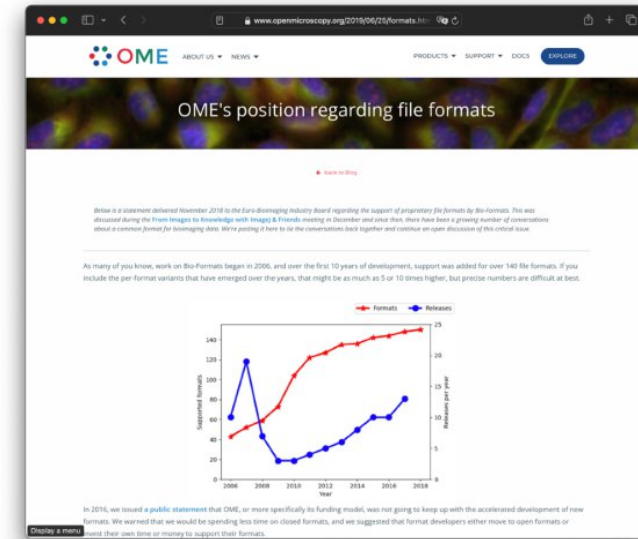




ome-zarr and transformations

Johannes Soltwedel

“Put simply, the format landscape has scaled beyond a manageable level”
(2019)



Paying for the initial cost of a format is not enough.

But paying for the initial cost of a format is not enough. The need for indefinite support carries a larger, longer-lived price tag that leaves data written in a given format constantly at risk. These costs are exacerbated by format variants. Even when a format is defined following standards like DICOM, there is a need to contend with multiple implementations as is the case in the radiology domain. The same happened with the Olympus OIB format added in 2017 in partnership with Olympus Europe. Following public release, the community has periodically reported breakages caused by new variants of the format. ^{1,4,7,8,9}

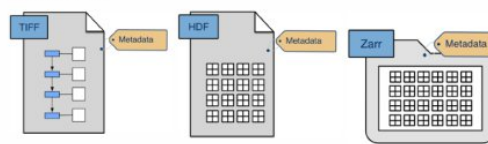
Put simply, the format landscape has scaled beyond a manageable level.

Put simply, the format landscape has scaled beyond a manageable level. The result is that scientists end up blocked in accessing and properly handling their data, and thus blocked in their scientific endeavor. If Bio-Formats were to cease to exist, a large percentage of imaging data would immediately cease to be accessible at least until someone took on the burden of support.

We understand the push to develop new formats. From numerous interactions, we know how crucial it is for data producers to be able to write data quickly as well as it is for users to be able to access their data quickly, and both across as many platforms as possible. We also know that, optimally, this ecosystem should all just keep working for years to come. But while these requirements need to be fulfilled, something must give.

We think the only scalable way forward is to work together on an ever smaller number of formats.

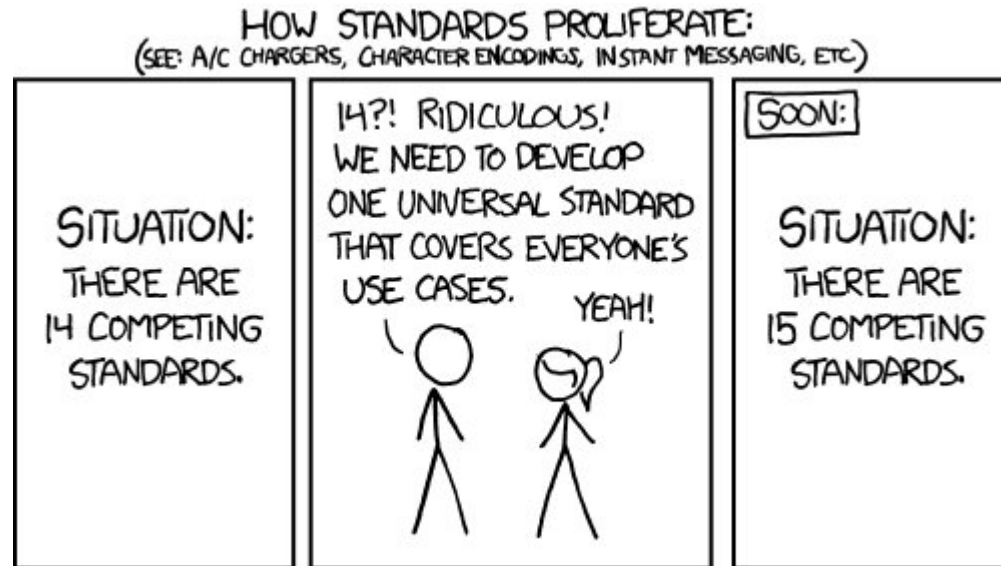
We think the only scalable way forward is to work together on an ever smaller number of formats. That's why we've been concentrating on open formats instead of adding new proprietary formats. For example, Bio-Formats 6.1 adds support for the open BigDataViewer (BDV) format, a strong candidate for support across the community.



BDV provides a tested for moving beyond the current single binary format of OME-TIFF. The OME Model will be extended to permit describing the multichannel, multidimensional data that is currently stored in BDV XMLAS. As a stable container format, HDF5 allows us a quick way to validate these concepts.

<https://www.openmicroscopy.org/2019/06/25/formats.html>

Why do we need a new (next-gen) file format?



