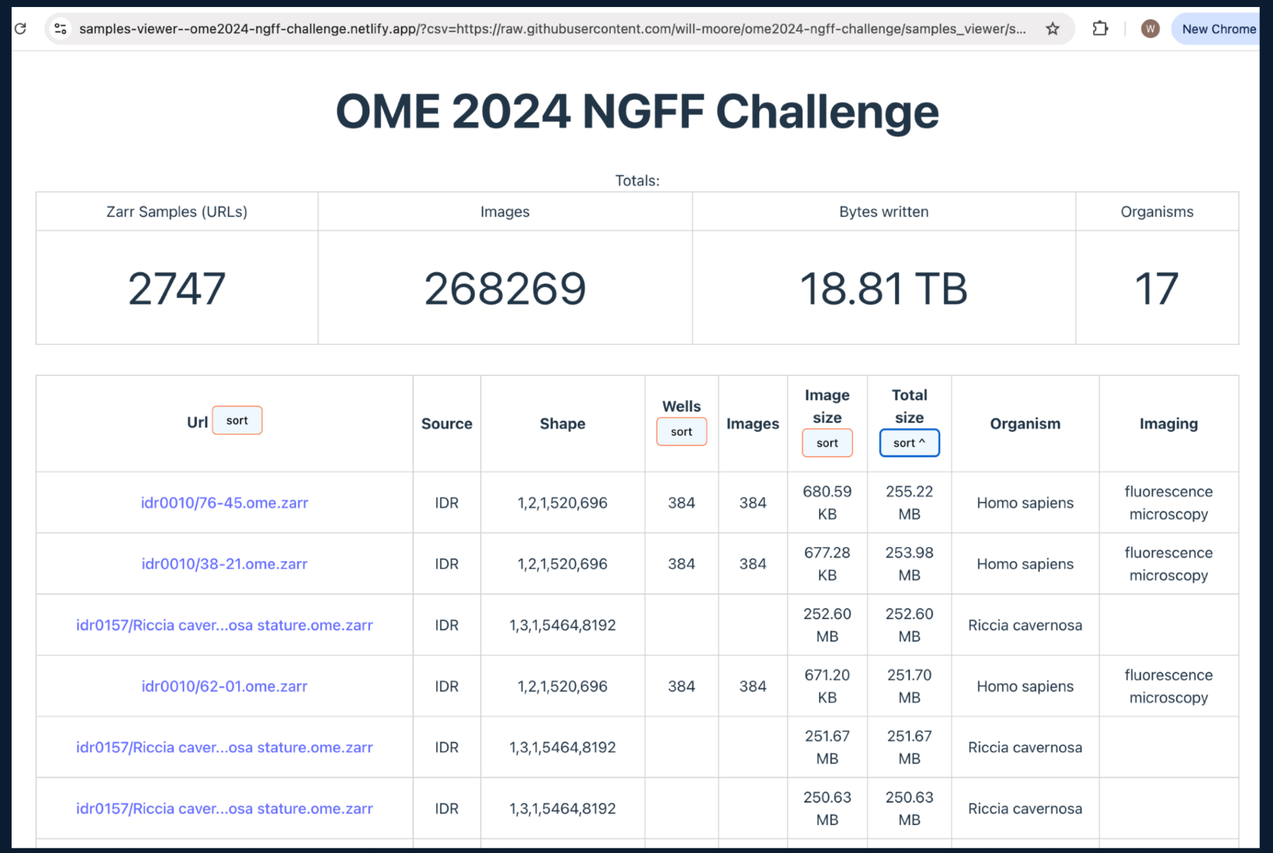
```
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/76-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/65-12.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/34-30.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/124-44.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/47-35.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/155-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/20-29.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/16-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/33-46.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/131-20.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/74-26.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/14-35.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/108-24.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/48-29.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/63-27.ome.zarr
```

samples2.csv

```
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/76-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/65-12.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/34-30.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/124-44.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/47-35.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/155-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/20-29.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/16-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/33-46.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/131-20.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/74-26.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/14-35.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/108-24.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/48-29.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/63-27.ome.zarr
```

samples3.csv

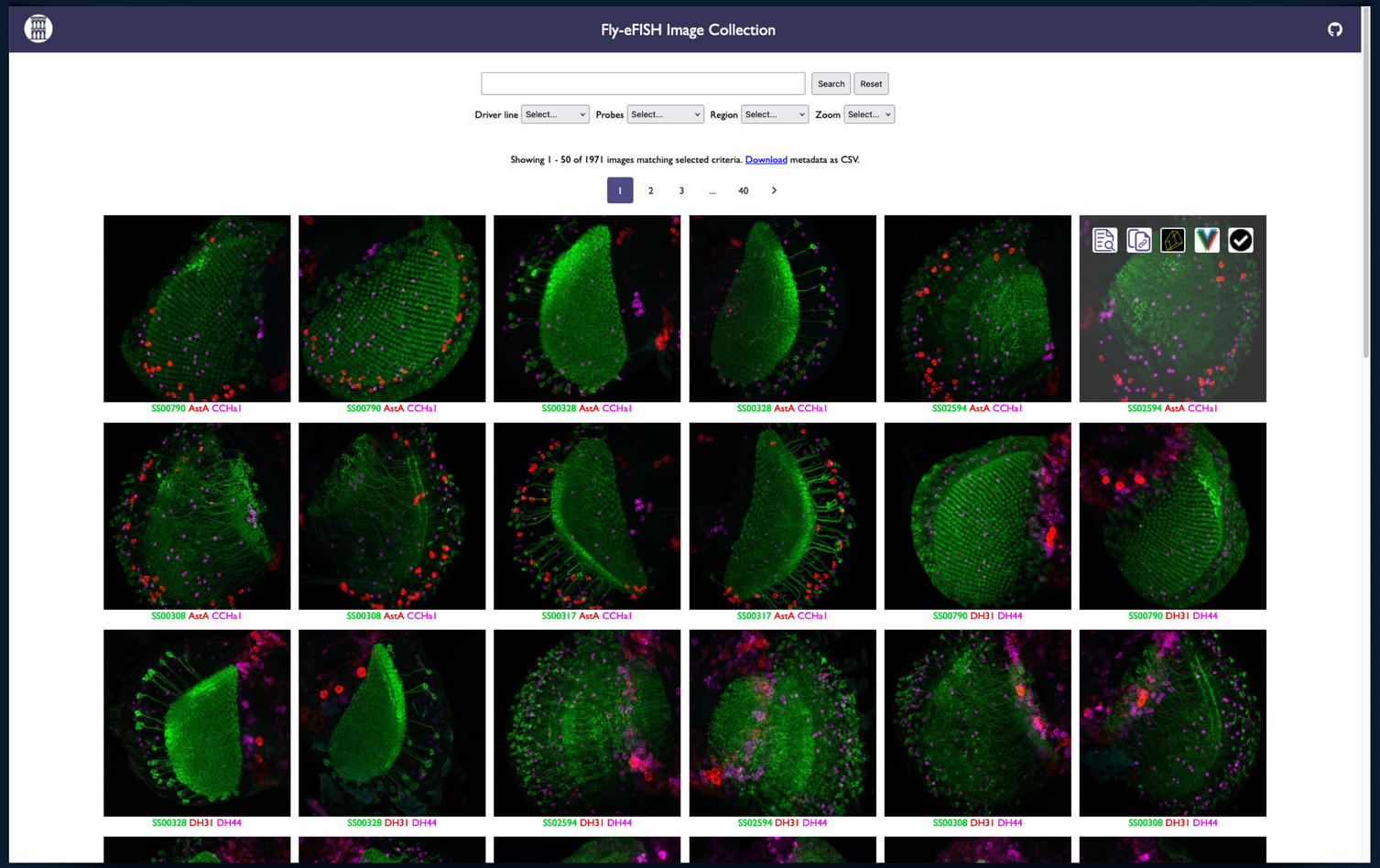
```
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/76-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/65-12.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/34-30.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/124-44.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/47-35.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/155-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/20-29.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/16-45.ome.zarr
https://uk1s3.embassy.ebi.ac.uk/idr//idr0010/33-46.ome.zarr
```



