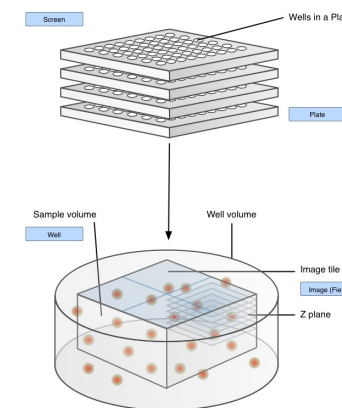# Next-generation file formats

Josh Moore

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

DANDI - Repository for Neurobio Data

BIDS - Brain Image Data Structure

NGFF - Next Generation File Format - successor to OME-TIFF

ZARR 3.0 - multidimensional array storage

Stack

Stack

Lee Kamentsky

Common Vocabulary

OME-FILES

BIO-FORMATS

OMERO

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

2

https://idr.openmicroscopy.org

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
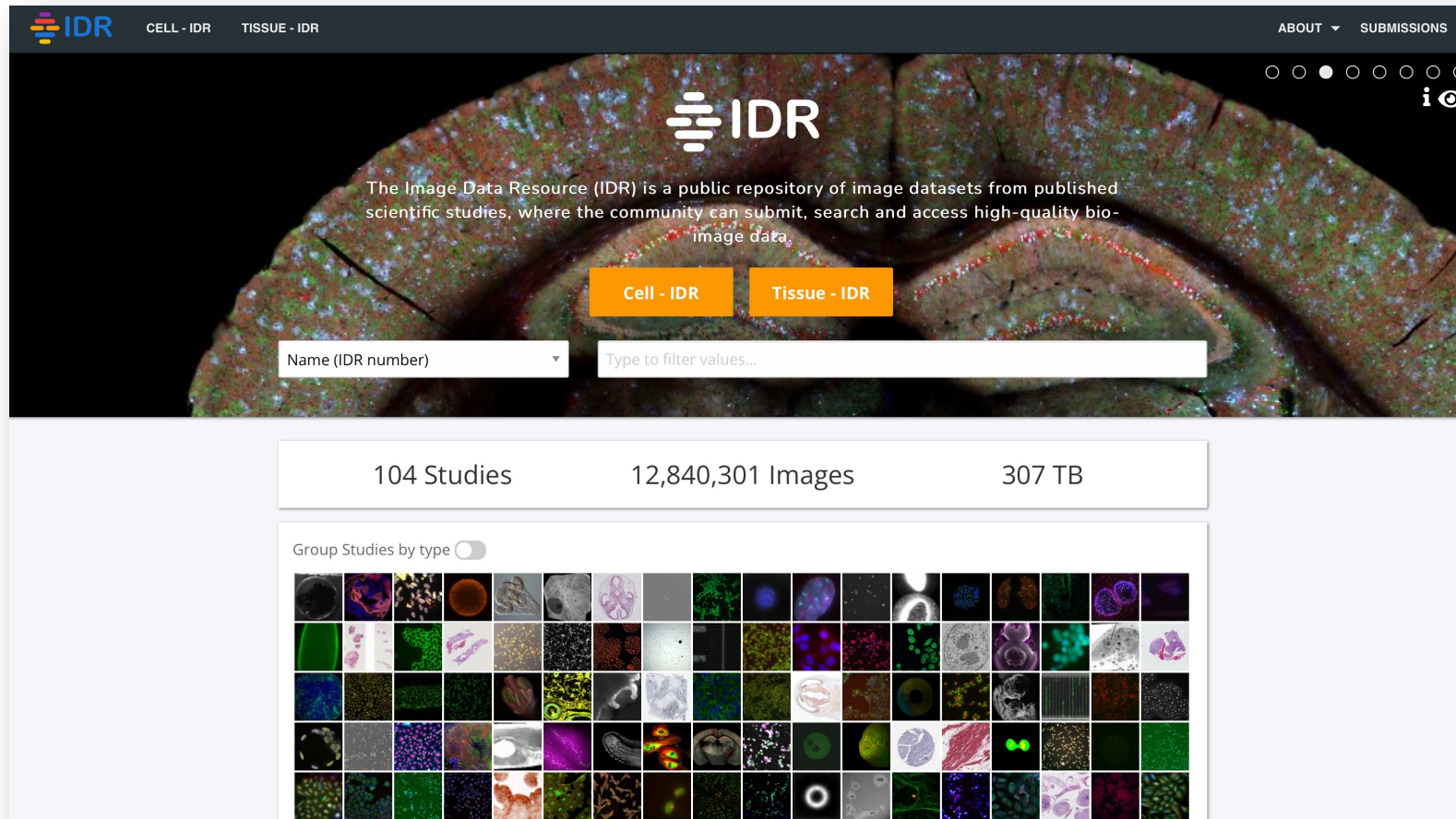Permission to share

