

Aktuální situace v oblasti manipulace biomedicínských obrazů, nastupující (nejen) cloud-ready Next-Generation-File-Format NGFF OME Zarr

Vladimír Ulman, Ph.D.

10th Dec, 2024

IT4Innovations



Spolufinancováno
Evropskou unií



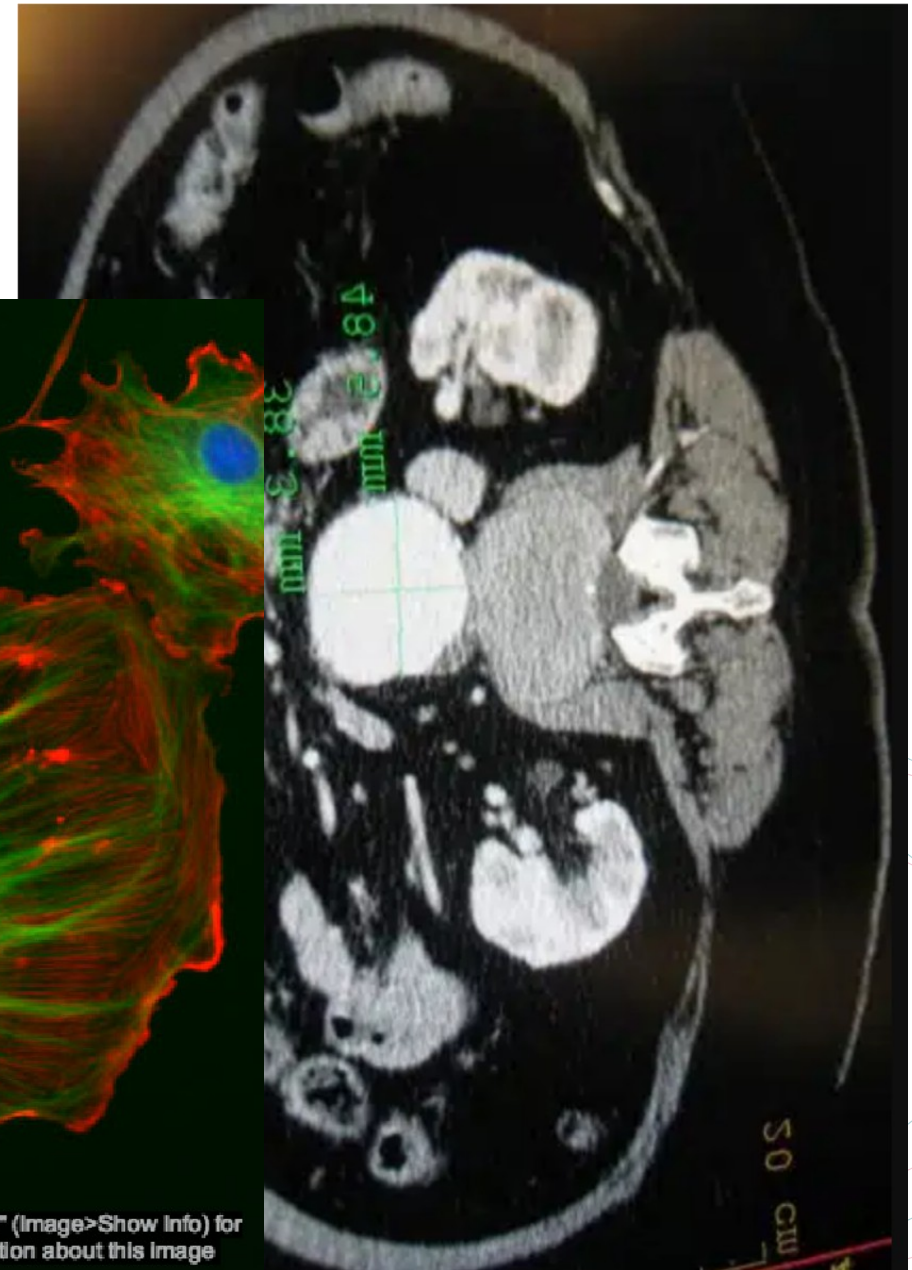
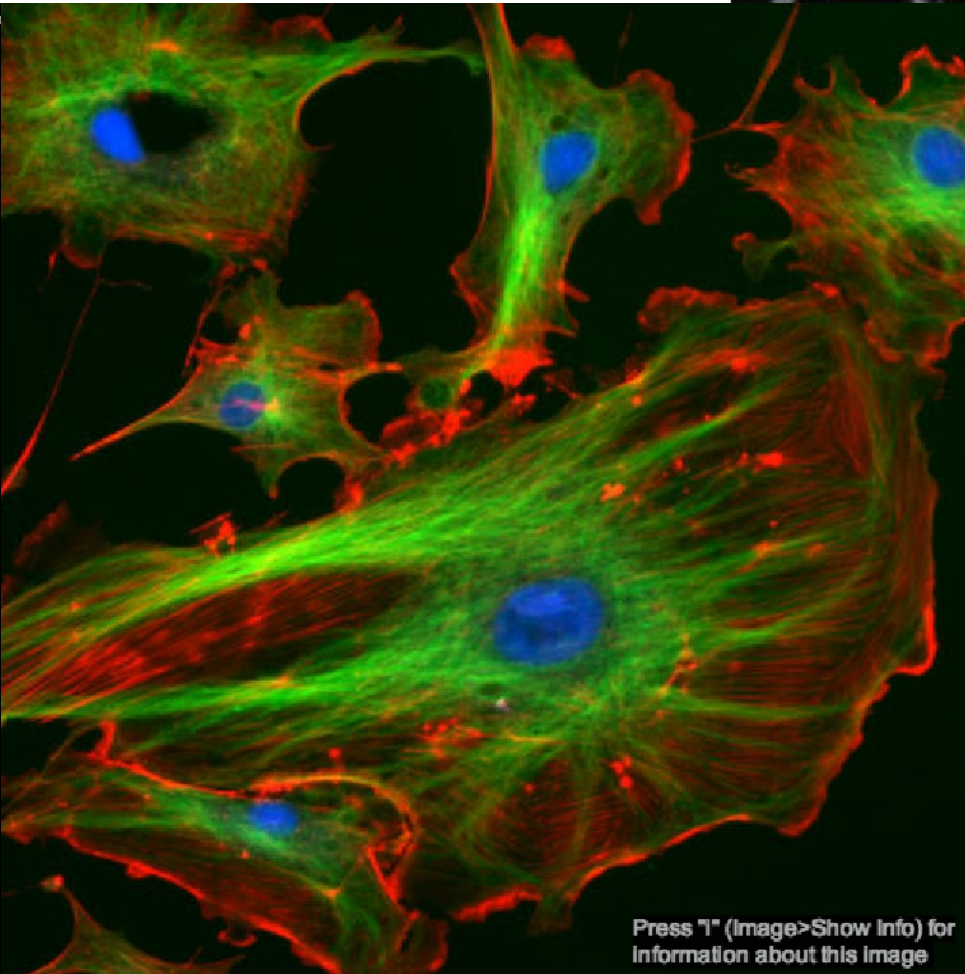
M U N I
I C S