https://ome.github.io/ome2024-ngff-challenge/

Will Moore (University of Dundee)

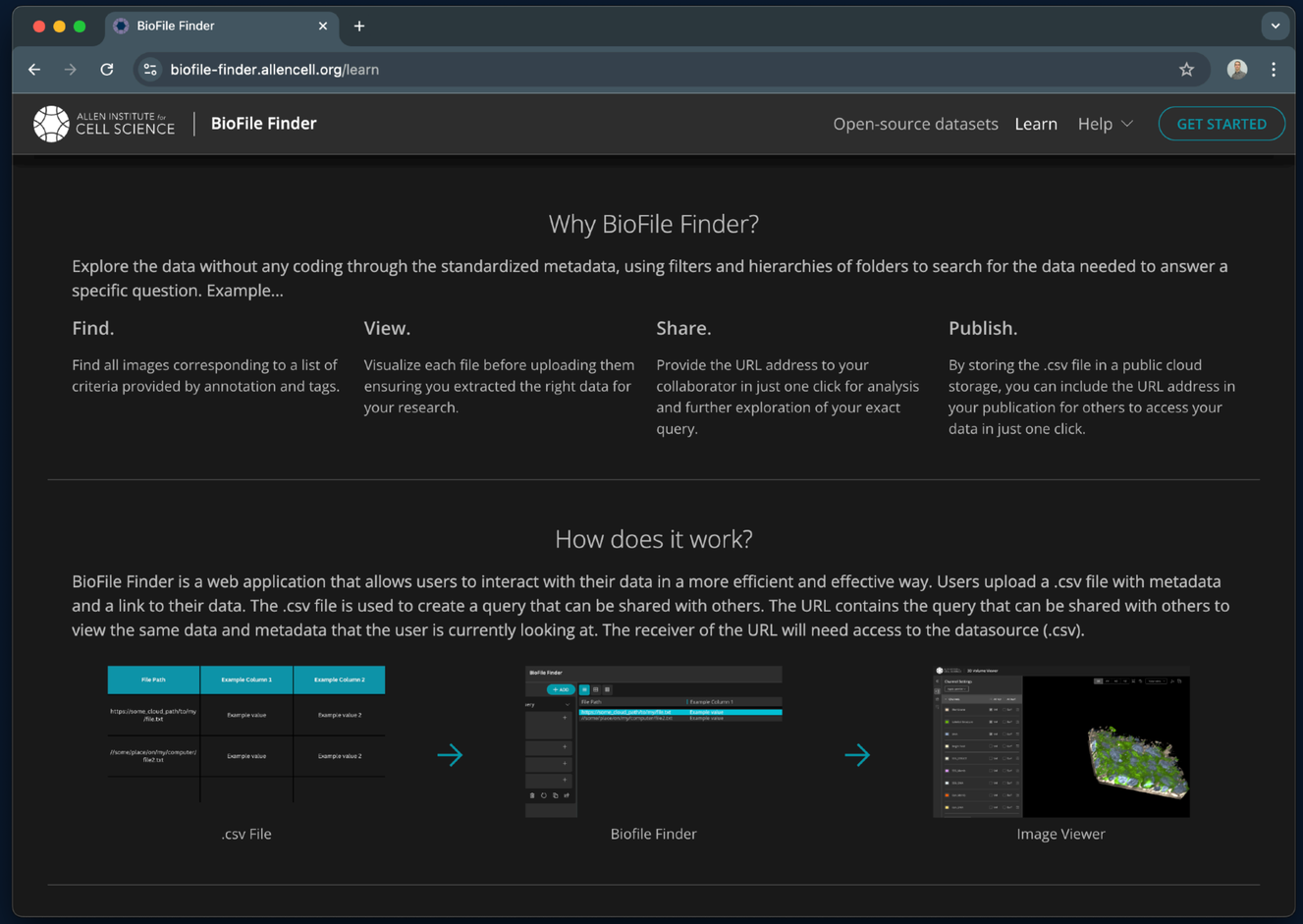
and **another**



https://github.com/JaneliaSciComp/zarrcade

Konrad Rokicki (Janelia)