Big Data Formats

BDV (HDF5)

Tomancak Lab (MPI-CBG)

KLB

Keller Lab (Janelia)
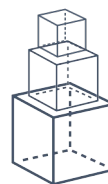
…

N5 [Java]

Saalfeld Lab (Janelia)

Zarr [Py]
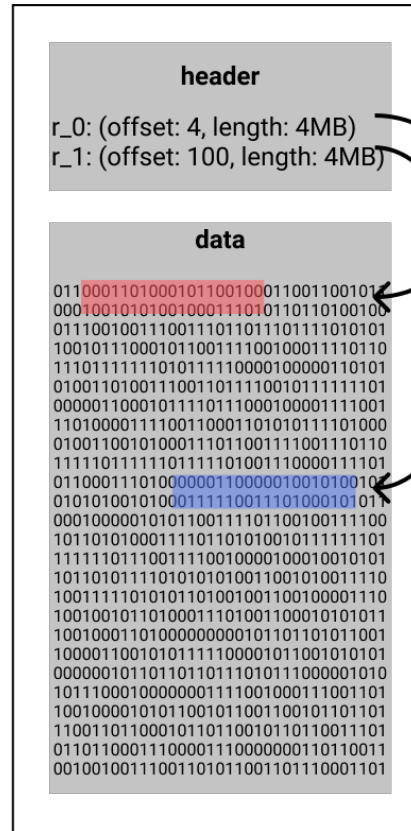
Alistair Miles (Oxford)

Next-generation
file formats (NGFF)

Multiscales

```
multiscales: [
  {
    version: 0.3,
    name: example,
    datasets: [ {path: 0}, {path: 1}, {path: 2} ],
    axes: [ t, c, z, y, x ],
    type: gaussian,
    metadata: {...}
  }
]
```

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

4

Monolithic file

header

r_0: (offset: 4, length: 4MB)
r_1: (offset: 100, length: 4MB)

data

Logical view

```
GROUP "ResolutionLevel_0":
  DATASET "Data":
    Shape: ( 64, 2048, 2048 )
    Chunks: ( 64, 256, 256)

GROUP "ResolutionLevel_1":
  DATASET "Data":
    Shape: ( 64, 1024, 1024 )
    Chunks: ( 64, 256, 256)
```

binary

file

folder

NGFF directory

.zgroup (json)   .zattrs (json)

"ResolutionLevel 0"

.zarray (json)   .zattrs (json)

0.0.0	0.0.1	...	0.0.x
0.1.0	0.1.1	...	0.1.x
...	...	...	...
0.y.0	0.y.1	...	0.y.x

"ResolutionLevel 1"

.zarray (json)   .zattrs (json)

| 0.0.0 | 0.0.1 | ... | 0.0.x |
| 0.1.0 | 0.1.1 | ... | 0.1.x |

...etc...

multiscales: [
  {
    version: 0.3,
    name: example,
    datasets: [{path: 0}, {path: 1}, {path: 2}],
    axes: [t, c, z, y, x],
    type: gaussian,
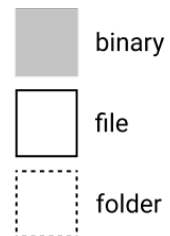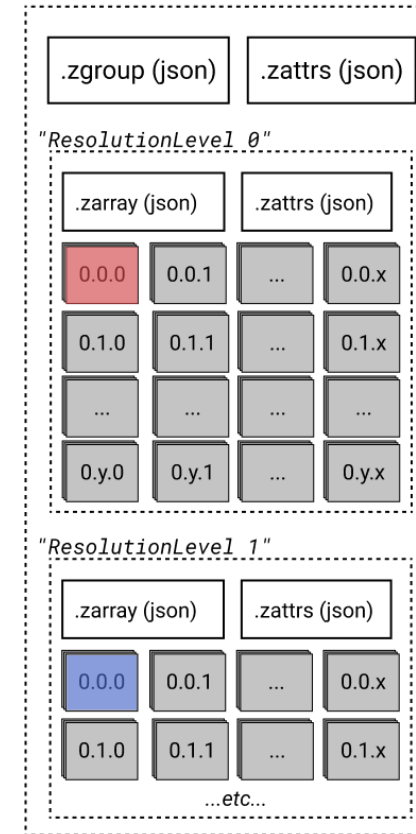   metadata: {...}
  }
]