cesnet
.....

VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

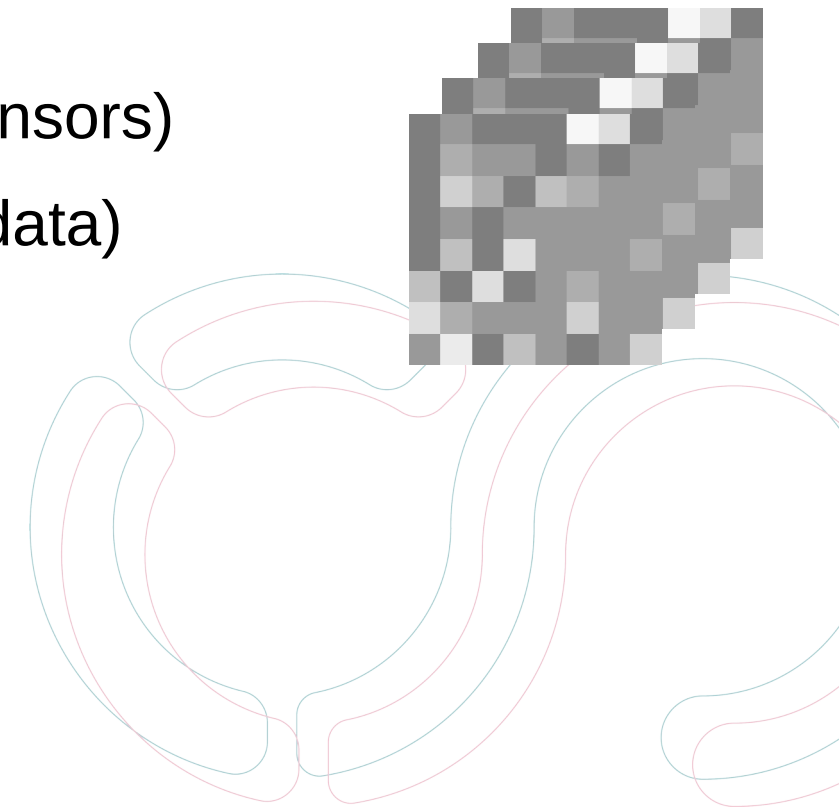
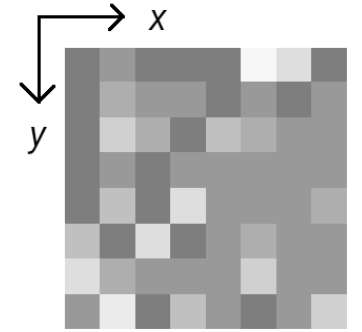
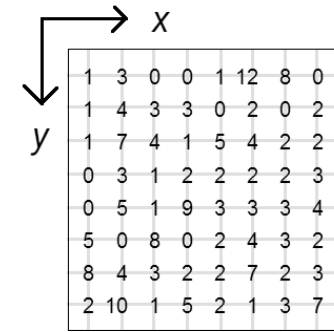
IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM

Biological and Medical Images



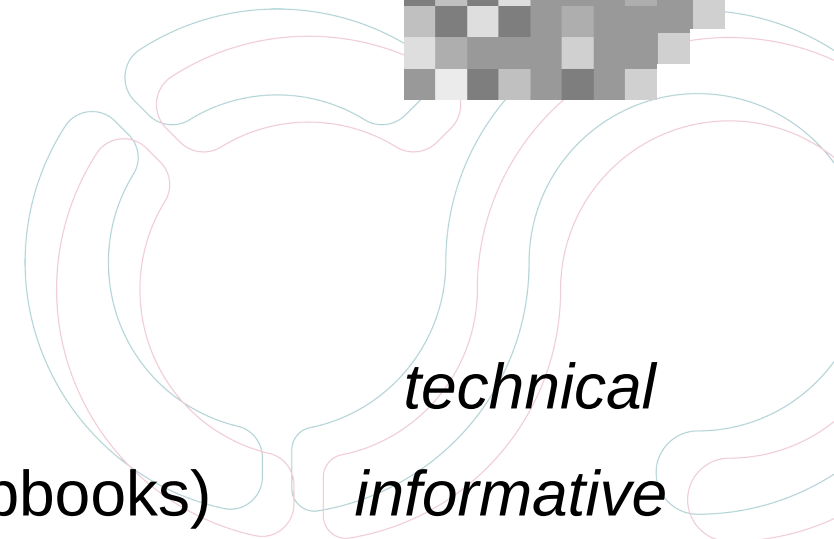
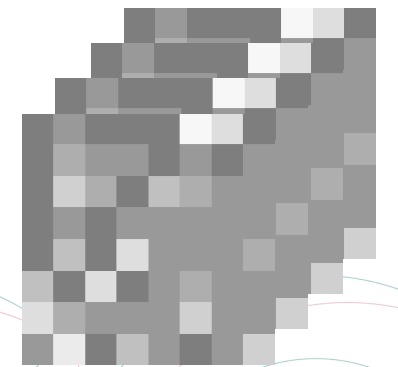
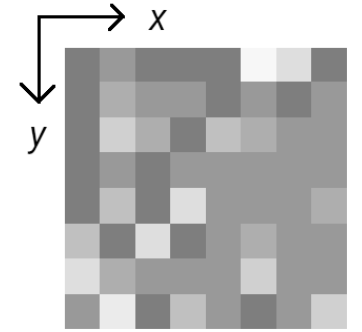
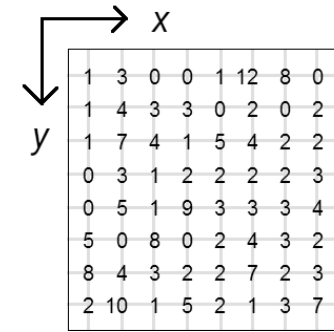
Biological and Medical Images

- Grids of multi-byte scalars per pixel
 - Rarely: Vectors per pixel -- vector fields
- Grid shaped as 2D, 3D, with “time”, with “channels” (Tensors)
 - Grid axes with semantics, with physical units (Metadata)
- Image sizes 1 MB → few TB (PB exists too, very rare)