Shared under CC BY-NC 2.5

<https://creativecommons.org/licenses/by-nc/2.5/>

So far, OME, bioformats have your back:

Supported Formats

Ratings legend and definitions

Format	Extensions	Pixels	Metadata	Openness	Presence	Utility	Expert	BSD	Multiple Images	Pyramidal
3i SlideBook	.sid	▲	▲	▼	▲	▼	✗	✗	✓	✗
3i SlideBook 7	.sldy	▲	▲	▲	▲	▲	✗	✓	✗	✗
Ander Bio-imaging Division (ABD) TIFF	.tif	▲	▲	▲	▼	▲	✗	✗	✓	✗
AIM	.aim	▲	▼	▼	▼	▼	✗	✗	✗	✗
Alicona 3D	.al3d	▲	▲	▲	▼	▲	✗	✗	✗	✗
Amersham Biosciences Gel	.gel	▲	▲	▲	▼	▼	✗	✗	✗	✗
Amira Mesh	.am, .amiramesh, .grey, .hx, .labels	▲	▼	▼	▼	▼	✗	✗	✗	✗
Amnis FlowSight	.cif	▲	▲	▲	▼	▼	✗	✓	✓	✗

Read more:

<https://www.openmicroscopy.org/bio-formats/>

NGFF/OME/ZARR?

NGFF documentation

Q Search Ctrl + K

Community

Contributing

Specifications

RFCs

Resources

Help Desk

Next-Generation File Formats (NGFF) + OME-Zarr

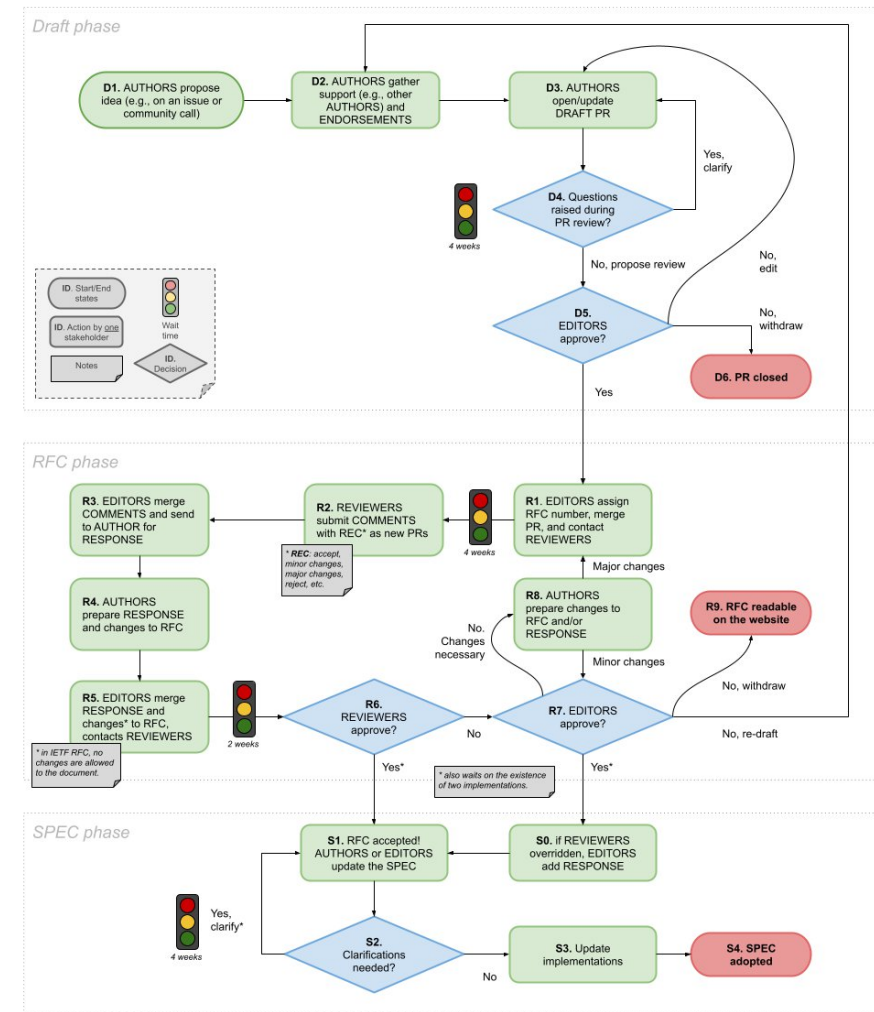
Welcome to the Next-Generation File Formats (NGFF) main page! This site is dedicated to providing resources for the NGFF community and those that are interested in getting started with OME-Zarr.

NGFF vs OME-Zarr, what is the difference?

OME-Zarr is the file format that the NGFF community has settled on to address issues of scalability and interoperability described below.