and yet **another**



https://biofile-finder.allencell.org/

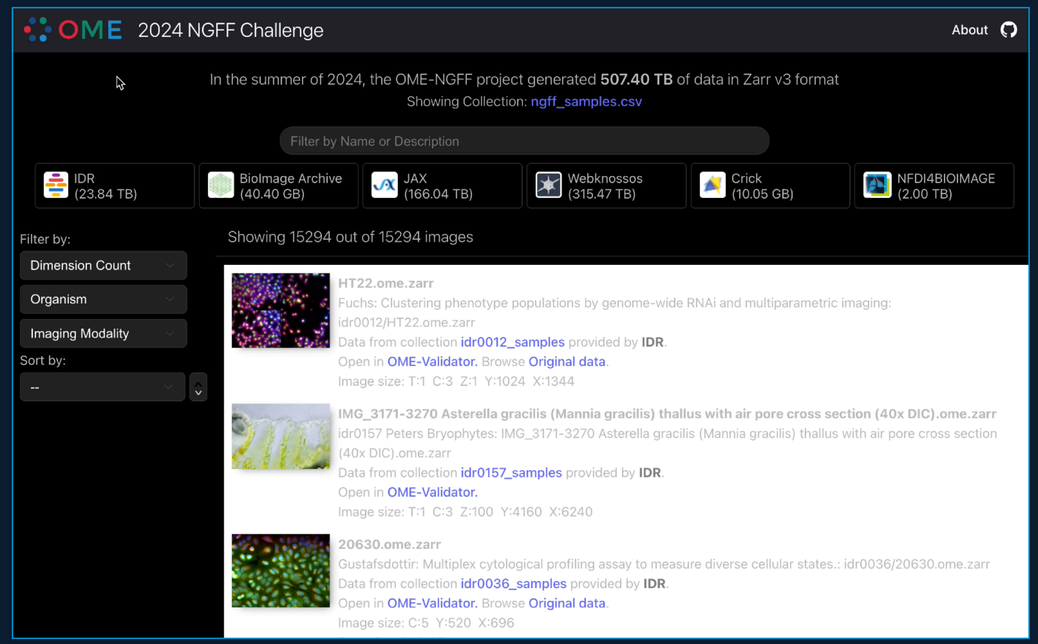
Allen Institute of Cell Science

ome2024-ngff-challenge

- 4 months
- ~9 conf. calls
- several PRs
- few Ks ($£¥€)
- 50% of goal

"most exciting thing about this challenge was big-picture vision of federated image repositories that speak common language"

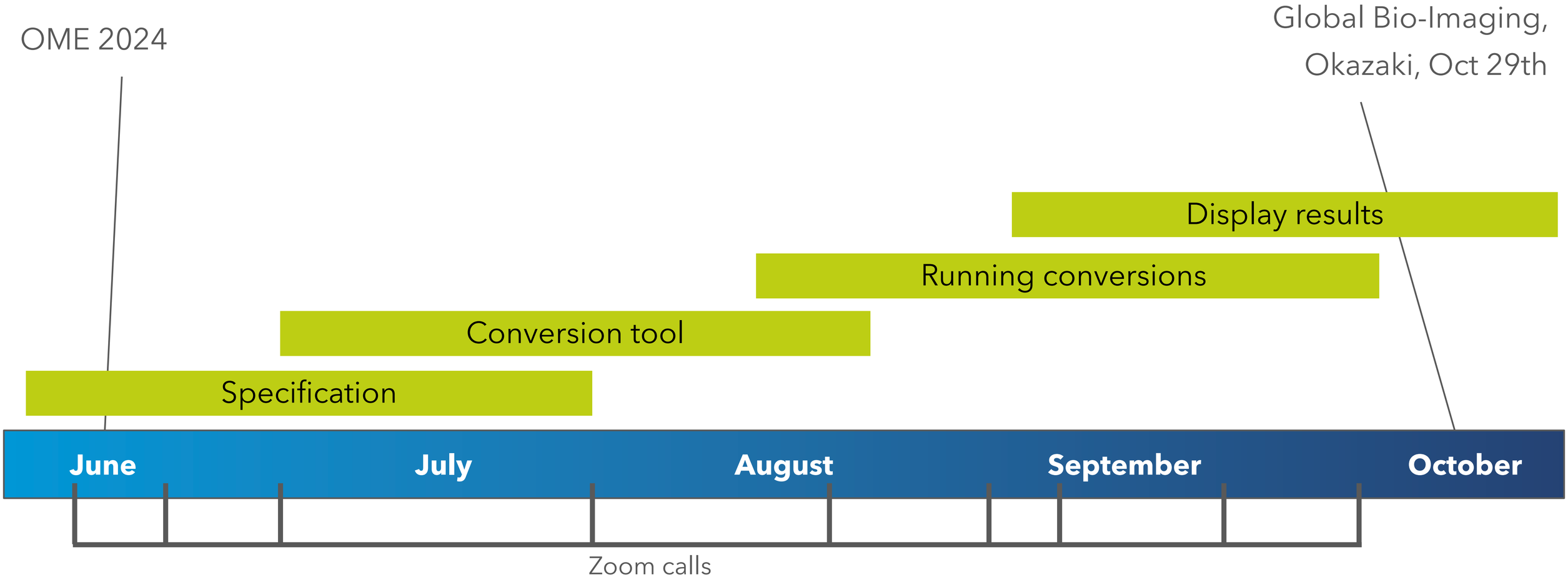
ome2024-ngff-challenge



https://ome.github.io/ome2024-ngff-challenge/

"most exciting thing about this challenge was big-picture vision of federated image repositories that speak common language"