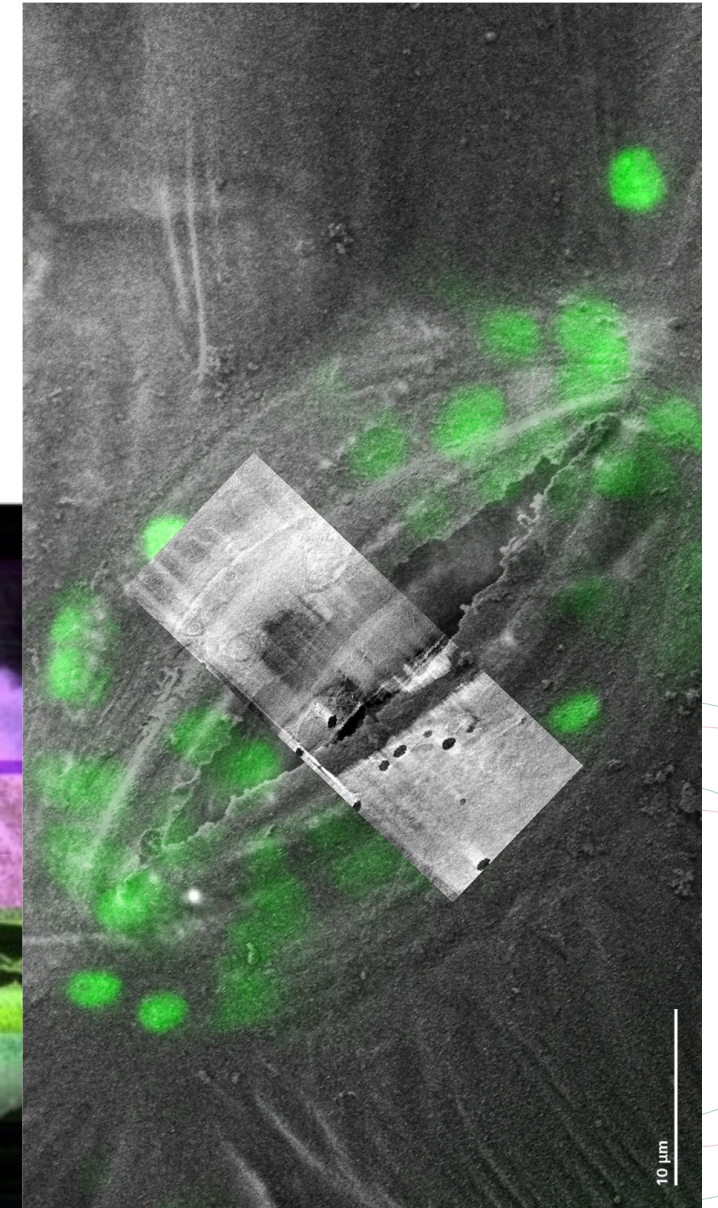
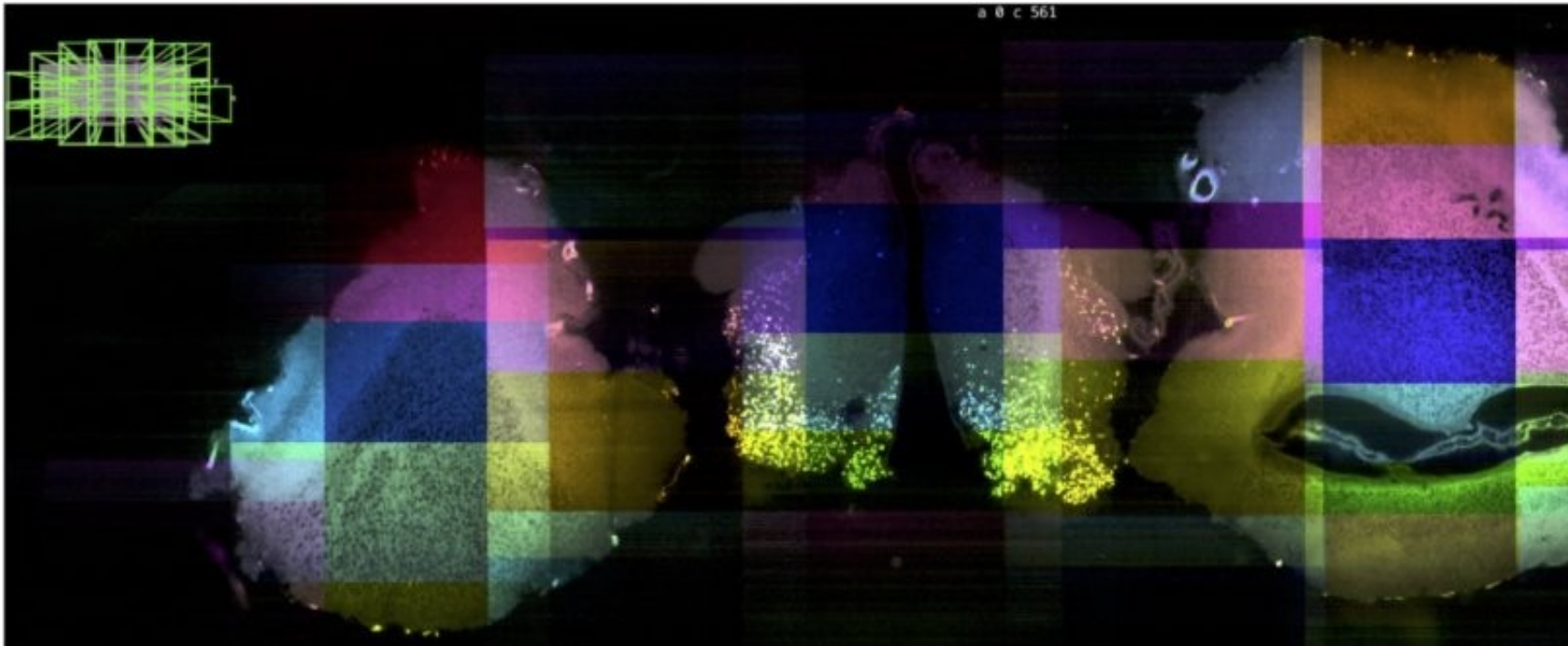
Biological and Medical Images

- Grids of multi-byte scalars per pixel
 - Rarely: Vectors per pixel -- vector fields
- Grid shaped as 2D, 3D, with “time”, with “channels” (Tensors)
 - Grid axes with semantics, with physical units (Metadata)
- Image sizes 1 MB → few TB (PB exists too, very rare)
- Metadata
 - Resolution, view transforms
 - Structured textual annotations (e.g. from electronic labbooks)



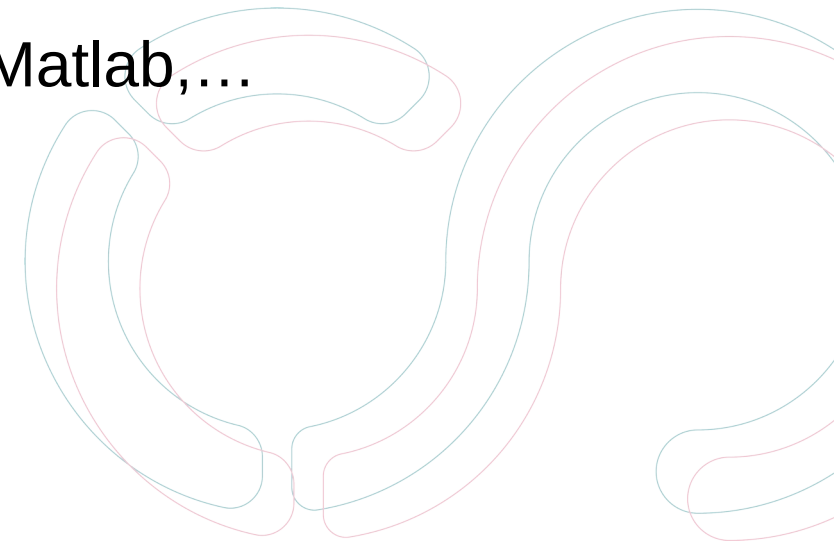
Biological and Medical Images

- Image grids overlaid
- View transforms (incl. pixel resolution)



Software for Processing Images

- Desktop applications
 - Open source: Fiji, Napari, QuPath, CellPose, CellProfiler, Ilastik,...
 - Closed source: Imaris, Zeiss Zen,...
- More coding, less GUI-clicking:
 - Jupyter notebooks, Fiji scripts, scikit-image scripts, Matlab,...
- Workflows:
 - Mostly containers around the above
 - BiaPy, Nextflow, Fractal,...