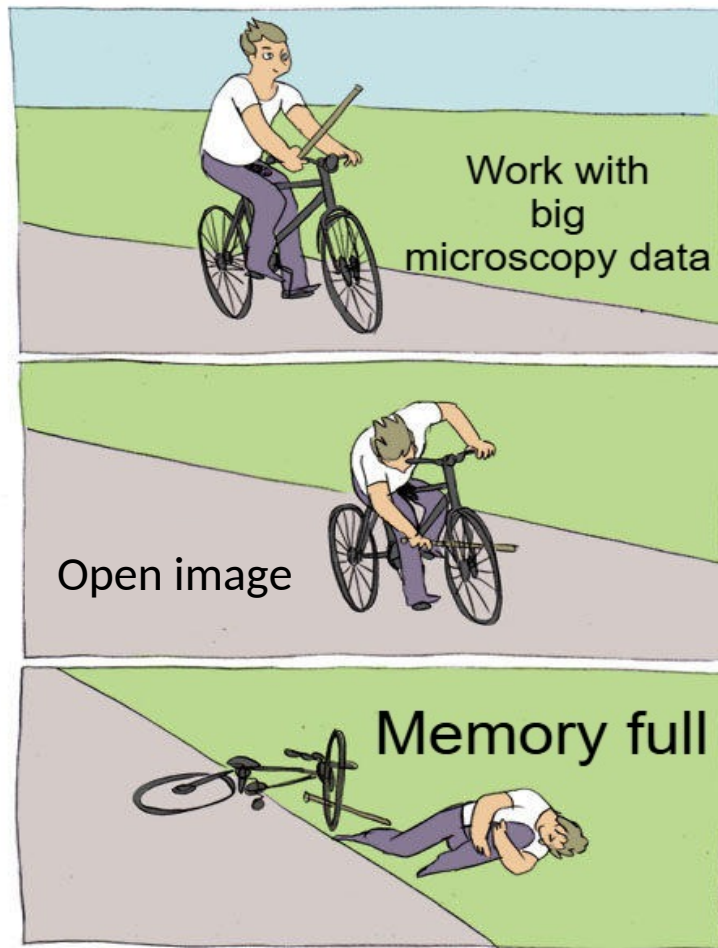
NGFF is the community-driven process for designing the next generation of bioimaging formats. NGFF brings together the community to define shared specifications, metadata standards, and best practices. OME-Zarr implements those decisions, providing a practical, open, and scalable way to store and share modern microscopy data. As the NGFF specifications evolve, OME-Zarr evolves with them — ensuring the format reflects the needs and experience of the wider community.

ngff.openmicroscopy.org



NGFF Process Diagram v3.0 | Last updated: 2024-05-03

Chunked file formats

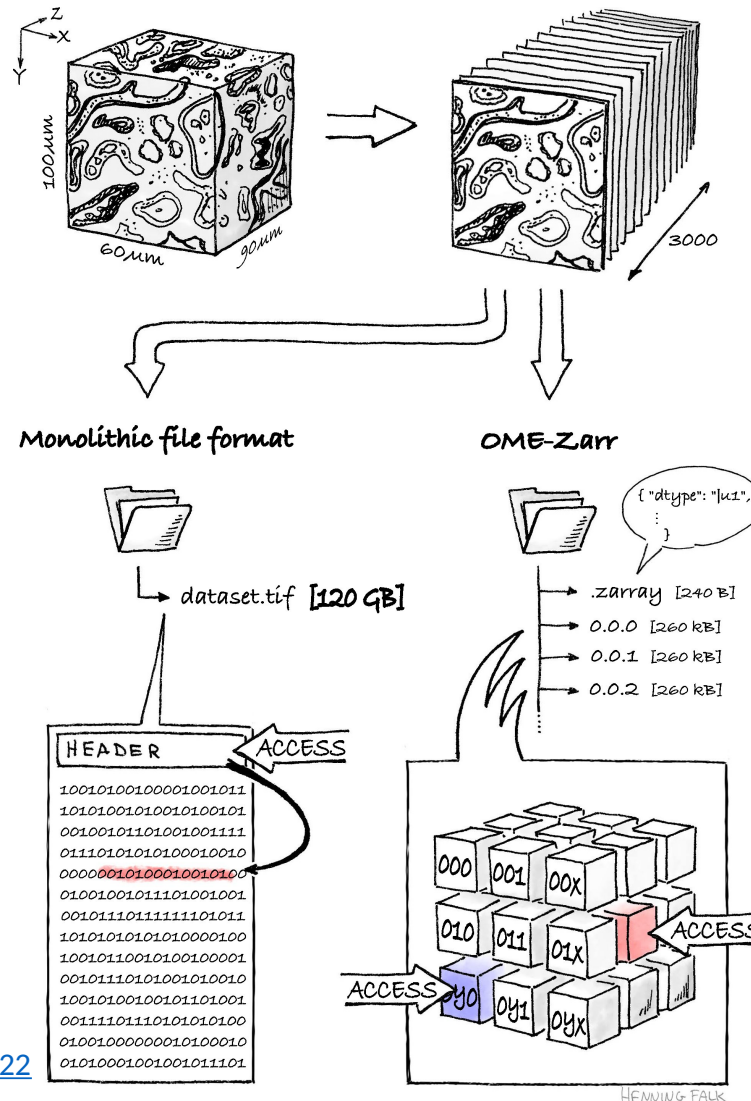


imgflip.com

Image „Monolithic vs. chunked“ by Henning Falk

<https://github.com/zarr-developers/zarr-illustrations-falk-2022>

(c) NumFOCUS (CC-BY-4.0)