Challenge Timeline

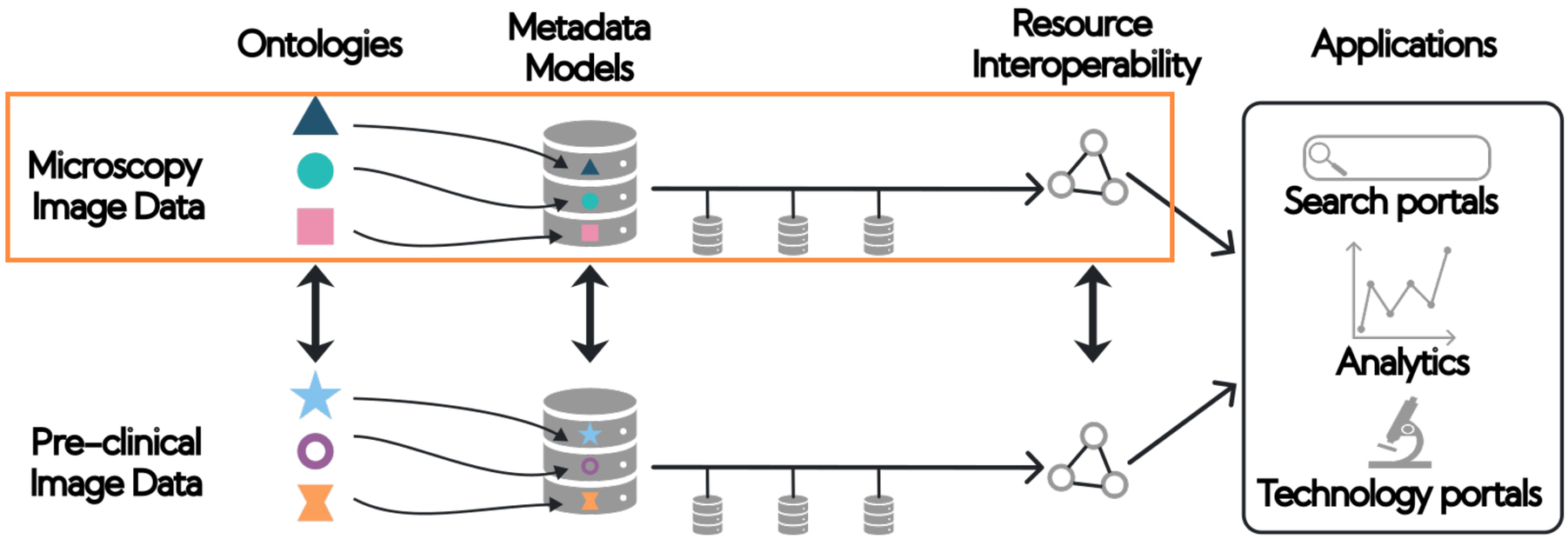


foundingGIDE

Global Image Data Ecosystem (2024-2026)

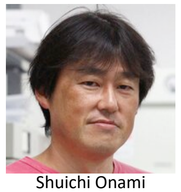
Funded by
the European Union

nfdi Nationale Forschungsdaten Infrastruktur



foundingGIDE is a Horizon Europe-funded project coordinated by **Euro-BioImaging ERIC** that will build strong foundations of a **Global Image Data Ecosystem** (GIDE) for image data exchange based on global coordination of technical developments among infrastructures and communities.

Matthew Hartley

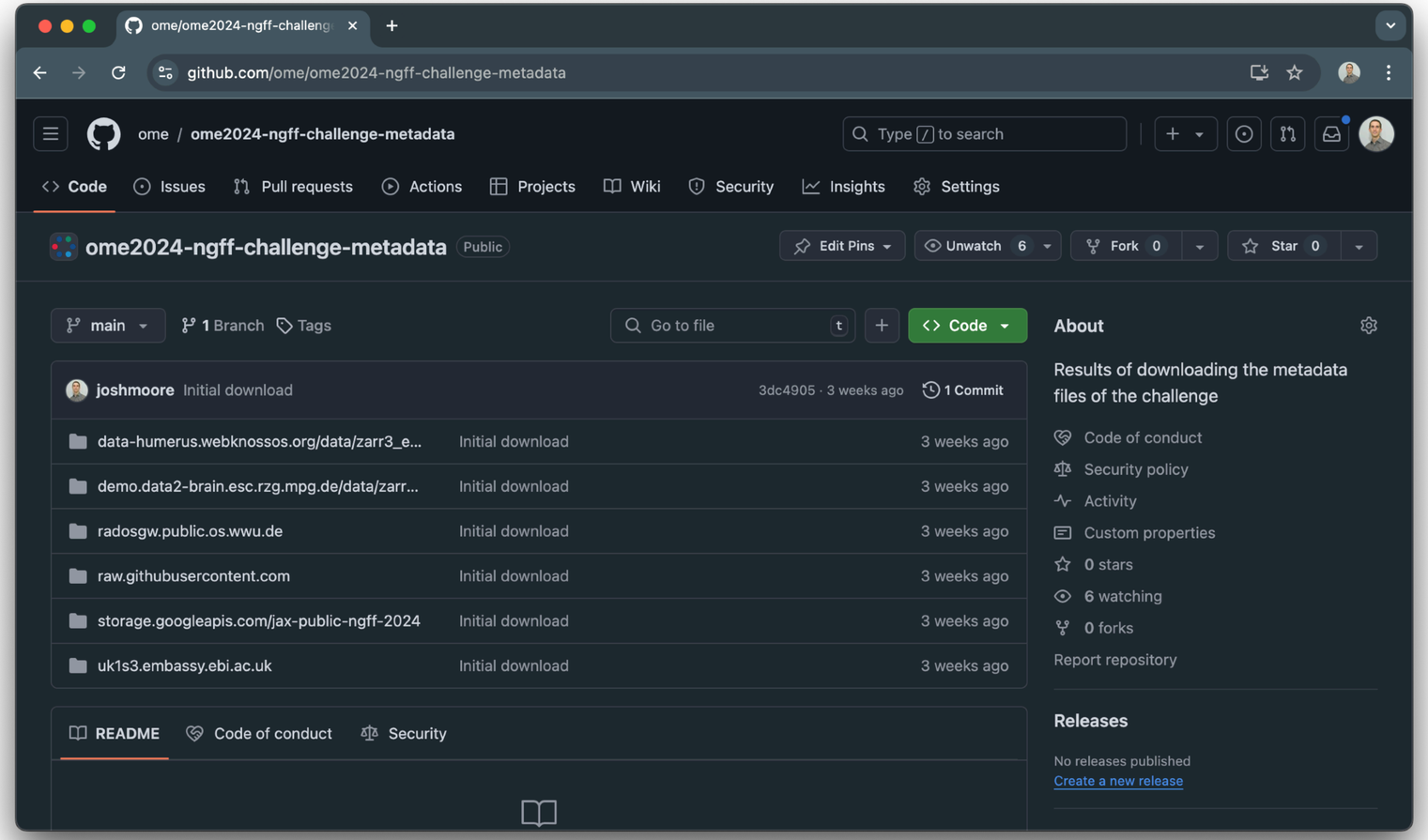

Shuichi Onami

foundingGIDE

Global Image Data Ecosystem (2024-2026)

