

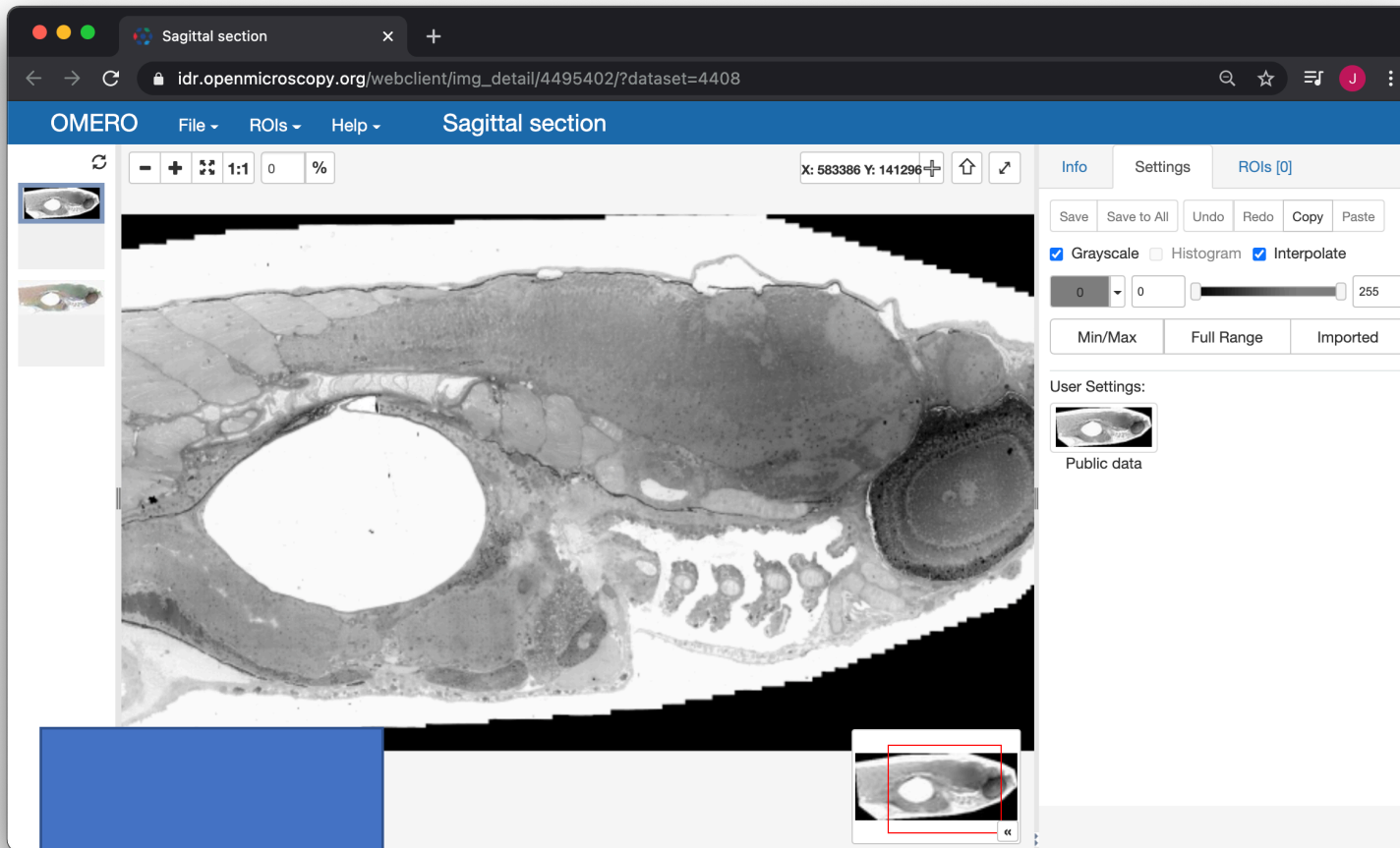
Accessing “cloud”- hosted image data

Day 3. Standard image file format for
sharing big image data in the cloud

Large data needs
chunks & pyramids to
be accessible.