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

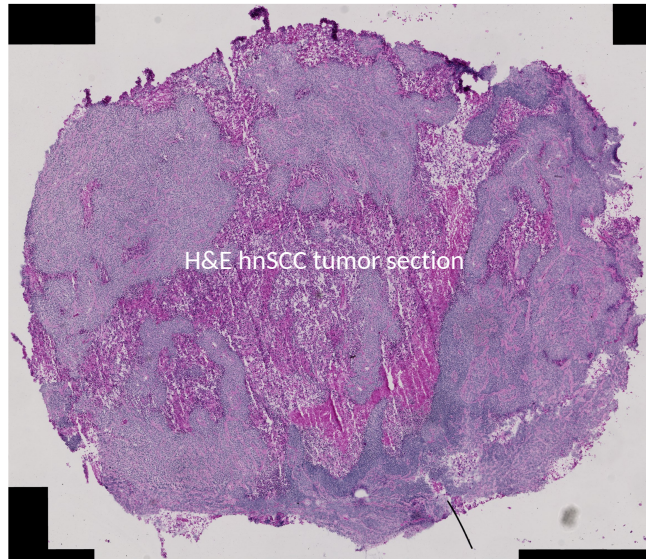
Pros:

- natively supports parallel read/write access to chunks
- avoid costly downloads or transfer of large data on cloud infrastructure (e.g., S3)
- web-/cloud-ready

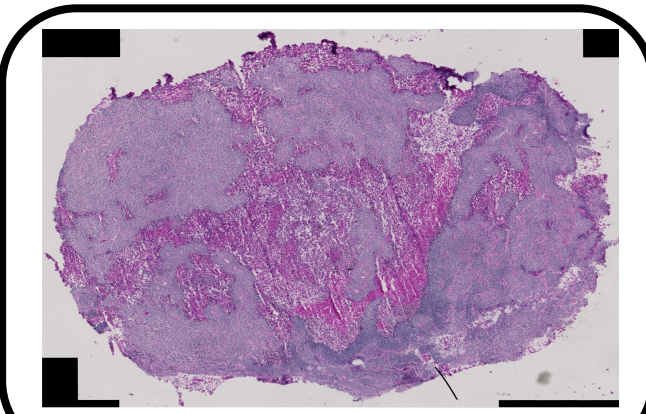
Cons:

- A lot of files on disk

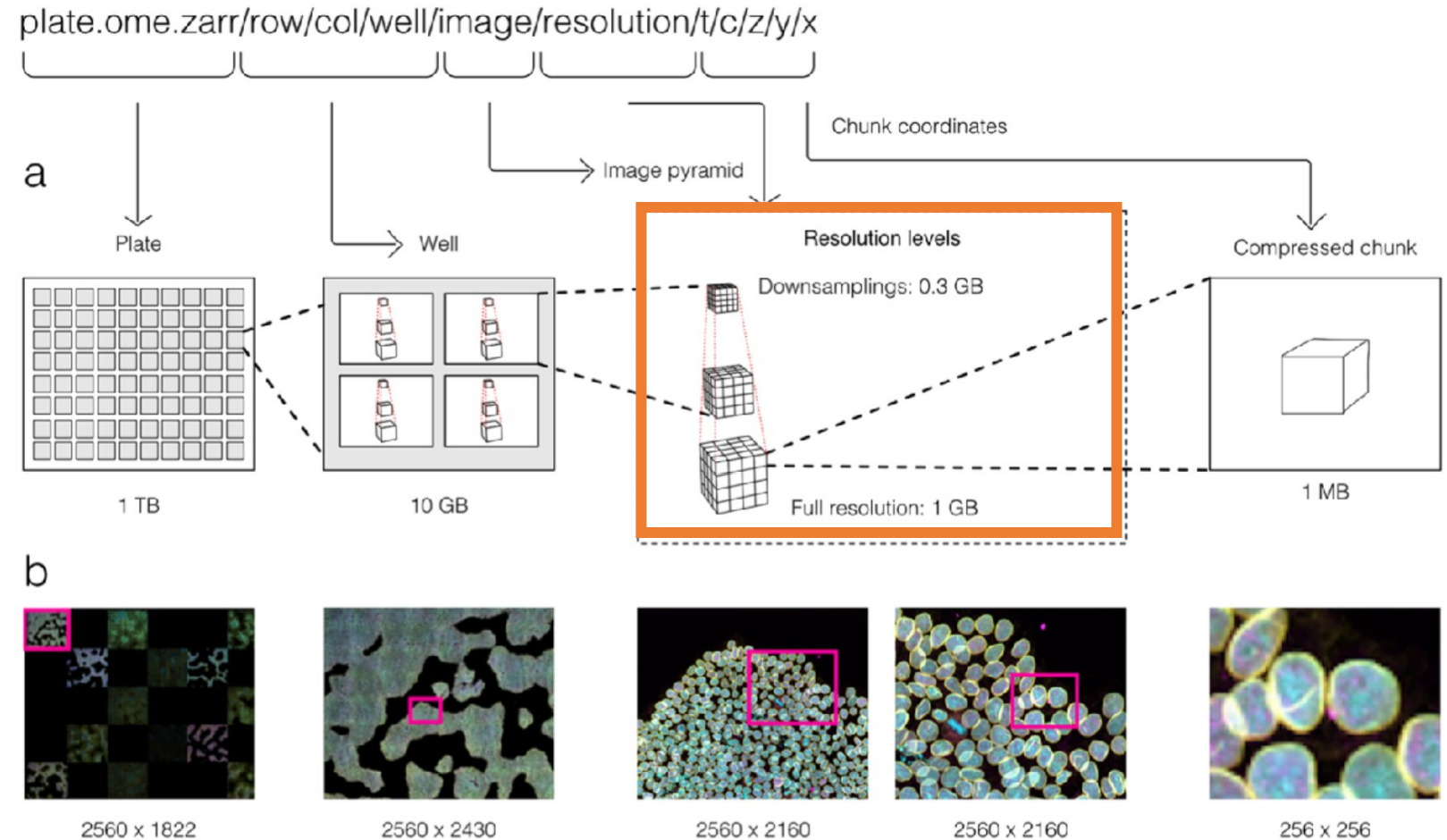
Pyramidal layout



Typical WSI image: ~30k x 30k pixels



Typical monitor: 2550 x 1440 pixels



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

S3 cloud storage

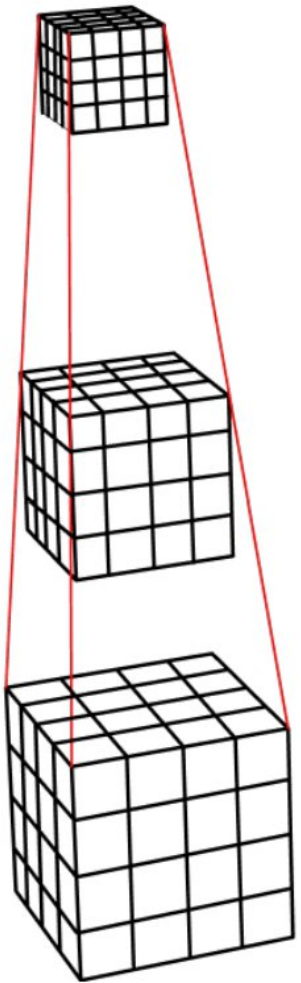
S3: **S**imple **S**torage **S**olution – first introduced by AWS

	Local	S3
Where's the data?	Local drive	Remote cloud
Data is stored as	Files	Objects with metadata
Hierarchy	Folders	None (<i>Data lake</i>)
Data can be accessed through	File browser (loading)	Webbrowser (streaming)
Sharing	Copy data	Via URL
Scalability	Buy more harddrives	Scales virtually indefinitely
Permissions	Per network-drive, per user locally	Data in buckets



Image „Clara shares“ by Henning Falk
<https://github.com/zarr-developers/zarr-illustrations-falk-2022>
(c) NumFOCUS (CC-BY-4.0)

Becoming OME-ZARR



```
"multiscales": [  
  {  
    "coordinateSystems": [ ... ]  
  },  
  "datasets": [  
    {  
      "path": "scale0/image",  
      "coordinateTransformations": [ ... ]  
    },  
    {  
      "path": "scale1/image",  
      "coordinateTransformations": [ ... ]  
    },  
    {  
      "path": "scale2/image",  
      "coordinateTransformations": [ ... ]  
    },  
    {  
      "path": "scale3/image",  
      "coordinateTransformations": [ ... ]  
    }  
  ]  
],
```


[item1, item2..]: Array of stuff
{"key": "value"}: Dictionary of stuff

Human-readable metadata header file contains all important information about image (axis ordering, pixel sizes, etc)

This is what „the spec“ defines:

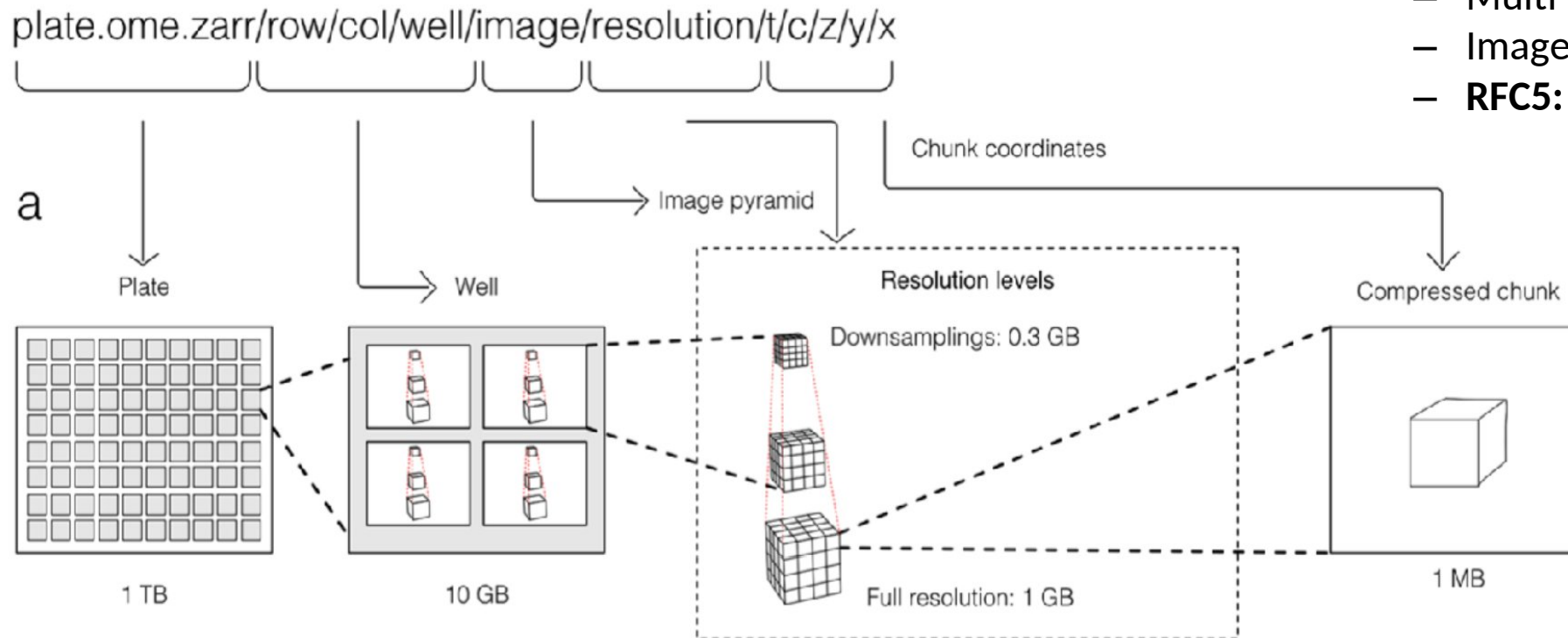
[..] "multiscales" contains a list of dictionaries where each entry describes a multiscale image. [...] Each "multiscales" dictionary MUST contain the field "datasets", which is a list of dictionaries [...]