Funded by
the European Union

NFDI4 BIOIMAGE



Challenge metadata

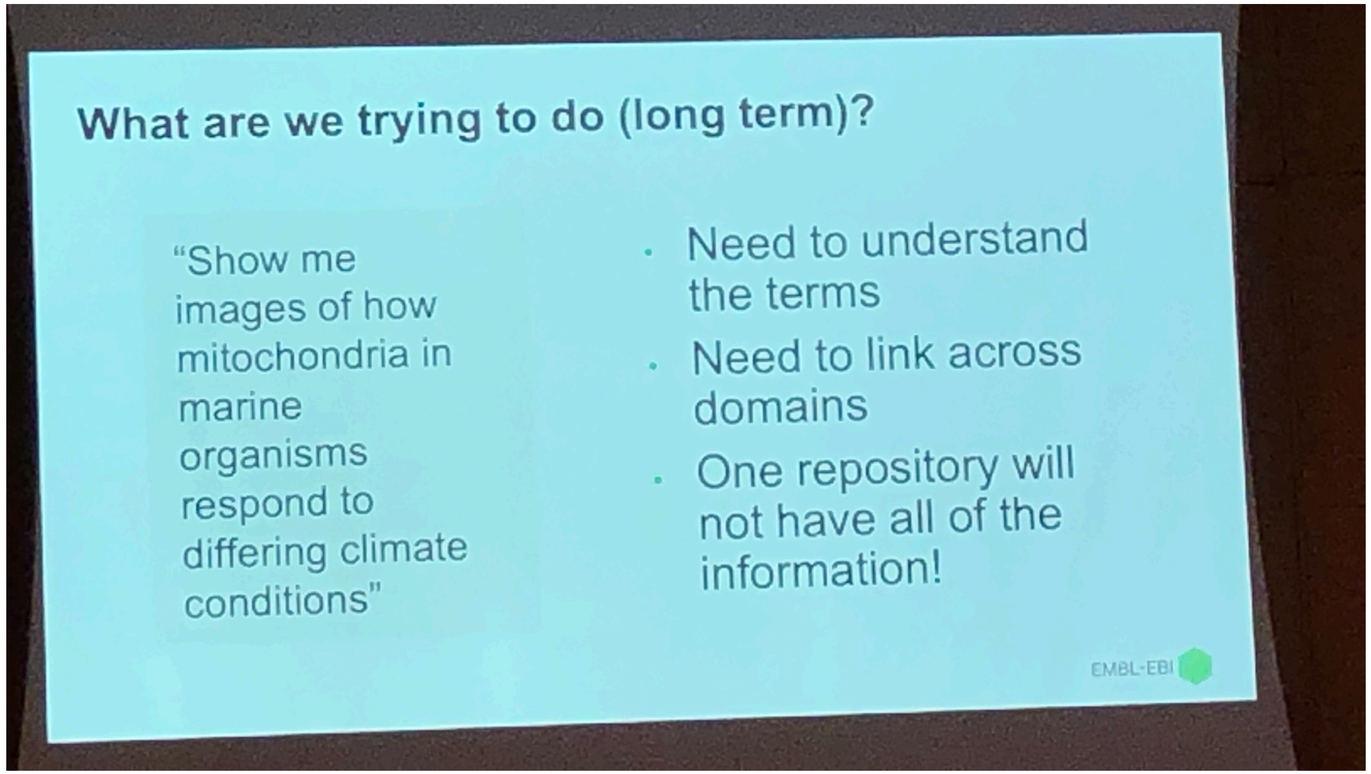
- 10K files
- 100 MBs
- Missing newer files etc.
- “Hackathon ready”

https://github.com/ome/ome2024-ngff-challenge-metadata

foundingGIDE

Global Image Data Ecosystem (2024-2026)

Funded by
the European Union



creative commons
10.5281/zenodo.14234608

Choose your own adventure (**20min**)

^metadata

Metadata Matters 2010

JCB: Comment

Metadata matters: access to image data in the real world

Melissa Linkert,[1,2,4] Curtis T. Rueden,[1,2] Chris Allan,[3,4] Jean-Marie Burel,[3] Will Moore,[3] Andrew Patterson,[3] Brian Loranger,[3] Josh Moore,[4] Carlos Neves,[4] Donald MacDonald,[3] Aleksandra Tarkowska,[3] Caitlin Sticco,[1,2] Emma Hill,[5] Mike Rossner,[5] Kevin W. Eliceiri,[1,2] and Jason R. Swedlow[3,4]

[1]Laboratory for Optical and Computational Instrumentation, Department of Molecular Biology and [2]Department of Biomedical Engineering, Graduate School, University of Wisconsin at Madison, Madison, WI 53711
[3]Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland, UK
[4]Glencoe Software, Inc., Seattle, WA 98101
[5]The Rockefeller University Press, New York, NY 10065

Data sharing is important in the biological sciences to prevent duplication of effort, to promote scientific integrity, and to facilitate and disseminate scientific discovery. Sharing requires centralized repositories, and submission to and utility of these resources require common data formats. This is particularly challenging for multidimensional microscopy image data, which are acquired from a variety of platforms with a myriad of proprietary file formats (PFFs). In this paper, we describe an open standard format that we have developed for microscopy image data. We call on the community to use open image data standards and to insist that all imaging platforms support these file formats. This will build the foundation for an open image data repository.

Recent letters and editorials have highlighted the importance of open access to the large datasets now being collected by biologists in laboratories around the world (COSEPUP, 2009; Field et al., 2009; Schofield et al., 2009). Researchers, universities, and funding bodies all agree that scientific data produced from public- and charity-funded research (not just the results, but complete workflows including raw data) should be shared and accessible. The arguments in favor of open access data are now well established, and protocols and principles for data sharing are emerging (http://sciencecommons.org/projects/publishing/open-access-data-protocol). However, access to and sharing of scientific data require substantial effort and investment to define specifications and build resources to support them. For the successful sharing of DNA sequence data, the genome communities built, maintained, and in some cases fought for the standards and resources that were ultimately accepted by the whole community. This effort laid the foundation for the release of genomic data and the development of online resources, accessible by anyone, for any purpose, that now underpin all modern biomedical research.

Correspondence to Kevin W. Eliceiri: eliceiri@wisc.edu; or Jason R. Swedlow: jason@lifesci.dundee.ac.uk

We believe the imaging community can achieve the same success for digital image data. In this paper, we review the current status of online biological image repositories and provide a set of recommendations to drive the use of open standardized data formats in biological microscopy as a prerequisite for creating a global image data repository.