Josh Moore
DANDI+BICCN Workshop, March 2022

University of Dundee

openmicroscopy
Permission to share

5

Monolithic file

header
r_0: (offset: 4, length: 4MB)
r_1: (offset: 100, length: 4MB)

data

Logical view

```
GROUP "ResolutionLevel_0":
    DATASET "Data":
        Shape: ( 64, 2048, 2048 )
        Chunks: ( 64, 256, 256)

GROUP "ResolutionLevel_1":
    DATASET "Data":
        Shape: ( 64, 1024, 1024 )
        Chunks: ( 64, 256, 256)
```
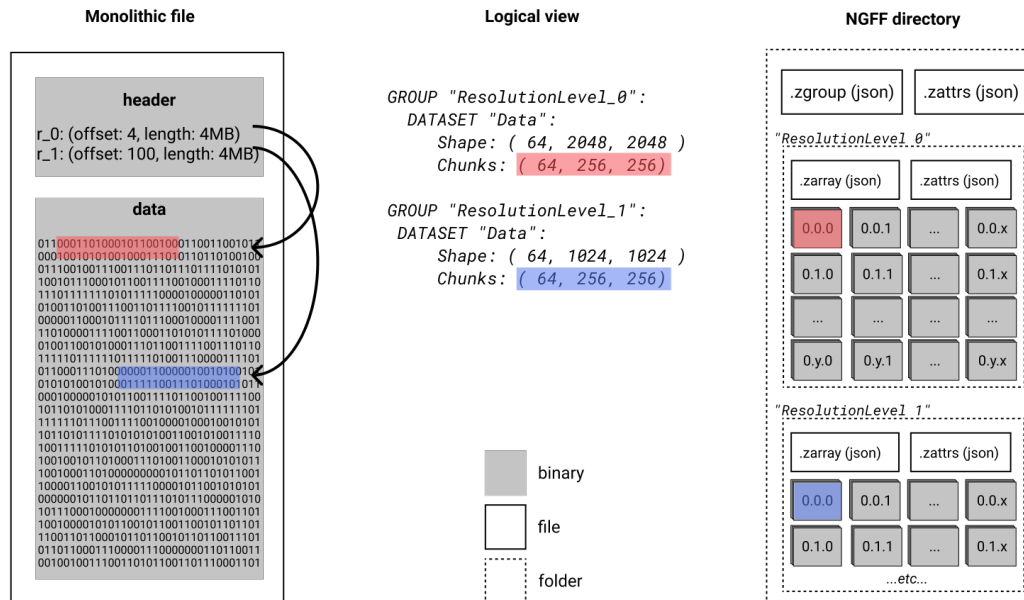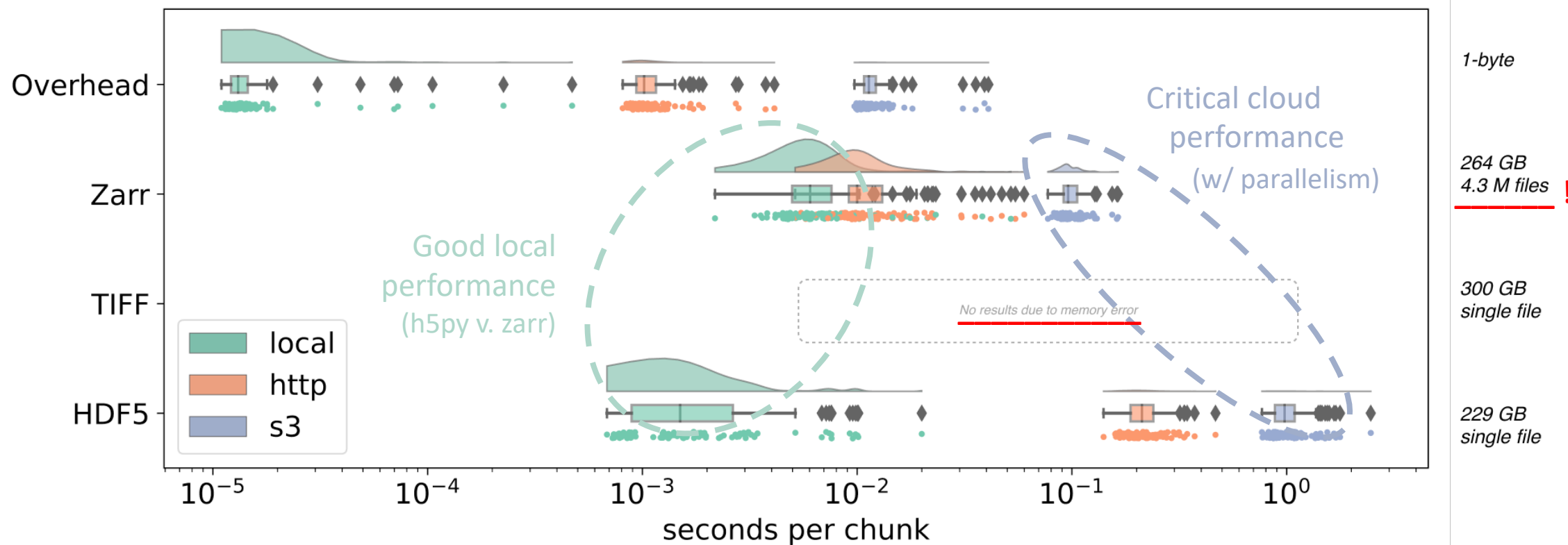
binary