RFCs

QUARTER	RFC	DESCRIPTION	LEAD	STATUS
Through Q3 2025	RFC-2	Zarr v3	scalableminds	Done
<div>  Q4 2025 </div>	RFC-3	5D+	Monash	In progress
	RFC-5	Transforms	Janelia / GerBI	In progress
	RFC-6	Simplify multiscales	scalableminds	In progress
	RFC-8	Collections	scalableminds	In progress
	RFC-9	Zips	SciLifeLab	In progress
Q1 2026	RFC-7	Provenance	Broad	Drafting
	RFC-10*	Roadmap + Editorial Board	GerBI	Drafting
	RFC-11*	Core concepts	BioVisionCenter	Discussion
Q2 2026	RFC-12*	Metadata	GerBI	Discussion
	RFC-13*	Clean-up	Allen	Discussion
	RFC-14*	<u>v1.ALPHA</u>	TBD	Discussion

OME-ZARR provides support for collections of images:

- Multi-resolution pyramids
- Multi-well plates
- Image + label
- **RFC5: Spatial relationships**

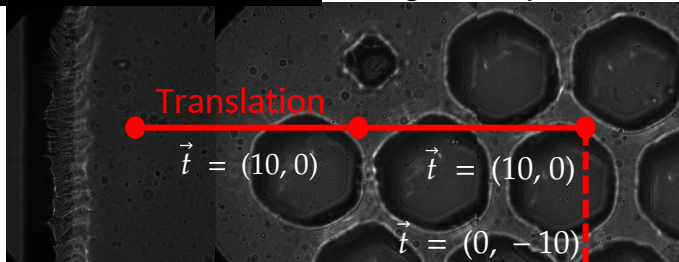


Moore et. al: Histochemistry and Cell Biology (2023)

ses

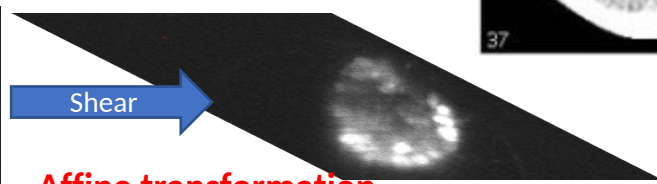
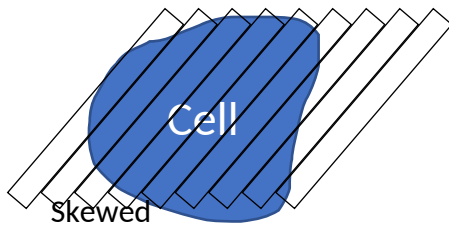
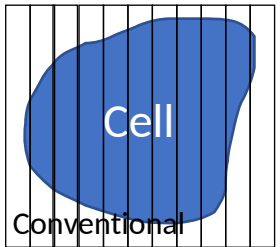
tion/Stitching