Current formats for
chunked access tends
to be *monolithic*.

Monolithic formats
are difficult to access
remotely.



Single TIFF file
437 GB

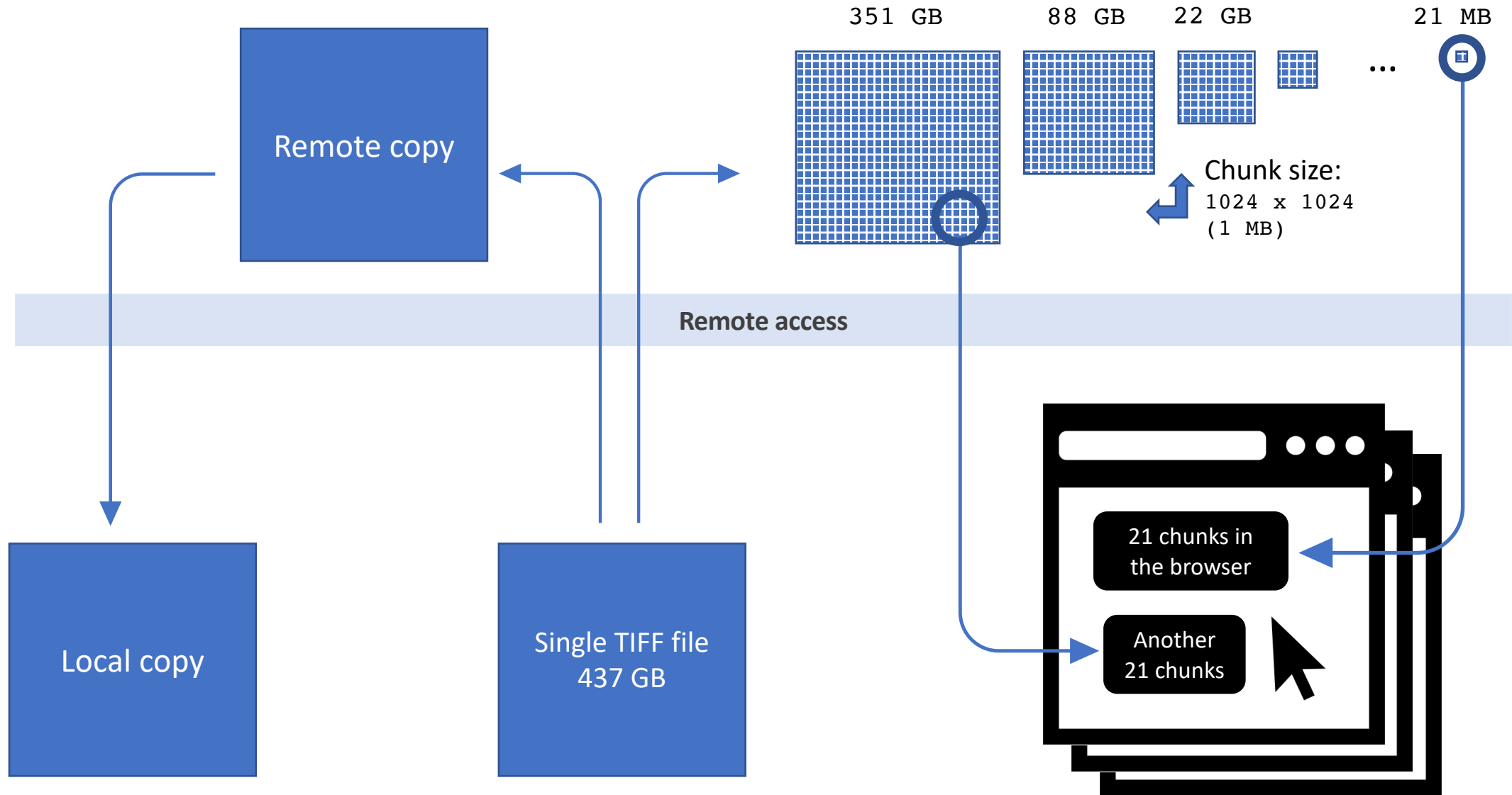
Faas *et al.* ([idr0053](#), CC BY-NC-SA 3.0)
J Cell Biol (2012)
921600 x 380928 pixels