Software for Processing Images

- Desktop applications
 - Open source: Fiji, Napari, QuPath, CellPose, CellProfiler, Ilastik,...
 - Closed source: Imaris, Zeiss Zen,...
- More coding, less GUI-clicking:
 - Jupyter notebooks, Fiji scripts, scikit-image scripts, Matlab,...
- Workflows:
 - Mostly containers around the above
 - BiaPy, Nextflow, Fractal,...

Python, Java (Java family), little of C++

Where's AI?

→ In each and every one...

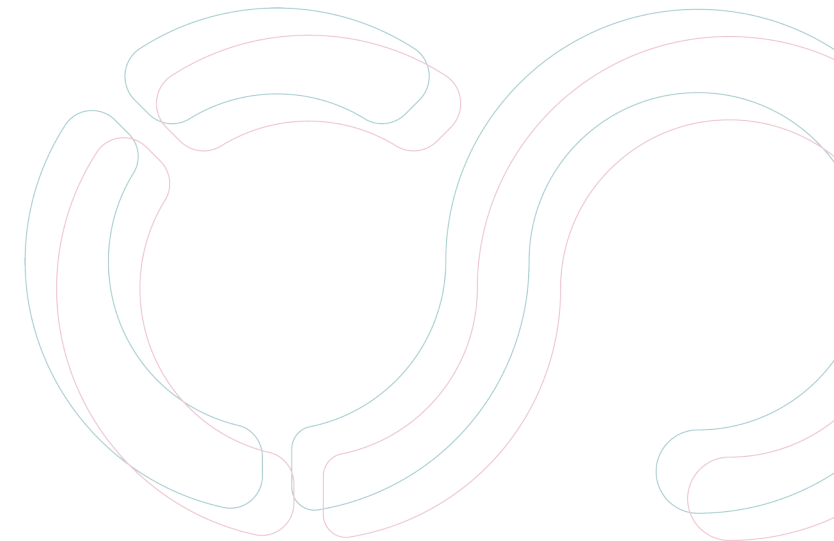
Processing Images

...is a busy Adjust - Test - Eval loop.

- One-time full-data upload

- ROI (re-)training
- Another ROI inference -- testing/evaluation
- Looks good on ROIs?

- Full-data inference & download
- During analysis: Found issues again?
- Publish (somewhat) full-data



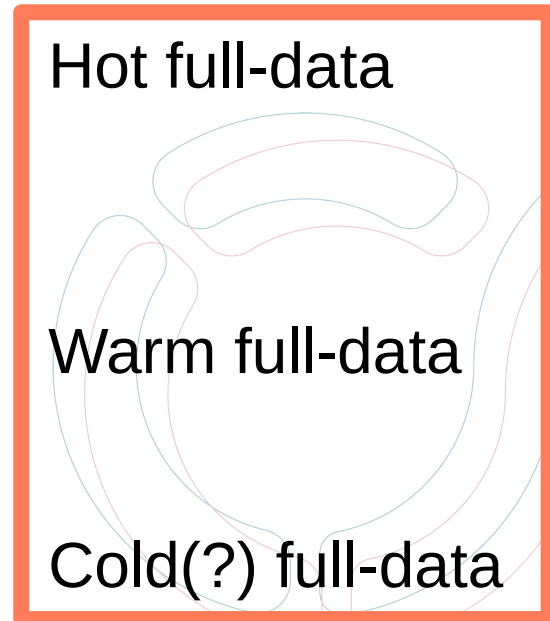
Processing Images

...is a busy Adjust - Test - Eval loop.

- One-time full-data upload

- ROI (re-)training
- Another ROI inference -- testing/evaluation
- Looks good on ROIs?

- Full-data inference & download
- During analysis: Found issues again?
- Publish (somewhat) full-data



Processing Images

...is a busy Adjust - Test - Eval loop.

- One-time full-data upload

- ROI (re-)training
- Another ROI inference -- testing/evaluation
- Looks good on ROIs?

- Full-data inference & download
- During analysis: Found issues again?
- Publish (somewhat) full-data

- Data-parallelism friendly
- Compute-graphs
- “Micro-scheduling”