Scientific image data repositories for the life sciences

In December 2008, the *Journal of Cell Biology* (JCB) launched the *JCB DataViewer*, an online repository for original image data in the life sciences (Fig. 1). To our knowledge, this system is the first open repository that enables routine archiving and sharing of original image datasets supporting published scientific articles. One key attribute of the *JCB DataViewer* that distinguishes it from past and current data repositories is that the original binary data and metadata, additional information captured by acquisition software about an image, such as the instruments used, acquisition settings, image size, and resolution, are preserved and accessible by the community. As of this writing, the *JCB DataViewer* contains 6,446 multidimensional (5D; including space, channel, and time) images in support of 186 published articles. The *JCB DataViewer* is a customized application based on the open source and open development Open Microscopy Environment (OME) Remote Objects (OMERO) and Bio-Formats projects, released by the OME Consortium (http://openmicroscopy.org).

One goal of the *JCB DataViewer* was to initiate the development of a functional, scientifically valuable online image repository. The first step was to make original data available alongside a publication, available for examination by reviewers and readers of a submitted or published manuscript. Currently, the *JCB DataViewer* allows access to original data for viewing, simple measurement, and review, but users cannot download

© 2010 Linkert et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see http://www.rupress.org/terms). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 3.0 Unported license, as described at http://creativecommons.org/licenses/by-nc-sa/3.0/).

The Rockefeller University Press $30.00
J. Cell Biol. Vol. 189 No. 5 777–782
www.jcb.org/cgi/doi/10.1083/jcb.201004104

Downloaded from http://rupress.org/jcb/article-pdf/189/5/777/1471712/jcb_201004104.pdf by guest on 22 May 2024

THE JOURNAL OF CELL BIOLOGY

https://doi.org/10.1083/jcb.201004104
(Most cited OME paper)

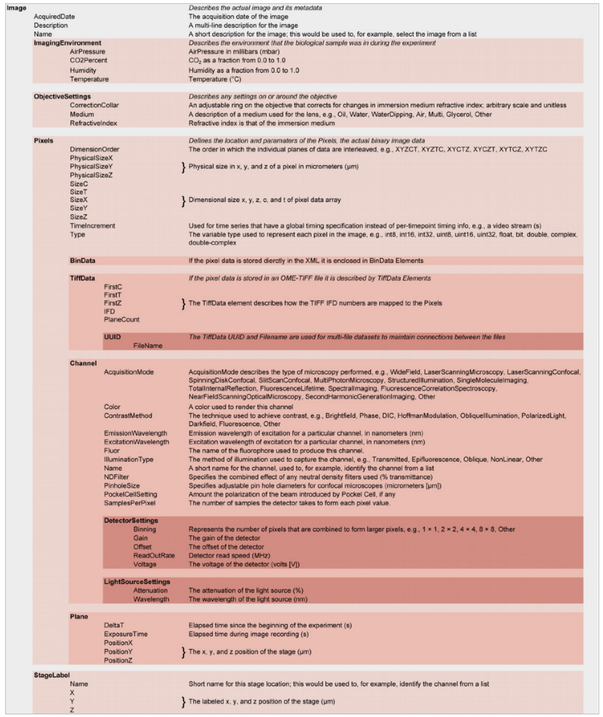


Image			*Describes the actual image and its metadata*
	AcquiredDate		The acquisition date of the image
	Description		A multi-line description for the image
	Name		A short description for the image; this would be used to, for example, select the image from a list
	ImagingEnvironment		*Describes the environment that the biological sample was in during the experiment*
		AirPressure	AirPressure in millibars (mbar)
		CO2Percent	CO_2 as a fraction from 0.0 to 1.0
		Humidity	Humidity as a fraction from 0.0 to 1.0
		Temperature	Temperature (°C)
	ObjectiveSettings		*Describes any settings on or around the objective*
	CorrectionCollar		An adjustable ring on the objective that corrects for changes in immersion medium refractive index; arbitrary scale and unitless
	Medium		A description of a medium used for the lens, e.g., Oil, Water, WaterDipping, Air, Multi, Glycerol, Other
	RefractiveIndex		Refractive index is that of the immersion medium
	Pixels		*Defines the location and parameters of the Pixels, the actual binary image data*
	DimensionOrder		The order in which the individual planes of data are interleaved, e.g., XYZCT, XYZTC, XYCTZ, XYCZT, XYTCZ, XYTZC
	PhysicalSizeX	}	Physical size in x, y, and z of a pixel in micrometers (µm)
	PhysicalSizeY		
	PhysicalSizeZ		
	SizeC	}	Dimensional size x, y, z, c, and t of pixel data array
	SizeT		
	SizeX		
	SizeY		
	SizeZ		
	TimeIncrement		Used for time series that have a global timing specification instead of per-timepoint timing info, e.g., a video stream (s)
	Type		The variable type used to represent each pixel in the image, e.g., int8, int16, int32, uint8, uint16, uint32, float, bit, double, complex, double-complex
	BinData		If the pixel data is stored directly in the XML it is enclosed in BinData Elements
	TiffData		*If the pixel data is stored in an OME-TIFF file it is described by TiffData Elements*
	FirstC	}	The TiffData element describes how the TIFF IFD numbers are mapped to the Pixels
	FirstT		
	FirstZ		
	IFD		
	PlaneCount		
	UUID		*The TiffData UUID and Filename are used for multi-file datasets to maintain connections between the files*
	FileName		
	Channel		
	AcquisitionMode		AcquisitionMode describes the type of microscopy performed, e.g., WideField, LaserScanningMicroscopy, LaserScanningConfocal, SpinningDiskConfocal, SlitScanConfocal, MultiPhotonMicroscopy, StructuredIllumination, SingleMoleculeImaging, TotalInternalReflection, FluorescenceLifetime, SpectralImaging, FluorescenceCorrelationSpectroscopy, NearFieldScanningOpticalMicroscopy, SecondHarmonicGenerationImaging, Other
	Color		A color used to render this channel
	ContrastMethod		The technique used to achieve contrast, e.g., Brightfield, Phase, DIC, HoffmanModulation, ObliqueIllumination, PolarizedLight, Darkfield, Fluorescence, Other
	EmissionWavelength		Emission wavelength of excitation for a particular channel, in nanometers (nm)
	ExcitationWavelength		Excitation wavelength of excitation for a particular channel, in nanometers (nm)
	Fluor		The name of the fluorophore used to produce this channel
	IlluminationType		The method of illumination used to capture the channel, e.g., Transmitted, Epifluorescence, Oblique, NonLinear, Other
	Name		A short name for the channel, used to, for example, identify the channel from a list
	NDFilter		Specifies the combined effect of any neutral density filters used (% transmittance)
	PinholeSize		Specifies adjustable pin hole diameters for confocal microscopes (micrometers [µm])
	PockelCellSetting		Amount the polarization of the beam introduced by Pockel Cell, if any
	SamplesPerPixel		The number of samples the detector takes to form each pixel value
	DetectorSettings		
	Binning		Represents the number of pixels that are combined to form larger pixels, e.g., 1 × 1, 2 × 2, 4 × 4, 8 × 8, Other
	Gain		The gain of the detector
	Offset		The offset of the detector
	ReadOutRate		Detector read speed (MHz)
	Voltage		The voltage of the detector (volts [V])
	LightSourceSettings		
	Attenuation		The attenuation of the light source (%)
	Wavelength		The wavelength of the light source (nm)
	Plane		
	DeltaT		Elapsed time since the beginning of the experiment (s)
	ExposureTime		Elapsed time during image recording (s)
	PositionX	}	The x, y, and z position of the stage (µm)
	PositionY		
	PositionZ		
StageLabel			
	Name		Short name for this stage location; this would be used to, for example, identify the channel from a list
	X	}	The labeled x, y, and z position of the stage (µm)
	Y		
	Z		