Use cases:

1. Sharing beyond current limits, e.g., downloading TBs from OMERO.
2. Remotely sharing without a server.

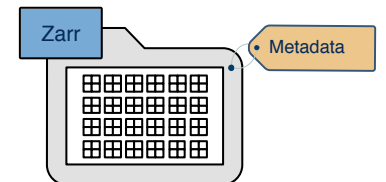
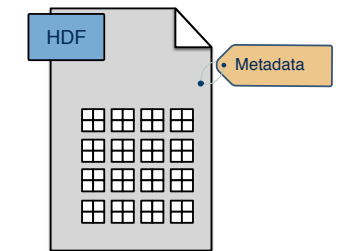
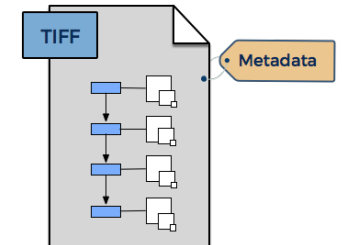
Monolithic format

Next-generation file format (NGFF)



Suite of open file formats

	Maturity	Expressivity	Multi-resolution	Performance
TIFF	Ubiquitous	2D tiles	TIFF spec	Linear slowdown
HDF	Well-supported	3D+ chunks	No spec yet	Slow writes *
Zarr/N5	Next-generation	3D+ chunks	NGFF spec	Highly parallel **



“Cloud” → “Storage”

Object

Commercial providers



Scientific providers



“Cloud” providers



Filesystems

- NFS/SMB
- GPFS
- HDD
- SDD

Filesystem	Object storage
1 €/GB	0.01 €/GB
Gbps	Tbps
10 μs	1 ms
I/O intensive	Immutable
High frequency	Versioned
...	...

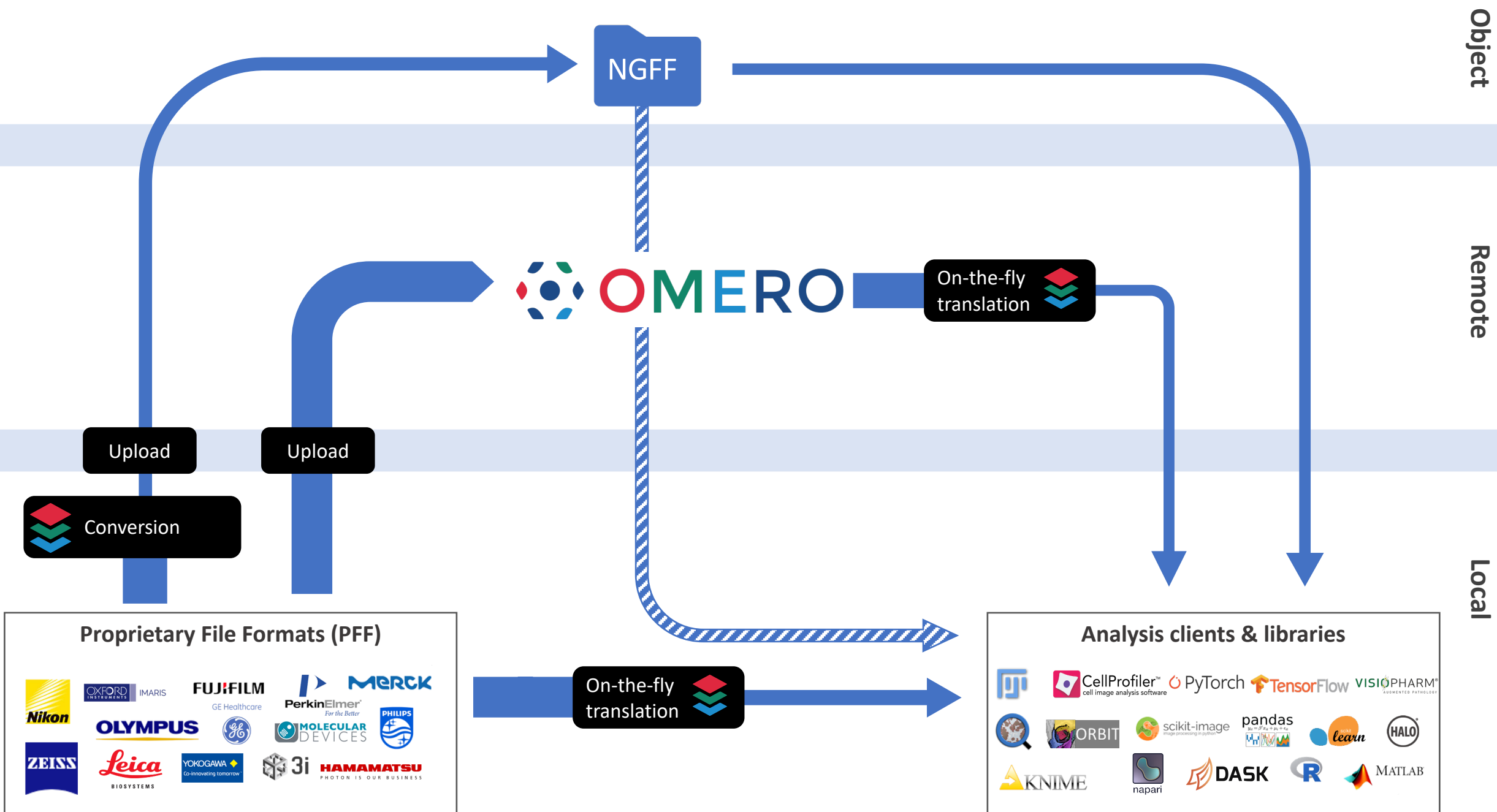
<https://www.openio.io/blog/block-file-object-storage-evolution-computer-storage-systems>