file

folder

NGFF directory

.zgroup (json)   .zattrs (json)

"ResolutionLevel 0"

.zarray (json)   .zattrs (json)

0.0.0	0.0.1	...	0.0.x
0.1.0	0.1.1	...	0.1.x
...	...	...	...
0.y.0	0.y.1	...	0.y.x

"ResolutionLevel 1"

.zarray (json)   .zattrs (json)

| 0.0.0 | 0.0.1 | ... | 0.0.x |
| 0.1.0 | 0.1.1 | ... | 0.1.x |

...etc...

```json
{
    ".zgroup": "{\n    \"zarr_format\": 2\n}",
    ".zattrs": "{\n    \"Conventions\": \"UGRID-0.9.0\n\"}",
    "x/.zattrs": "{\n    \"_ARRAY_DIMENSIONS\": [\n        \"node\"\n ...",
    "x/.zarray": "{\n    \"chunks\": [\n        9228245\n    ],\n    \"compressor\": null,\n
    "x/0": ["s3://bucket/path/file.nc", 294094376, 73825960]
}
```

https://github.com/fsspec/kerchunk

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

6

Scalability

LSM-like synthetic benchmark

1k x 1k x 1k, 100 time points
chunk size: 32 x 32 x 32

http://github.com/ome/bioimage-latency-benchmark

Moore, et al. **OME-NGFF: a next-generation file format for expanding bioimaging data-access strategies.** Nat Methods. 2021 Nov 29. (ahead of print) doi: 10.1038/s41592-021-01326-w PMID: 34845388

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

7

Infrastructure

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

8

Infrastructure

Cloud-Native

Upload

Upload

Conversion

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 updated.

Operations are atomic on chunks.

Proprietary File Formats (PFF)

On-the-fly translation

Analysis clients & libraries

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share
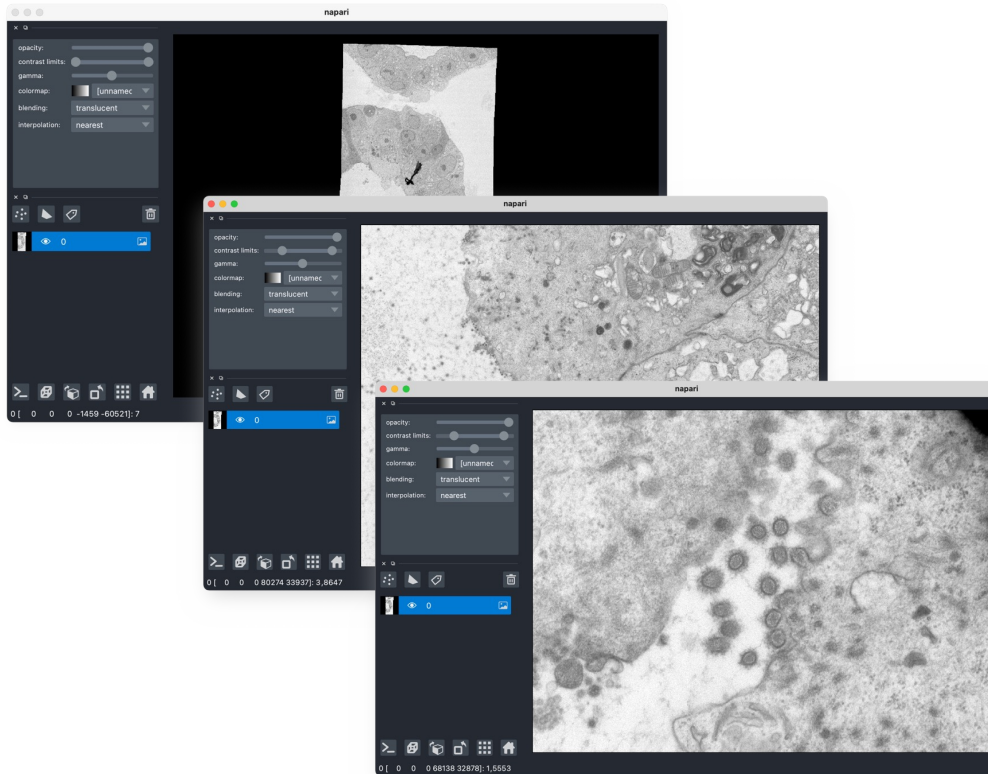
Accessibility