the image tile by tile



Microscope stage movement

Skewed acquisition

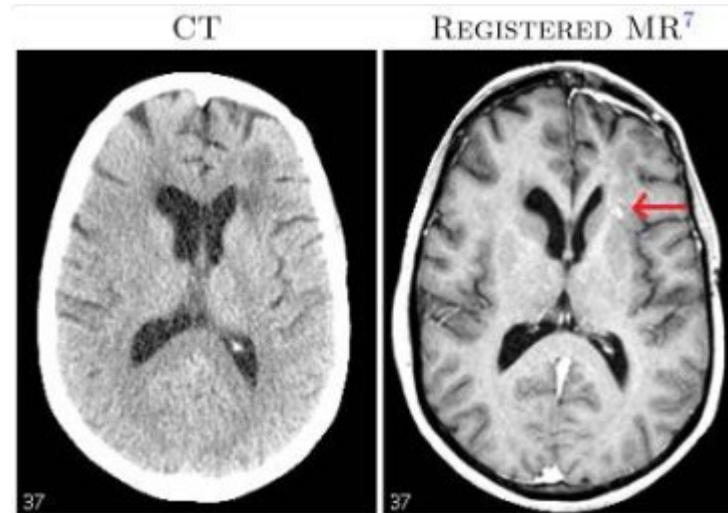


Affine transformation

Reconstruction of 3D volume
from individual 2D images

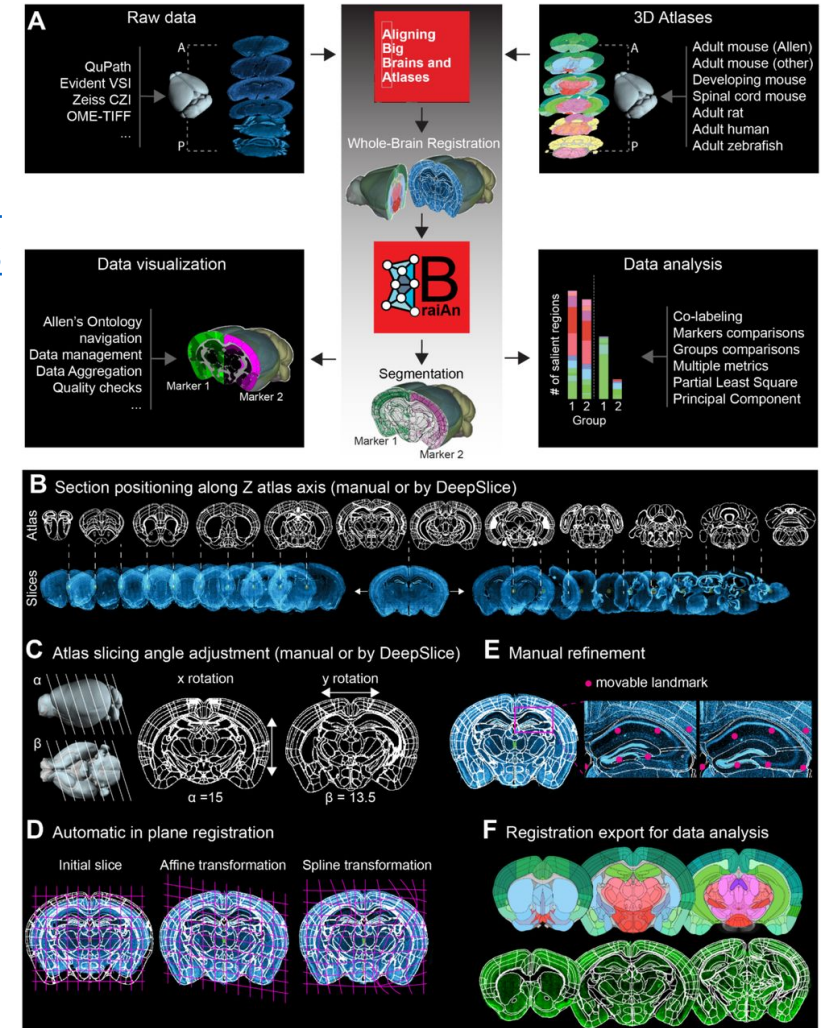
<https://github.com/BIOP/ijp-imagetoatlas>

Multi-modal image registration



Lee et al (2021)

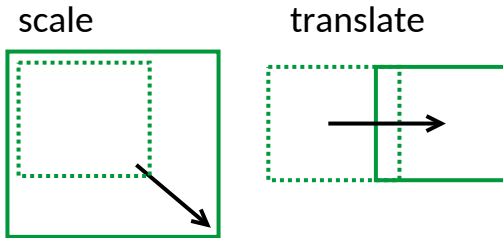
Atlas registration/3D reconstruction



Chiaruttini et al. (2024), shared under [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

RFC5 proposal

Current transformations



```
"multiscales": [  
  {  
    "coordinateSystems": [...]  
  },  
  "datasets": [  
    {  
      "path": "scale0/image",  
      "coordinateTransformations": [...]  
    }  
  ]  
},
```

New transformations:

scale, translate, rotate, affine,
sequence, displacements,
coordinates, bijection,
byDimension

Project lead:
John Bogovic,
HHMI Janelia, USA



```
{  
  "type": "affine",  
  "output": "sheared",  
  "input": "physical",  
  "name": "shear-transformation",  
  "affine": [  
    [  
      3,  
      0.4,  
      30  
    ],  
    [  
      0.3,  
      2,  
      20  
    ]  
  ]  
}
```

identity		The identity transformation is the do-nothing transformation and is typically not explicitly defined.
mapAxis	"mapAxis": List[number]	an axis permutation as a transpose array of integer indices that refer to the ordering of the axes in the respective coordinate system.
translation	one of: "translation": List[number], "path": str	Translation vector, stored either as a list of numbers (translation) or as a zarr array at a location in this container (path).
scale	one of: "scale": List[number], "path": str	Scale vector, stored either as a list of numbers (scale) or as a zarr array at a location in this container (path).
affine	one of: "affine": List[List[number]], "path": str	2D affine transformation matrix stored either with JSON (affine) or as a zarr array at a location in this container (path).
rotation	one of: "rotation": List[List[number]], "path": str	2D rotation transformation matrix stored as an array stored either with json (rotation) or as a zarr array at a location in this container (path).