Software needs to be smarter.

Each object is a chunk.

Remote

Local



Object

Remote

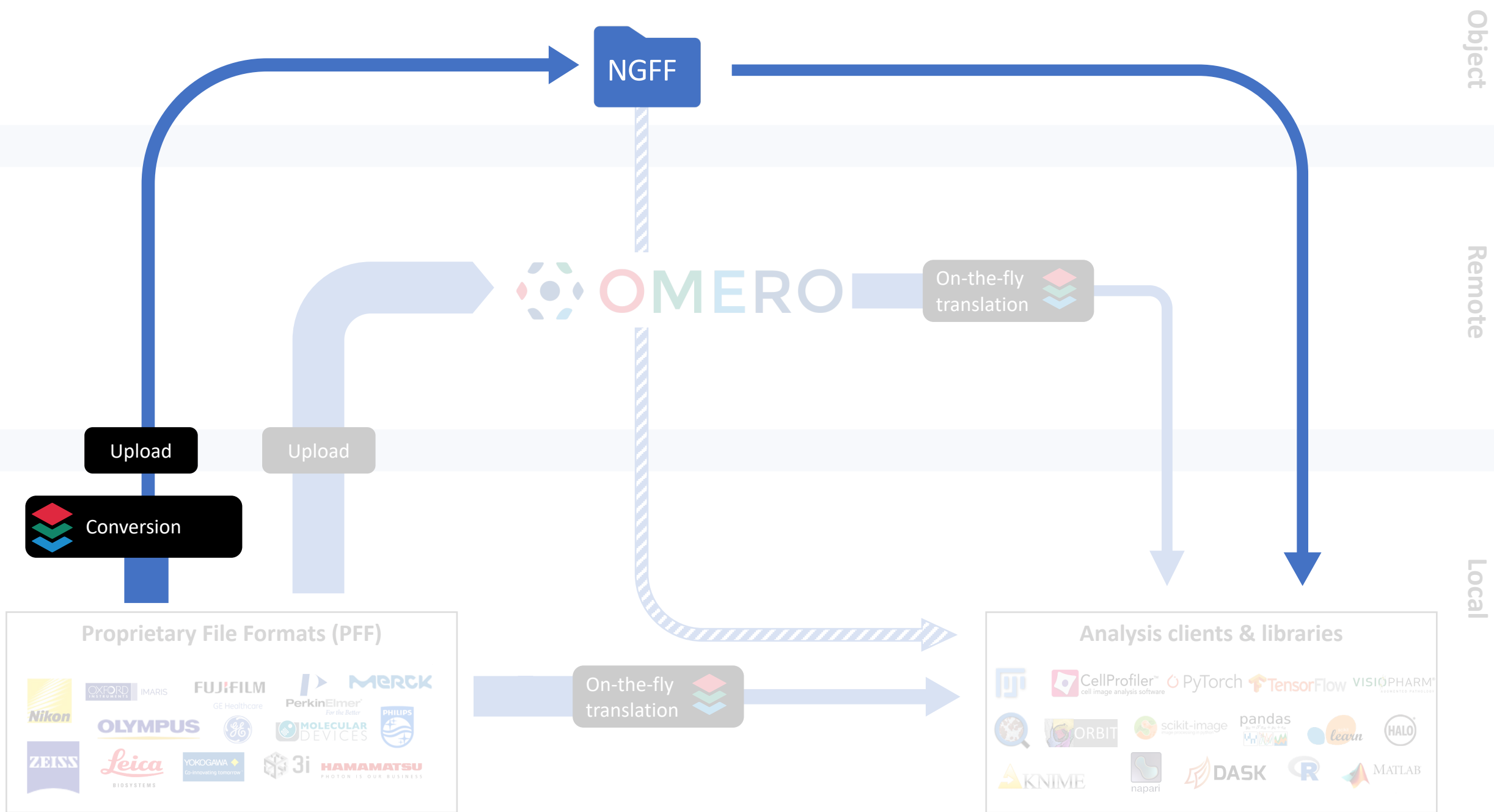
Local

Proprietary File Formats (PFF)