Data: S3 in EBI's Embassy Cloud

Lamers *et al.* (idr0083) **Science (2020)** CC BY 4.0

Highest resolution Image of SARS-CoV-2 (May 2020)
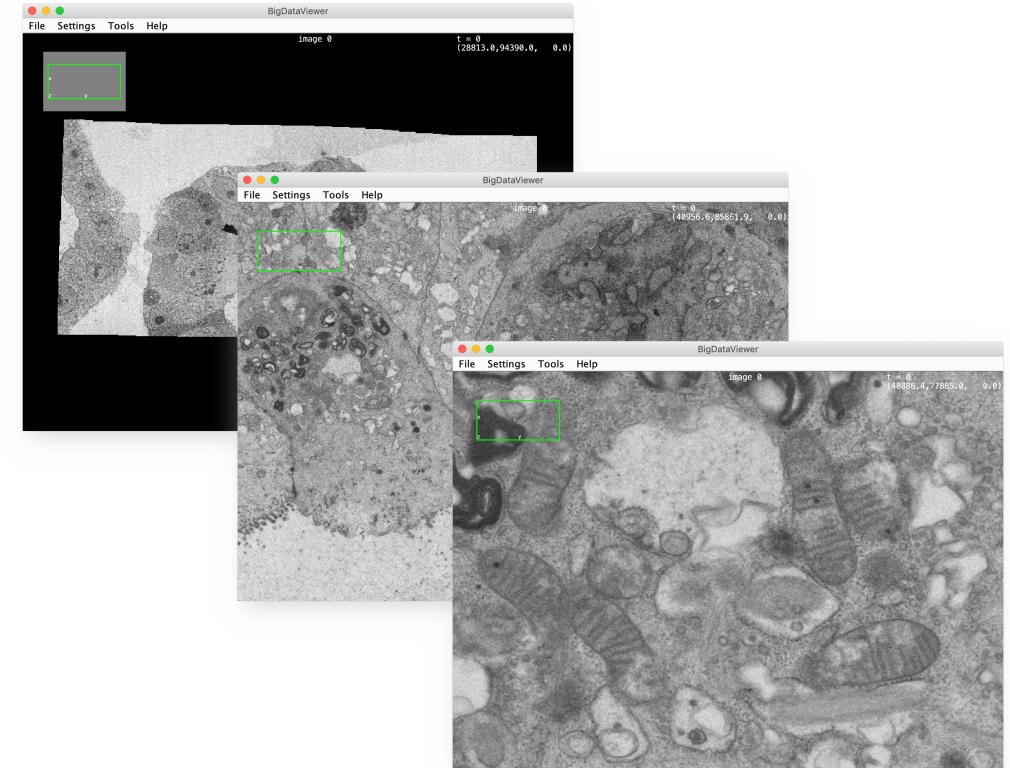
30 GB from EBI's S3 to your browser

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

10

Cross-platform

ome-zarr-py plugin
Will Moore, **U. Dundee** *et al.*
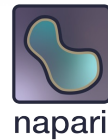
napari

MoBIE Plugin
Christian Tischer, **EMBL Heidelberg**

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

Specifications:

Multiscales   Labels   HCS Plates   Scale   (Transforms)