Recommendations for OME Compliant image metadata. The Image and Instrument Elements from the OME Data Model, with attributes and hierarchies shown in diagrammatic form. The Image Element contains core metadata that can be used for display and processing of the associated binary image data. Currently, an OME Compliant image completes all of the metadata in the Image Element. By the end of 2010, we aim to include the Instrument Element in the OME Compliant specification. The Bio-Formats library provides support for writing OME-XML either as a stand-alone file or within the header of an OME-TIFF file. The full XML Schema version of the OME Data Model is available at http://ome-xml.org/browser/Schemas/OME/2010-04/ome.xsd. Updates to the OME Data Model are announced on the project's roadmap site (http://ome-xml.org/roadmap).

Box 1. Recommendations for use of PFFs

1. Image metadata must be associated with the binary image data, preferably as a single file.

2. Microscope systems must not store metadata in proprietary databases that are available only on the data acquisition system.

3. Metadata must be readable by third party software using a common, openly accessible software package or library. PFF developers must work with developers of open translation libraries to ensure their format is correctly interpreted.

4. Scientists must use image processing and analysis tools that preserve image metadata.

5. Image data must reflect the original measurement. If compression is supported, the user must be given the option of saving uncompressed or losslessly compressed images (which allows the exact original data to be reconstructed after compression). If compression or encryption is used, the algorithm and parameters must be stated and stored in the metadata.

6. Commercial software programs must provide data export to an open metadata specification. To ensure that commercial software writes these formats correctly, open, freely available libraries and format validators must be available to enable compliance.

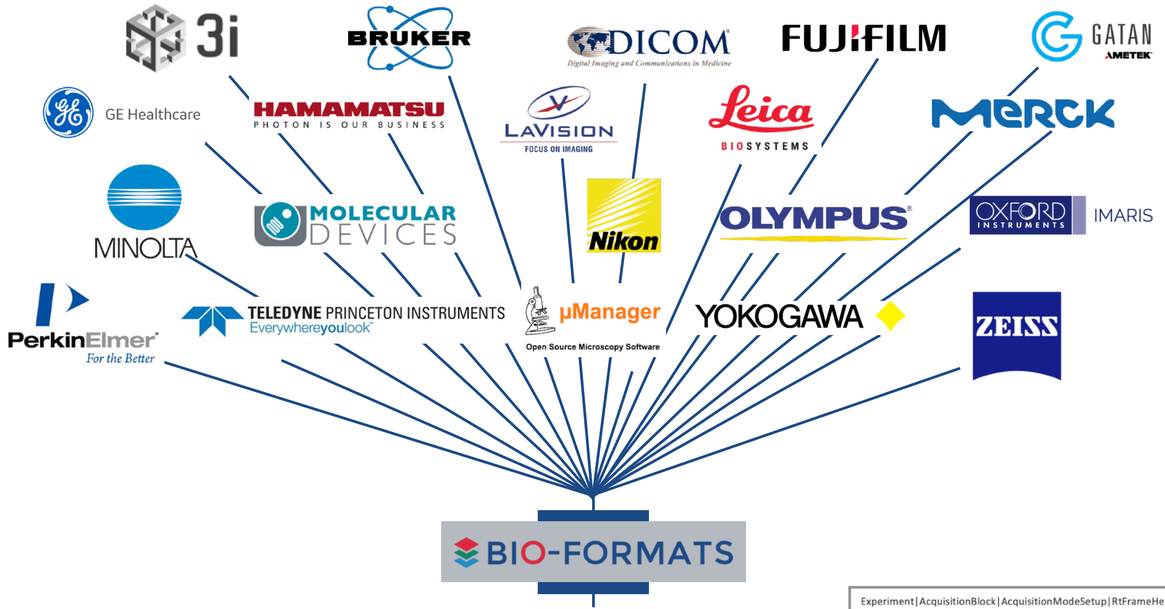
7. Public and charity funding for imaging systems must include a requirement that the system writes data in an open, accessible format, wherever possible.

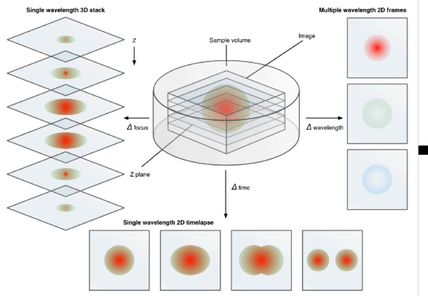
8. All file formats must use versioning to reflect any changes in the data model.

9. When PFFs must be used, new versions must be announced to the scientific community, and users and funding bodies must predicate their purchases on this type of support for the scientific community.

10. Once a standardized repository is available, journals must require deposition of original data supporting scientific manuscripts.