Hot full-data

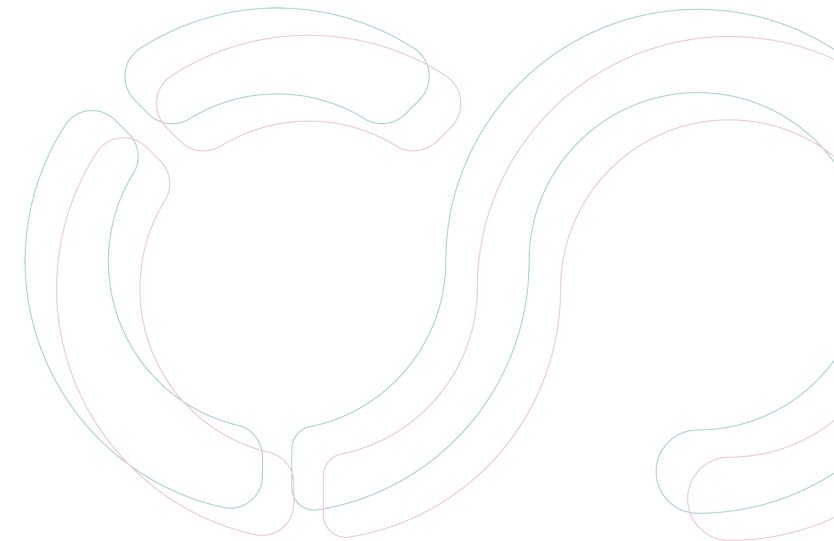
Warm full-data

Cold(?) full-data

Processing Images

...can be cruel to data formats.

- Datasets incrementally grow
 - AI annotations, Results
 - Versions of results
- Quick access to ROIs
 - Locally, remotely
 - Rewritable
 - Addressable



Processing Images

...can be cruel to data formats.

- Datasets incrementally grow
 - AI annotations, Results
 - Versions of results
- Quick access to ROIs
 - Locally, remotely
 - Rewritable
 - Addressable

- TIFF for <2 GB (no ROI)
- TIFFs for <100 GB
- HDF5, DICOM (ROIs easy)
- Rising star: **Zarr**

- TIFF → OME TIFF
- Zarr → OME Zarr

Open Microscopy Environment

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

Storing Images

- OME Zarr, OME NGFF, NGFF Zarr, NGFF (Next Generation File Format)
 - <https://ngff.openmicroscopy.org/latest/>

- Chunk-based + Pyramids
- Metadata specs
- Community-driven

- “Everybody waits for it”

Josh Moore [joshmoore(.bsky.social)]

OME & Zarr steering councils

- Head of RDM at German BioImaging, e.V.
- Funded in Germany to work on Next-Generation File Formats (NGFF)
- Generally would like to see a bioimaging standard.

My main goal for the hackathon:

- Find & help folks who would like to contribute to OME-Zarr and/or a future standard.



(very old photo)

Things I “built”:



Things I help manage:



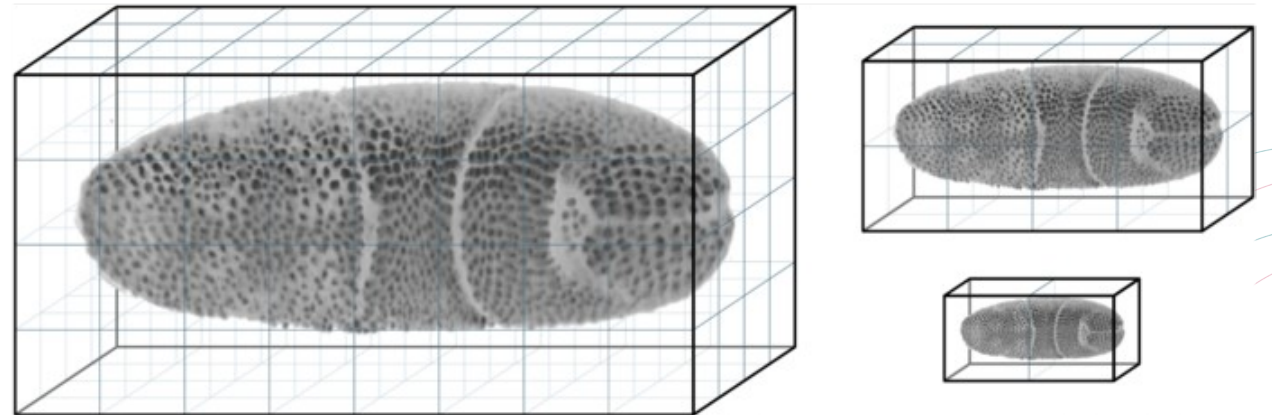
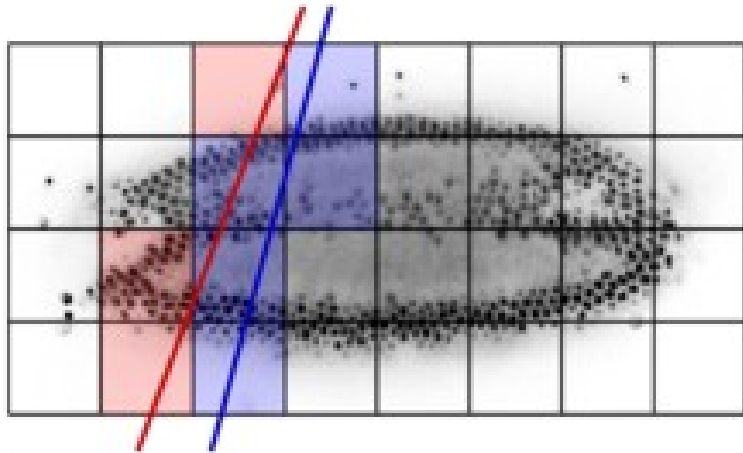
Non-profit representing imaging facilities in Germany