<https://ngff.openmicroscopy.org/rfc/5>

Past & future

Coordinate systems and new coordinate transformations proposal

227



#138 by bogovicj was closed on Oct 31 • Draft 0.6

Rfc5 review response ✓

252

#350 by jo-mueller was merged on Oct 31 • Review required

RFC-5: Transformations and Coordinate systems ✓

61

#255 by bogovicj was merged on Oct 8, 2024 • Review required

WIP: Rfc5 review replies pt. 2 ✗ rfc

6

1

#389 opened 2 weeks ago by jo-mueller • Draft

Present

- Release specification 0.6 (transforms + minor stuff)
- Update popular read/write tools (ngff-zarr, ome-zarr-py, etc)

January

More resources

ngff.openmicroscopy.org

Acknowledgements

<https://image.sc>



Rueden CT, Ackerman J, Arena ET, Eglinger J, Cimini BA, Goodman A, et al. (2019)
Scientific Community Image Forum: A discussion forum for scientific image software.
PLoS Biol 17(6): e3000340. <https://doi.org/10.1371/journal.pbio.3000340>

Acknowledgements



NFDI 4
BIOIMAGE



OME-Zarr Contributors

BE **VIB** Benjamin Pavie // **CH** **Bern** Ana Stojiljkovic **BioVisionCenter** Caterina Fuster Barceló, Joel Lüthi, Lorenzo Cerrone, Valentina Marcelli, Virginie Uhlmann **EPFL** Rémy Dornier **ETH** Kevin Yamauchi, Leonardo Schwarz **FMI** Franziska Moos, Gustavo de Medeiros, Jan Eglinger, Joel Lüthi, Prisca Liberali **Zürich** Fabio Steffen, Flurin Sturzenegger, Grégory Paul, Lucas Pelkmans, Reto Gerber, Ruth Hornbachner // **CZ** **MUNI** Vladimír Ulman // **DE** **EMBL** Benedikt Best, Christian Tischer, Erick Ratamero, Luca Marconato, Oane Gros, Oliver Stegle, Wouter-Michiel Vierdag **Euro-Biolmaging** Bugra Özdemir **German Biolmaging** Josh Moore, Johannes Soltwedel, Tom Boissonnet, Chris Barnes, Tiago Lubiana, Janina Hanne, Stefanie Weidtkamp-Peters **Göttingen** Constantin Pape **Helmholtz München** Fabian J. Theis, Giovanni Palla, Isaac Virshup **TU Dresden** Stefan Hahmann // **FR** **France Biolmaging** Guillaume Gay **Pasteur** Daniel Krentzel, Marvin Albert // **JP** **RIKEN** Hiroya Itoga, Koji Kyoda, Shuichi Onami // **SE** **Lund University** Nils Norlin **SciLifeLab** Jonas Windhager, Wei Ouyang, Xavier Casas Moreno // **UK** **Crick** Joost de Folter, Kenneth Ho, Martin Jones **Dundee** David Gault, Dominik Lindner, Frances Wong, Jason R. Swedlow, Jean-Marie Burel, Josh Moore, Khaled Mohamed, Petr Walczysko, Sébastien Besson, Will Moore **EBI** Aybüke Küpcü Yoldaş, Francois Sherwood, Kolawole Babalola, Liviu Anita, Matthew Hartley **Sanger** Daniela Basurto-Lozada, Dave Horsfall, Martin Prete, Muzlifah Haniffa, Omer Bayraktar, Tong Li **UCL** David Stansby // **US** **Allen** Dan Toloudis, Eva M. Brown, Gabor Kovacs, Sean Meharry **Broad** Beth A. Cimini, Erin Weisbart **CZ Biohub SF** Ahmet Can Solak, Jordão Bragantini, Loic A. Royer, Merlin Lange **CZI** Jason R. Swedlow **Dartmouth** Yaroslav O. Halchenko **HMS** Ilan Gold, Mark S. Keller, Nils Gehlenborg, Trevor Manz **Janelia** John Bogovic, Konrad Rokicki, Lila Schweinfurth, Mark Kittisopikul, Stephan Preibisch, Stephan Saalfeld, Tobias Pietzsch **JAX** Fernando Cervantes Sanchez, Kiya Govek, Peter Sobolewski **MIT** Satrajit S. Ghosh **NCI** Kedar Narayan **NIH** Hythem Sidky, Nathan Hotaling, Nicholas Schaub, Sameeul Samee **Pittsburgh** Alan M. Watson **UMN** Rachael Huxford // **company** **Fideus** Matthew McCormick **Glencoe** Chris Allan, David R. Stirling, Erin Diel, Jason R. Swedlow, Kevin Kozlowski, Melissa Linkert **Google** Jeremy Maitin-Shepard **Roche** Dominique Roux, Maximilian Schulz **scalable minds** Albane le Tournoux de la Villegeorges, Hannes Spitz, Jonathan Striebel, Norman Rzepka **Yikes** Eric Perlman **Zeiss** Sebastian Rhode, Stephan Wagner-Conrad // **plus** **Many more** Christoph Engwer, Davis Bennett, Dmytro Horyslavets, Filip Mroz, Igor Zubarev, Szymon Stoma, Vera Galinova

ALONG WITH 120+ MEMBERS OF THE IMAGE.SC GROUP:

Aafke Gros Alan Liddell Andreas Eisenbarth Andrzej Oles Artür Manukyan Caleb Hulbert Cameron Lloyd Chris Wood Christophe Avenel Curtis Rueden David Feng Doug Shepherd Eduardo Brito Freitez Elise Liu Elliott Magnuson Eric Leroy Farzad Felix Meyenhofer Florian Levet Guillaume Witz Haowen Hugo_Blanc Igor Tatarnikov Jean-Karim Hériché Jens Wendt Jenu Chacko Jeremy Delahanty Johannes Elferich Jules Scholler Kimberly Meechan Ko Sugawara Kyle I S Harrington Lachlan Deakin Lorenzo Gaifas Luis Kuhn Cuellar Marco Dalla Vecchia Mark Tsuchida Matt Smith giovanni jni | Juan Nunez-Iglesias

...