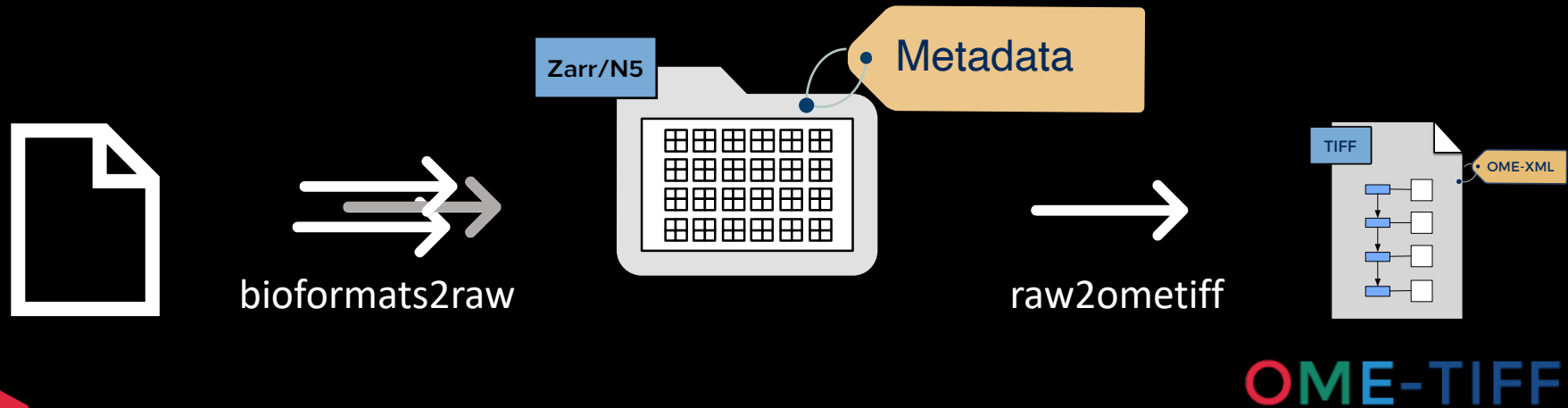


Analysis clients & libraries





Conversion

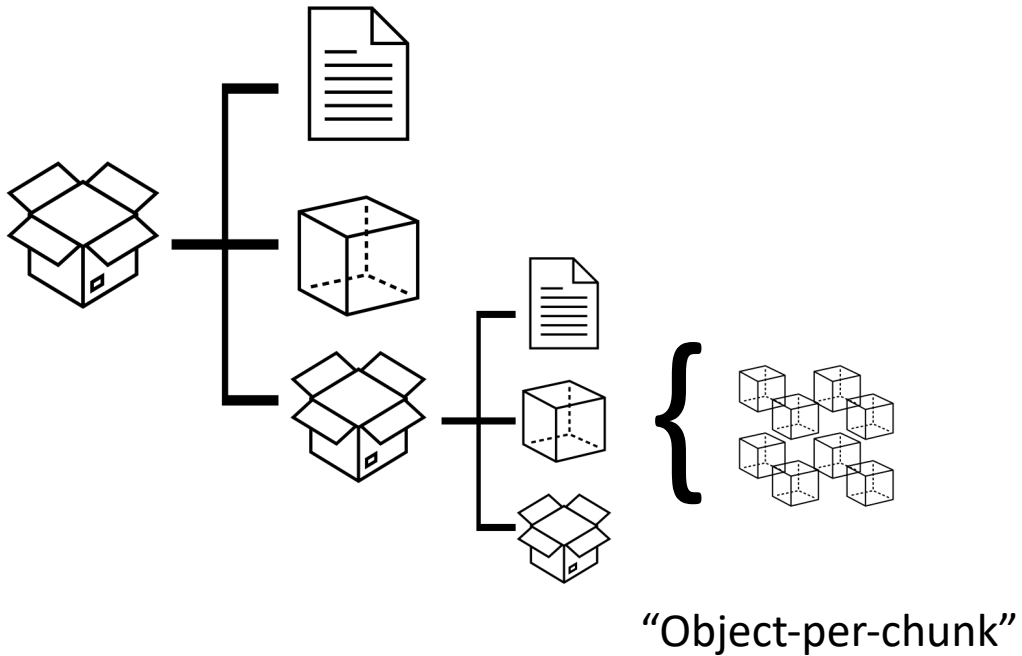
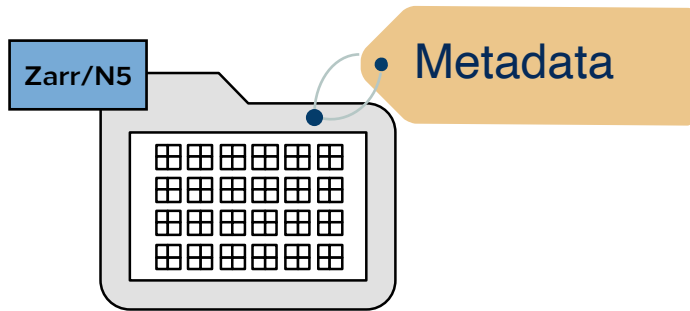


Chris Allan & Glencoe
Converting Whole Slide
Images to OME-TIFF: A New
Workflow



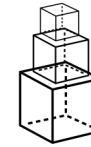
[glencoesoftware/bioformats2raw](https://github.com/glencoesoftware/bioformats2raw)
[glencoesoftware/raw2ometiff](https://github.com/glencoesoftware/raw2ometiff)





Specifications:

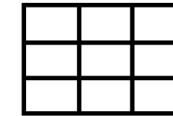
- Multiscales



- Labels



- HCS Plates



Process:

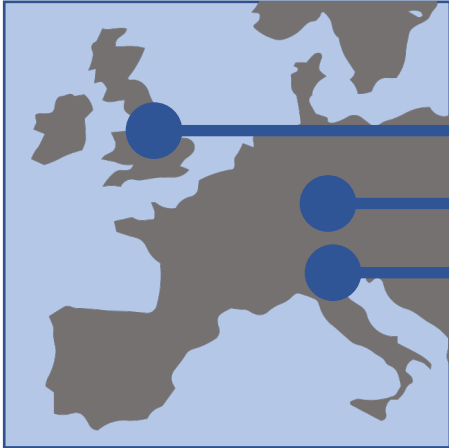
- Discussions: <https://image.sc>
- Publication: <https://ngff.openmicroscopy.org>
- Samples: <https://s3.embassy.ebi.ac.uk/idr/zarr>

Practical

<http://bit.ly/gbi-2021-ngff-1>

Download from S3
Convert your data
Share your data on S3

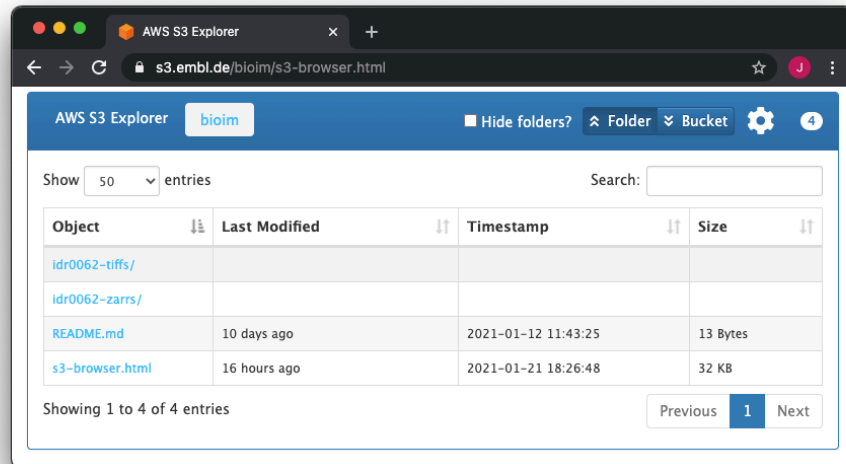
A few last words on S3



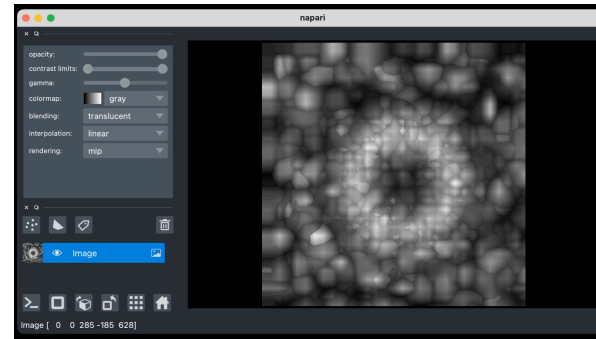
Endpoint	Bucket	Key	(Access)
https://s3.embassy.ebi.ac.uk	idr	zarr/v0.1/zarr/9822151.zarr	download
https://s3.embl.de	bioim	idr0062-tiffs/...	download, upload
https://s3.eu-south-1.amazonaws.com