Process:
Discussions: https://image.sc/tags/ome-ngff
Publication: https://ngff.openmicroscopy.org
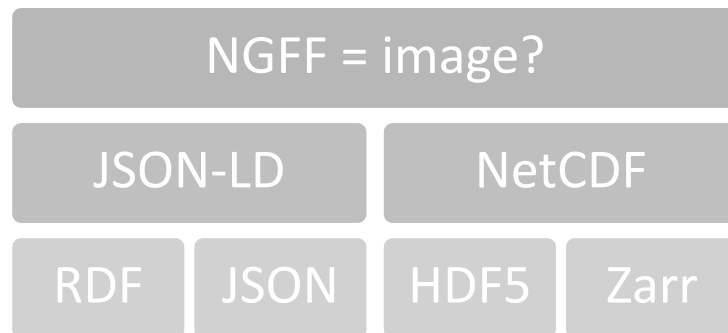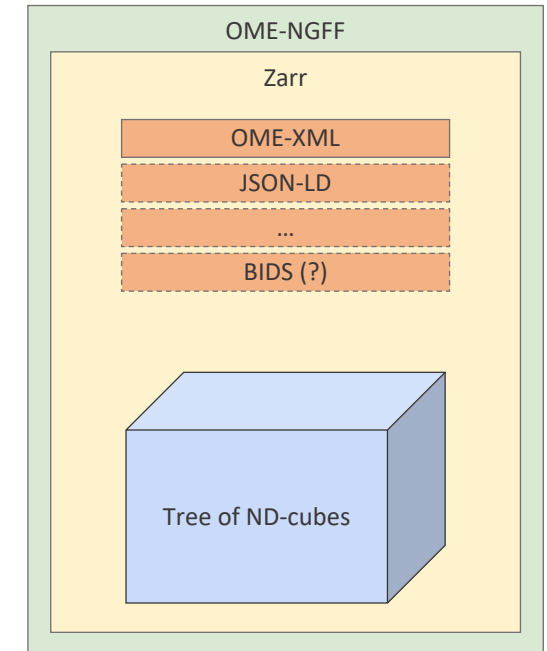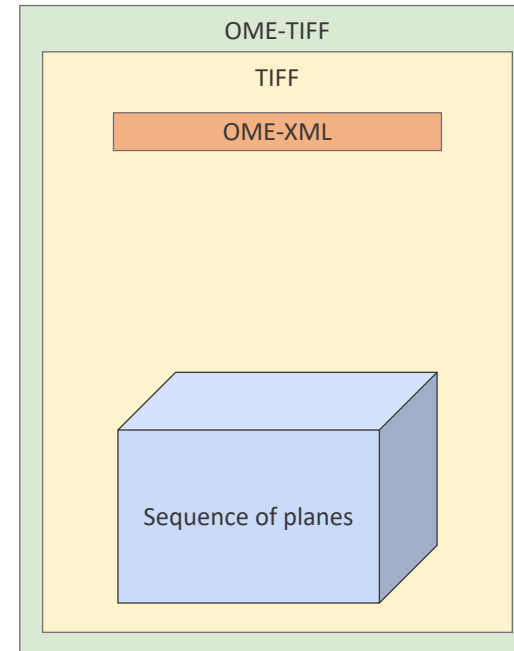Samples: https://uk1s3.embassy.ebi.ac.uk/idr/zarr

Other talking points:

- In our experience, the fewer formats the users need to contend with, the easier the dissemination.

- If we can't agree on a single format, let's try to agree on the highest level of abstraction possible.

- Archival use cases (write-once) are substantially simpler than read-write ones.

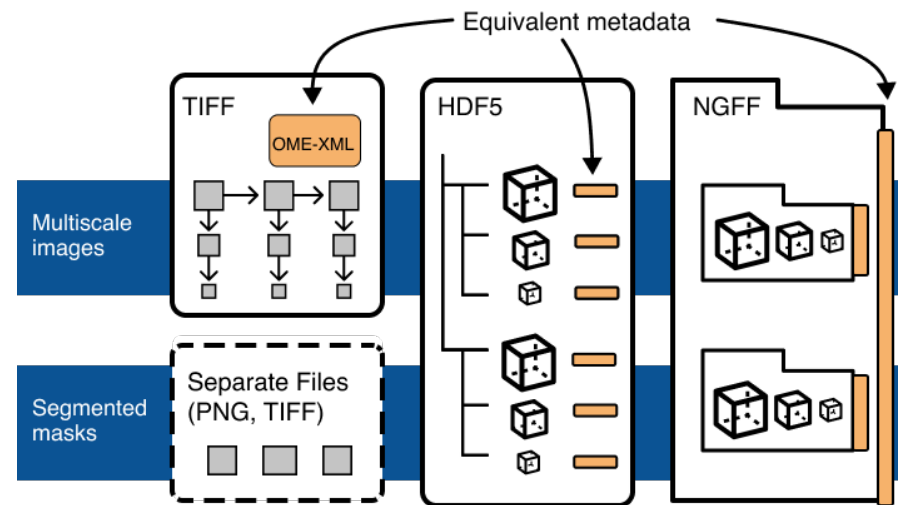- Similarly for public data since S3 security is less than straight-forward.