National Research Data Infrastructure consortium for bioimaging

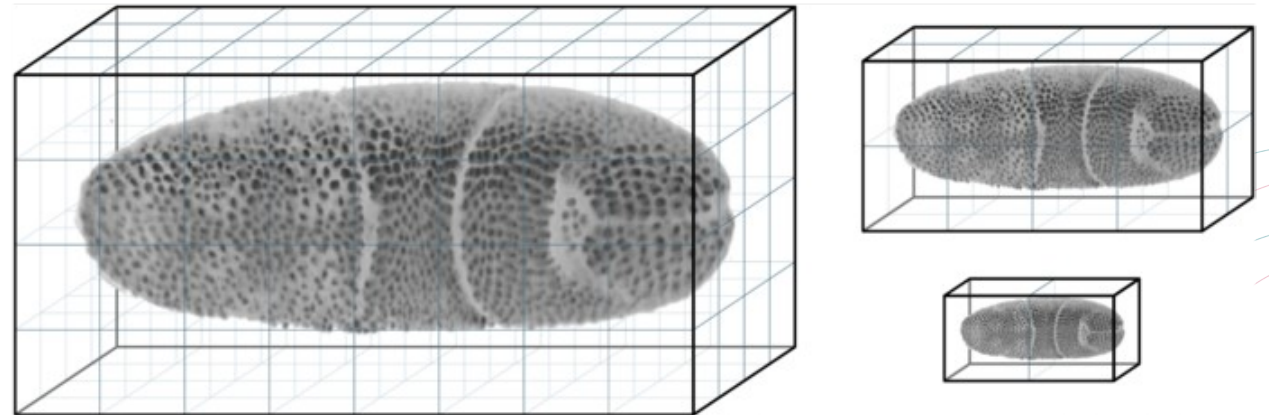
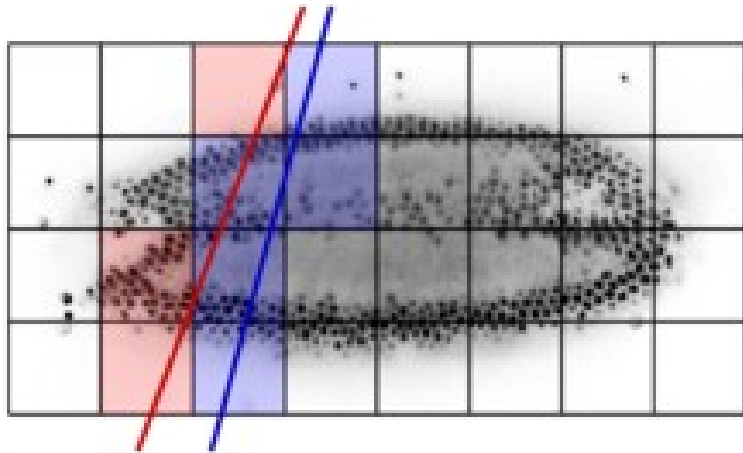
Images with Chunks and Pyramids

- Chunks enable ROIs, Viewing, HPC processing
- Pyramids enable fast-”searching”



Images with Chunks and Pyramids

- Chunks enable ROIs, Viewing, HPC processing
- Pyramids enable fast-”searching”



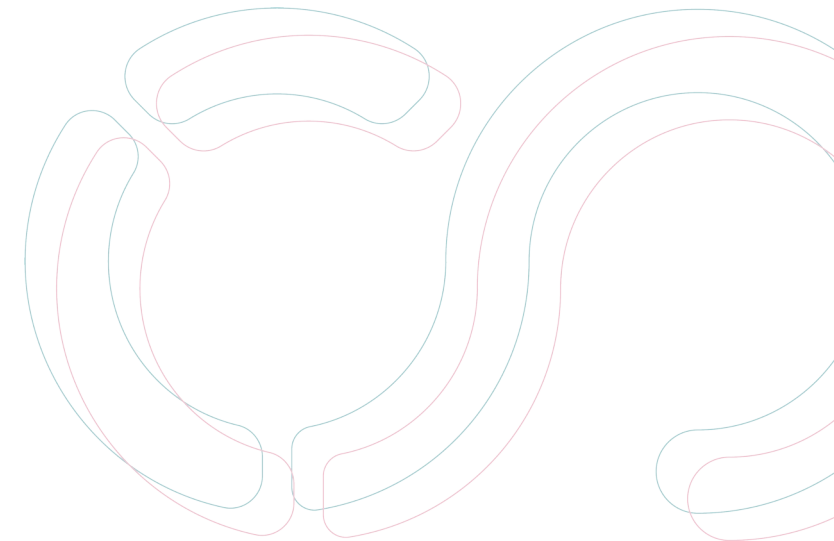
- **Any Key-Value store seems to be good for chunk-based images.**
- Key = Path-like address of a chunk

NGFF OME Zarr

- Started formally in 2018
- 202x: Big buzz
 - Developers wanted to adopt it (and improve it, e.g. “tables story”)
 - Too democratic
- 2024: Development now RFC-based: <https://ngff.openmicroscopy.org/rfc/>
 - To have a formal process for adding new specs/features
- 2025: V1.0 shall be released
 - Companies are waiting for it
 - “Less is more” principle (careful)

NGFF OME Zarr

- Where is it used?
 - IDR image database
 - EMBL EBI archive
 - <https://alleninstitute.org/>
 - <https://www.janelia.org/>
 - Research teams, Individuals
 - Webknossos (Scalableminds)
- CEITEC and IT4Innovations can be part of it too...