“6. Data export to an open metadata specification”





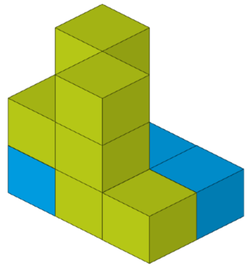
OME-XML

\+

```
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtFrameHeight #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtFrameWidth #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtLinePeriod #1=3.0000000000000001e-005
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtOffsetX #1=0
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtOffsetY #1=0
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtRegionHeight #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtRegionWidth #1=512
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtSuperSampling #1=1
Experiment|AcquisitionBlock|AcquisitionModeSetup|RtZoom #1=1
Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingX #1=4.1513291872589499e-007
Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingY #1=4.1513291872589499e-007
Experiment|AcquisitionBlock|AcquisitionModeSetup|ScalingZ #1=9.9999999999999995e-007
Experiment|AcquisitionBlock|AcquisitionModeSetup|SimRotations #1=3
Experiment|AcquisitionBlock|AcquisitionModeSetup|TimeSeries #1=false
Experiment|AcquisitionBlock|AcquisitionModeSetup|TrackMultiplexType #1=ZStack
Experiment|AcquisitionBlock|AcquisitionModeSetup|UseRois #1=false
Experiment|AcquisitionBlock|AcquisitionModeSetup|ZoomX #1=1
Experiment|AcquisitionBlock|AcquisitionModeSetup|ZoomY #1=1
Experiment|AcquisitionBlock|Laser|LaserName #1=HeNe633
Experiment|AcquisitionBlock|Laser|LaserName #2=Diode 405-30
Experiment|AcquisitionBlock|Laser|LaserName #3=Argon
Experiment|AcquisitionBlock|Laser|LaserName #4=HeNe594
Experiment|AcquisitionBlock|Laser|LaserPower #1=0.005000000000000001
Experiment|AcquisitionBlock|Laser|LaserPower #2=0.029999999999999
Experiment|AcquisitionBlock|Laser|LaserPower #3=0.025000000000000
Experiment|AcquisitionBlock|Laser|LaserPower #4=0.002
```

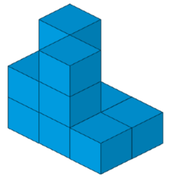
Uninterpretable metadata

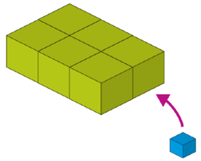
Complex internal state



Current Bio-Formats parsinging

Overall goal





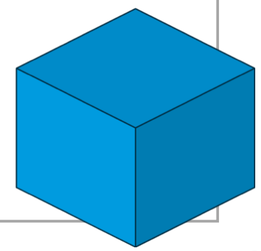
Uninterpretable information

>

RO-Crate

```
{ "@context": {“ROC”: “https://w3id.org/ro/crate/1.1/context”,
               “OME”: “https://openmicroscopy.org/context”,
               “QL”: “https://quarep.org/context”,
               “EG”: “https://vendor.com/context”},
  "@graph": [
    {
      "@type": “ROC:CreativeWork”,
      "@id": "ro-crate-metadata.json",
      "conformsTo": {"@id":
        “https://w3id.org/ro/crate/1.1"},
      "about": {"@id": "./"}
    },
    {
      "@id": "./",
      "@type": [
        “Dataset”, “OME:Image”,
      ],
      “EG:acquisitionModeSetup”: [
          “EG:UseRois”: false
      ], …
  }
```

Critical to have complete metadata if we are ever to have a unified format.



Creating metadata representations is easy.

The hard part is building a common one.

Acknowledgments

University of Dundee

German BioImaging
Gesellschaft für Mikroskopie und Bildanalyse

openmicroscopy.org/about/#consortium
NGFF community

Challenge participants USA **Allen** Daniel Toloudis **CZBioHub** Jordao Bragantini **HMS** Trevor Manz **JAX** Eric Perlman, Erick Ratamero, Kiya Govek, Peter Sobolewski Europe **Crick** Joost de Folter,Kenneth Ho, Martin Jones **Dundee** Will Moore **EBI** Aybuke Kupcu Yoldas, Francois Sherwood, Kola Babalola, Matthew Hartley **EMBL** Bugra Oezdemir **FBI** Guillaume GAY **FZ-Jülich** Julia Thoennissen Timo Dickscheid **Georg-August-Universität Göttingen** Constantin Pape, Jan Huisken **HHU** Tom Boissonnet **LIN Magdeburg** Torsten Stöter **Sanger** Tong Li **TU-Dresden** Cornelia Wetzker **UZH** Joel Lüthi, Lorenzo Cerrone **Uni Münster** Jens Wendt Japan **RIKEN** Hiroya Itoga, Koji Kyoda, Shuichi Onami Companies **Glencoe** Melissa Linkert, Sébastien Besson **Google** Jeremy Maitin-Shepard **Scalable Minds** Normal Rzepka

image.sc
AGAVE
AICSImageIO
Arkitekt
Aydin
BAND
BIAFLOWS
BiaPy
BIII
BiofilmQ
Bio-Formats
BioImageIO
BoneJ
BrainGlobe
Cellpose
CellProfiler
CLIJ
CytoMAP
Cytomine
DAIM
DeepLabCut
Fiji
FLIMLib
GerBI
GloBIAS
Icy
IDR
ilastik
ImageJ
ImageJ2
ImgLib2
ImJoy
ImSwitch
iRODS
JIPipe
JuliaImages
Mars
MCMICRO
MIA
MIB
µManager
MoBIE
ModularImageAnalysis
MorphoGraphX
MorphoNet
napari
NEUBIAS
NFDI4BIOIMAGE
OME
OMERO
OmeSliCC
OpenIRIS
OpenSPIM
Orbit
Piximi
PolusAI
PYME
Python-Microscope
QUAREP-LiMi
QuPath
scenery
SCIFIO
scikit-image
sciview
SmartMicroscopy
SR-Tesseler
StarDist
TeamTomo
TissUUmaps
vedo
VVDViewer
webKnossos
ZeroCostDL4Mic

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrasstructure – NFDI 46/1 – 501864659

DFG Deutsche Forschungsgemeinschaft

In cooperation with

nfdi

The NFDI4BIOIMAGE consortium comprises legally independent partners and does not act autonomously towards third parties. The authors represent the contributions from their respective affiliated institutions and work together for the project.

creative commons