Josh Moore
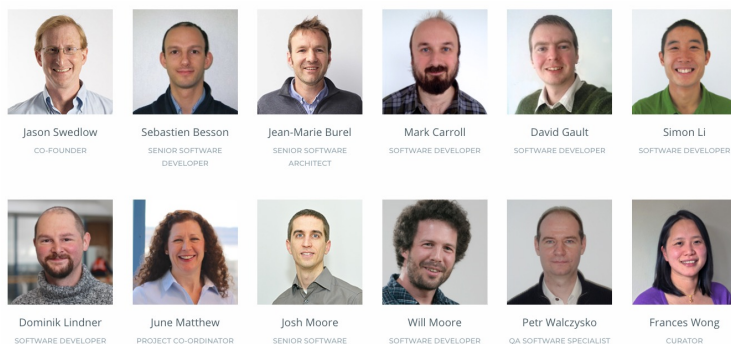DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

12

DANDI - Repository for Neurobio Data

BIDS - Brain Image Data Structure

NGFF - Next Generation File Format - successor to OME-TIFF

ZARR 3.0 - multidimensional array storage

Stack

Stack

Lee Kamentsky

OME-TIFF

TIFF

OME-XML

Sequence of planes

OME-NGFF

Zarr

OME-XML

JSON-LD

...

BIDS (?)

Tree of ND-cubes

NGFF = image?

JSON-LD   NetCDF

RDF   JSON   HDF5   Zarr

Potential connections & avenues:

- Brain (Dandi, BIDS, OpenMINDS, etc.)
- DICOM
- bioschemas.org
- Geo (xarray, HDF5/MET)

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

13

	TIFF	HDF5	NGFF
First release	1986	1998	2016
Maturity (in imaging)	Ubiquitous	Well-supported	Emerging
Base structure	Sequence of 2D planes	Hierarchy of ND arrays	Hierarchy of ND arrays
Multi-file support	With OME metadata	With internal links	Natively
Pyramidal images	With OME metadata	BDV, Imaris	OME-Zarr
Advantages	Tool support	Feature rich format	Simplicity
Limitations	Scalability	Parallel writes	Large number of small files
Ideal use case	Laptop	Powerful workstation, or cluster	Online archive, or public resource

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

14

Former members of the OME team in Dundee

Jason Swedlow — CO-FOUNDER
Sebastien Besson — SENIOR SOFTWARE DEVELOPER
Jean-Marie Burel — SENIOR SOFTWARE ARCHITECT
Mark Carroll — SOFTWARE DEVELOPER
David Gault — SOFTWARE DEVELOPER
Simon Li — SOFTWARE DEVELOPER

Dominik Lindner — SOFTWARE DEVELOPER
June Matthew — PROJECT CO-ORDINATOR
Josh Moore — SENIOR SOFTWARE ARCHITECT
Will Moore — SOFTWARE DEVELOPER
Petr Walczysko — QA SOFTWARE SPECIALIST
Frances Wong — CURATOR

Chris Allan
Gus Ferguson
Kelli Griffis
Roger Leigh
Brian Loranger
Andrew Patterson
Gabriella Rustici
Harald Waxenegger
Wilma Woudenberg

Colin Blackburn
Helen Flynn
Emma Hill
Simone Leo
Scott Loynton
Blazej Pindelski
Aleksandra Tarkowska
Simon Wells

Andrea Falconi
Stefan Frank
Kenny Gillen
Scott Littlewood
Donald MacDonald
Balaji Ramalingam
Joyce Walsh
Eleanor Williams

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

Glencoe Software · Baldock Lab · Bertrand Lab · Brazma Lab
Carazo-Salas Lab · Danuser Lab · Davis Lab · Eliceiri Lab
French Lab · Murphy Lab · Shorte Lab · Zanetti Lab

https://www.openmicroscopy.org/teams

MoBIE Fiji plugin in Java

napari-ome-zarr plugin in Python

Vizarr browser client in Javascript

BDV
Tomancak Lab (MPI-CBG)

KLB
Keller Lab (Janelia)