NGFF OME Zarr

- Where is it used?
- NGFF hackathon Nov 2024
- Python
- Java



10. prosince 2024

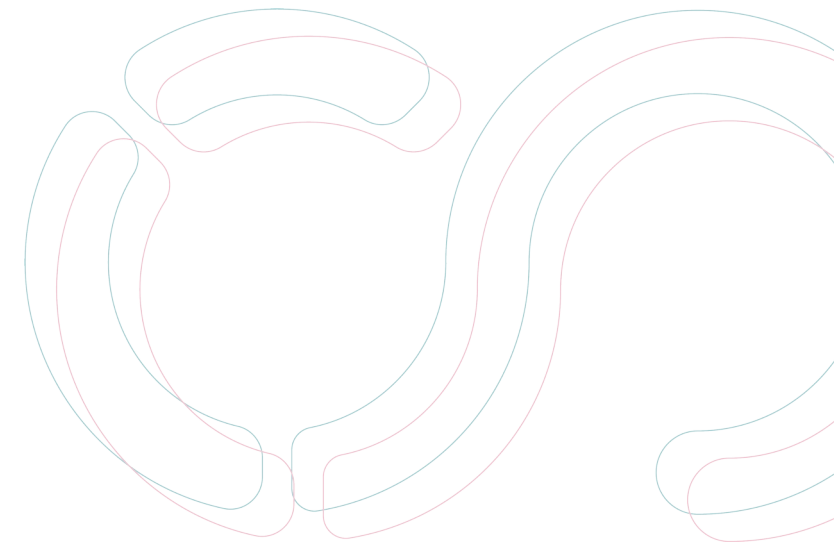
Biolmages, NGFF, Vladimír Ulman

18

NGFF OME Zarr

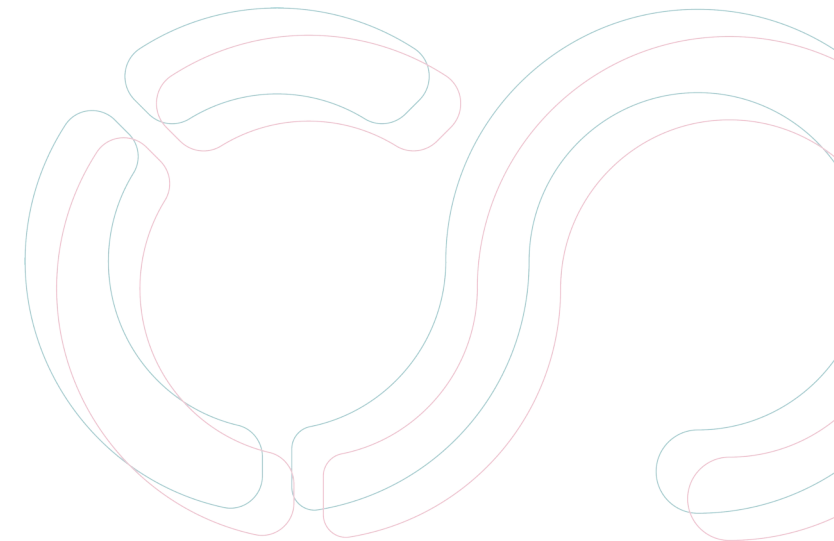
- Where is it used?
 - IDR image database
 - EMBL EBI archive
 - <https://alleninstitute.org/>
 - <https://www.janelia.org/>
 - Research teams, Individuals
 - Webknossos (Scalableminds)
- CEITEC and IT4Innovations can be part of it too...

It's unavoidable, (core) unchangable, unstoppable (at least in biology).



NGFF OME Zarr

- Anything left to be done?
 - RFCs – old and new ones
 - Tooling
 - Especially for Fiji (client side)
 - Integration to “data crunching systems”
 - Full-duplex (remote) chunks transfers
 - Transfer protocols
 - Server-side management



NGFF OME Zarr

- Anything left to be done?
 - RFCs – old and new ones
 - Tooling
 - Especially for Fiji (client side)
 - Integration to “data crunching systems”
 - Full-duplex (remote) chunks transfers
 - Transfer protocols
 - Server-side management

NGFF is only a format:

- Addressing chunks
- Serialization of pixels

Data access today:

- R/W local I/O
- R/O remote via S3

Data access TODO:

- Incremental building
- Desktop Apps

NGFF OME Zarr

- Corner-stones:
 - Make it work with existing tools
 - Python, Java API
 - BigDataViewer (Fiji)
 - Napari viewer
 - Servers:
 - OMERO
 - *TBA (solution for HPC)*

NGFF is only a format:

- Addressing chunks
- Serialization of pixels

Data access today:

- R/W local I/O
- R/O remote via S3

Data access TODO:

- Incremental building
- Desktop Apps

Biological and Medical Images

- 3,4,5D images are normal
- 1 TB sizes normal
- Multi-resolution and multi-view → Metadata important more than ever
- Data-parallelism to start easy, Task-graphs (very soon!) to follow

NGFF, OME Zarr

- Biology is going to adopt it, Medical I'm not sure
- Locally, S3, web browsers
- There's room to do more...



Biological and Medical Images

- 3,4,5D images are normal
- 1 TB sizes normal
- Multi-resolution and multi-view → Metadata important more than ever
- Data-parallelism to start easy, Task-graphs (very soon!) to follow

NGFF, OME Zarr

- Biology is going to adopt it, Medical I'm not sure
- Locally, S3, web browsers
- There's room to do more...

Thank you for your attention!