Unique identifier for an object

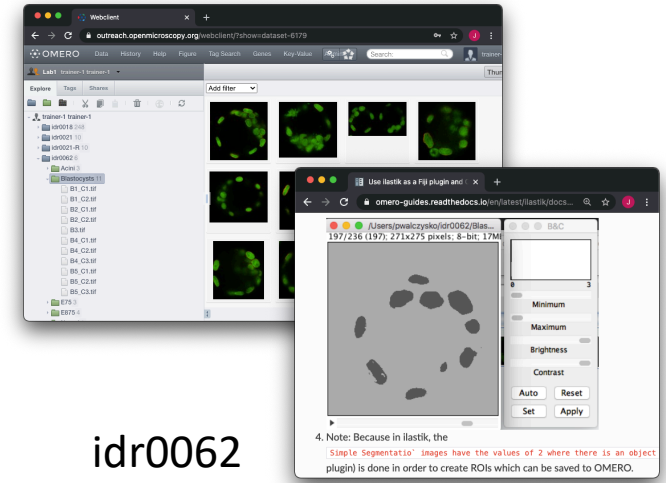
Z-stacks of TIFFs



<https://s3.embl.de/bioim/s3-browser.html>

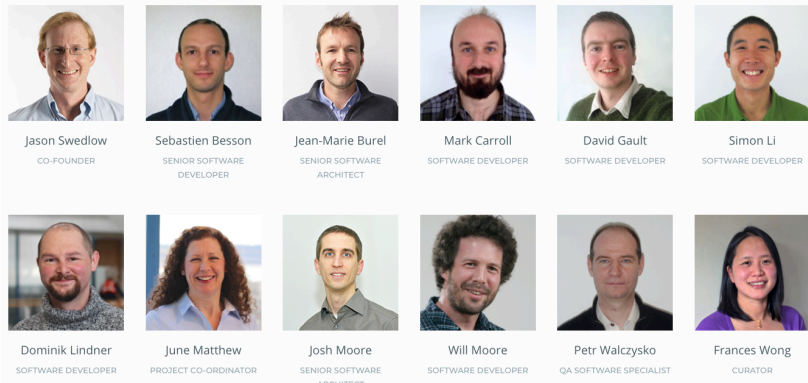


idr0023

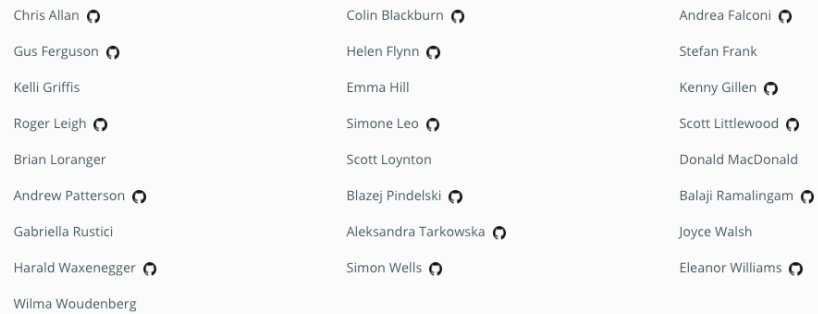


idr0062

Acknowledgements

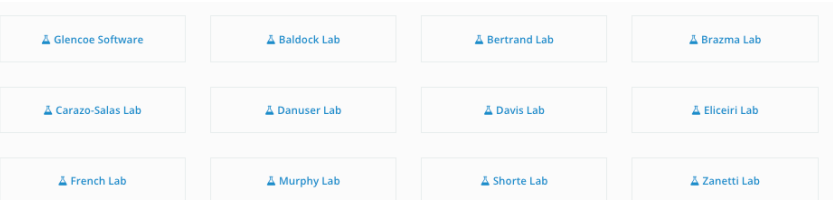


Former members of the OME team in Dundee

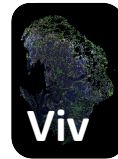


Development Teams

Other teams are also working on developing or integrating OME tools.



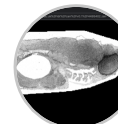
<https://www.openmicroscopy.org/teams>



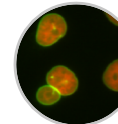
Gehlenborg Lab (HMS)



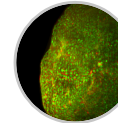
napari



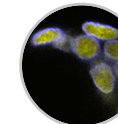
Faas *et al.* ([idr0053](#), CC BY-NC-SA 3.0)
J Cell Biol (2012)



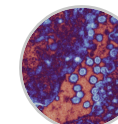
Hériché *et al.* ([idr0002](#), CC-BY 4.0)
MBoC (2016)



McDole *et al.* ([idr0044](#), CC BY 4.0)
Cell (2018)



Blin *et al.* ([idr0062](#), CC BY 4.0)
PLOS Biology (2019)



Lamers *et al.* ([idr0083](#), CC BY 4.0)
Science (2020)

BDV

Tomancak Lab
(MPI-CBG)

KLB

Keller Lab
(Janelia)

N5

Saalfeld Lab
(Janelia)

Zarr

Alistair Miles
(Oxford)



Chan
Zuckerberg
Initiative