N5
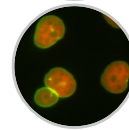Saalfeld Lab (Janelia)
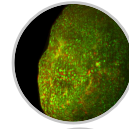
Zarr
Alistair Miles (Oxford)
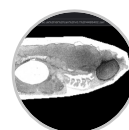
Faas *et al.* **J Cell Biol (2012)**
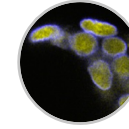idr0053, CC BY-NC-SA 3.0

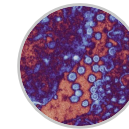Hériché *et al.* **MBoC (2016)**
idr0002, CC-BY 4.0

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

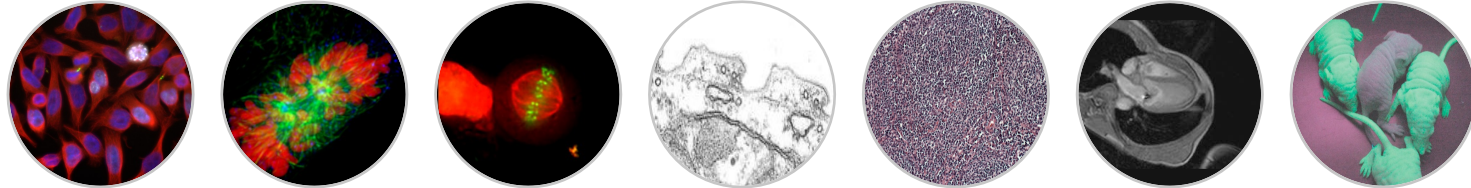Blin *et al.* **PLOS Biology (2019)**
idr0062, CC BY 4.0
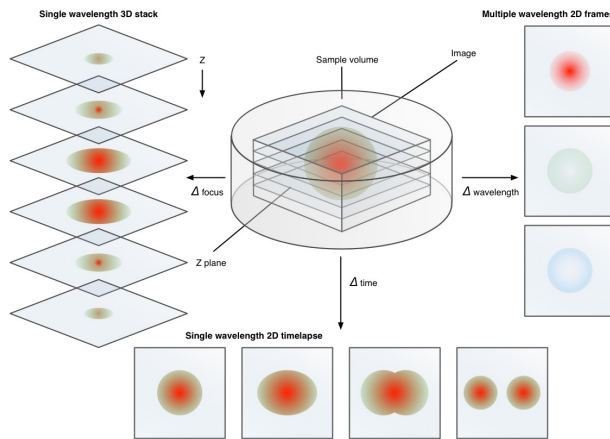
Lamers *et al.* **Science (2020)**
idr0083, CC BY 4.0

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

15

Bioimaging Data Model
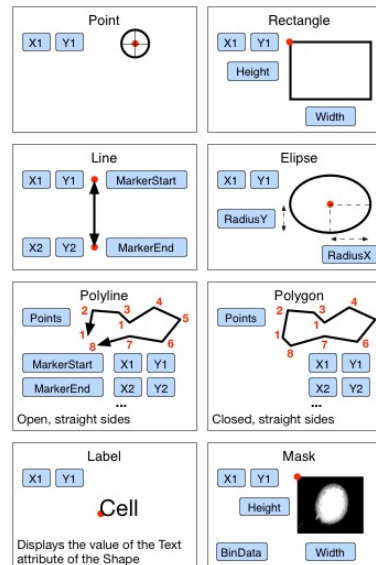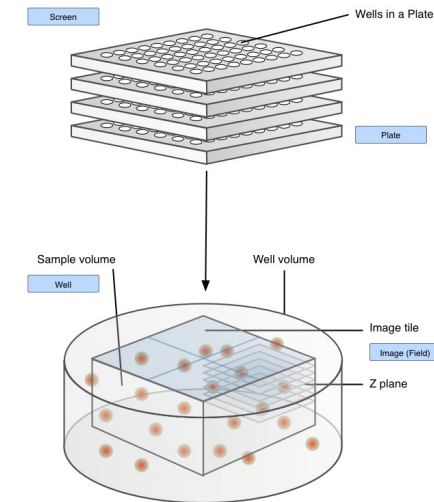

Imaging Modalities

5D Images

3D, multi-color, movies, or any combination thereof

Regions

Collections

High-content screening

Josh Moore
DANDI+BICCN Workshop, March 2022

Open Microscopy Environment
School of Life Sciences, University of Dundee
Dundee, Scotland, UK

University of Dundee

openmicroscopy
Permission to share

20