BioImage Data Handling and Analysis with OMERO

Workshop in 4 Modules

May 13th, 2024

Trainers: Michele Bortolomeazzi, Riccardo Massei, Christian Schmidt

Support: Lena Krämer & Tom Boissonnet

Programme

Module 1 (9 am - 10.15 am):
Basics of OMERO, data structuring and annotation
(Christian)

Module 2 (10.45 am - 12.45 pm):
OMERO and Fiji
(Michele)

Time to go to Lunch

Module 3 (1.45 pm - 3.45 pm):
OMERO and Jupyter Notebooks
(Riccardo)

Module 4 (4.15 pm - 6. pm):
Publication-ready figures and data with OMERO
(Christian , Riccardo , Michele)

Who are the trainers?

Dr. Riccardo Massei (Helmholtz-Center for Environmental Research, UFZ, Leipzig) - Data Steward for Bioimaging Data in NFDI4BIOIMAGE

Dr. Michele Bortolomeazzi (DKFZ, Single cell Open Lab, bioimage data specialist, bioinformatician, staff scientist in the NFDI4BIOIMAGE project)

Dr. Christian Schmidt (Science Manager for Research Data Management in Bioimaging, DKFZ, Heidelberg, Project Coordinator of the NFDI4BIOIMAGE project)

Introduction and expecations

- Join your nearest neighbors into groups of 3 – 5 people
- Find two things you have all in common
- Collect your expectations for this workshop (excalidraw)

5 min only!

Introduction to OMERO and structured annotations

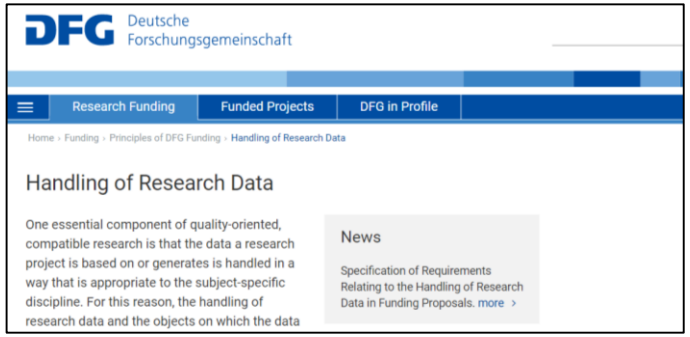
Workshop: **Bioimage data management and analysis with OMERO**

May 13th, 2024, Module 1

Trainers: Michele Bortolomeazzi, Riccardo Massei, **Christian Schmidt**

Significance of research data management

- **Intrinsic motivation**
- Impact and efficiency of my research
- **Extrinsic factors**
- Funding agency demands
- Good Scientific Practice
- **Ethical aspect**
- Make science more sustainable
- Promote *Open Science* and data sharing



- https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html
- https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf
- https://twitter.com/BMBF_Bund/status/1571801906074337280?s=20&t=krDcwOPMuPwjs-VisYBgVg
- https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0347

The FAIR principles for data and data stewardship

Findable **Accessible** **Interoperable** **Reusable**

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

- Wilkinson et al., 2016, Scientific Data, DOI: 10.1038/sdata.2016.18, CC-BY 4.0 (http://creativecommons.org/licenses/by/4.0)
- Jacobsen et al., 2020, FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence, DOI: 10.1162/dint_r_00024

***Guiding* principles**
(*not* a dictation of technical solutions)

Goal is machine-readability
(*not* only human ability to reuse)

Guidelines for data producers *and* publishers

Applies to data and their provenance

FAIR is *not*
- the same as „open“
- a standard

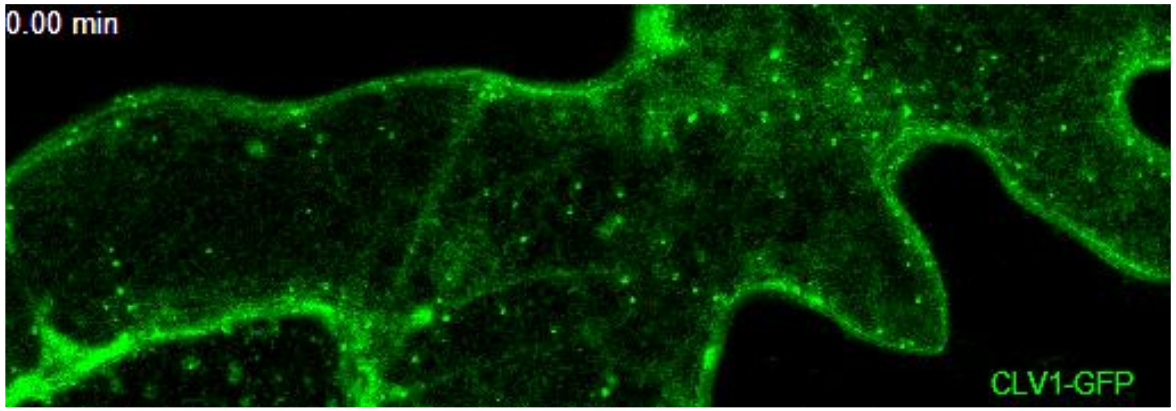
The bioimage data type

Microscopy data is (often):

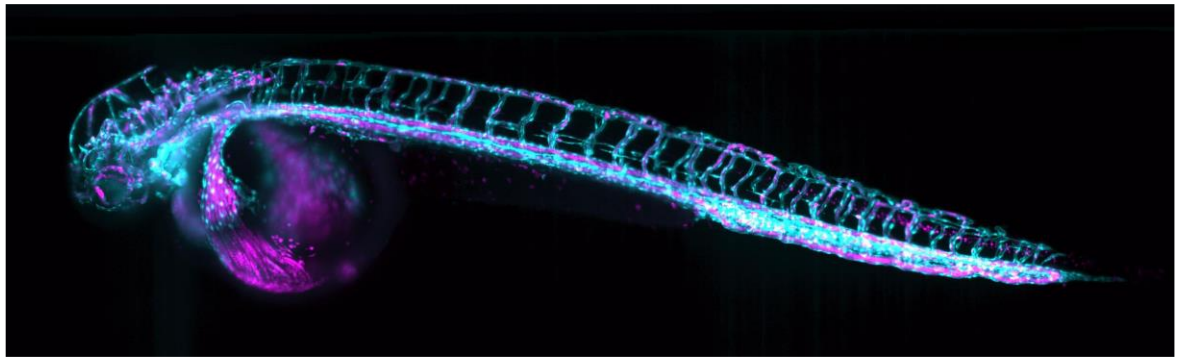
- high-dimensional (X, Y, Z, Channel, Time, …)
- saved in proprietary file formats
- of large file size

  (often in GB-, sometimes in TB-range)
- produced with complex experimental setups
- used for quantitative analysis → derived data

- … i.e. cumbersome to handle, store, and share?

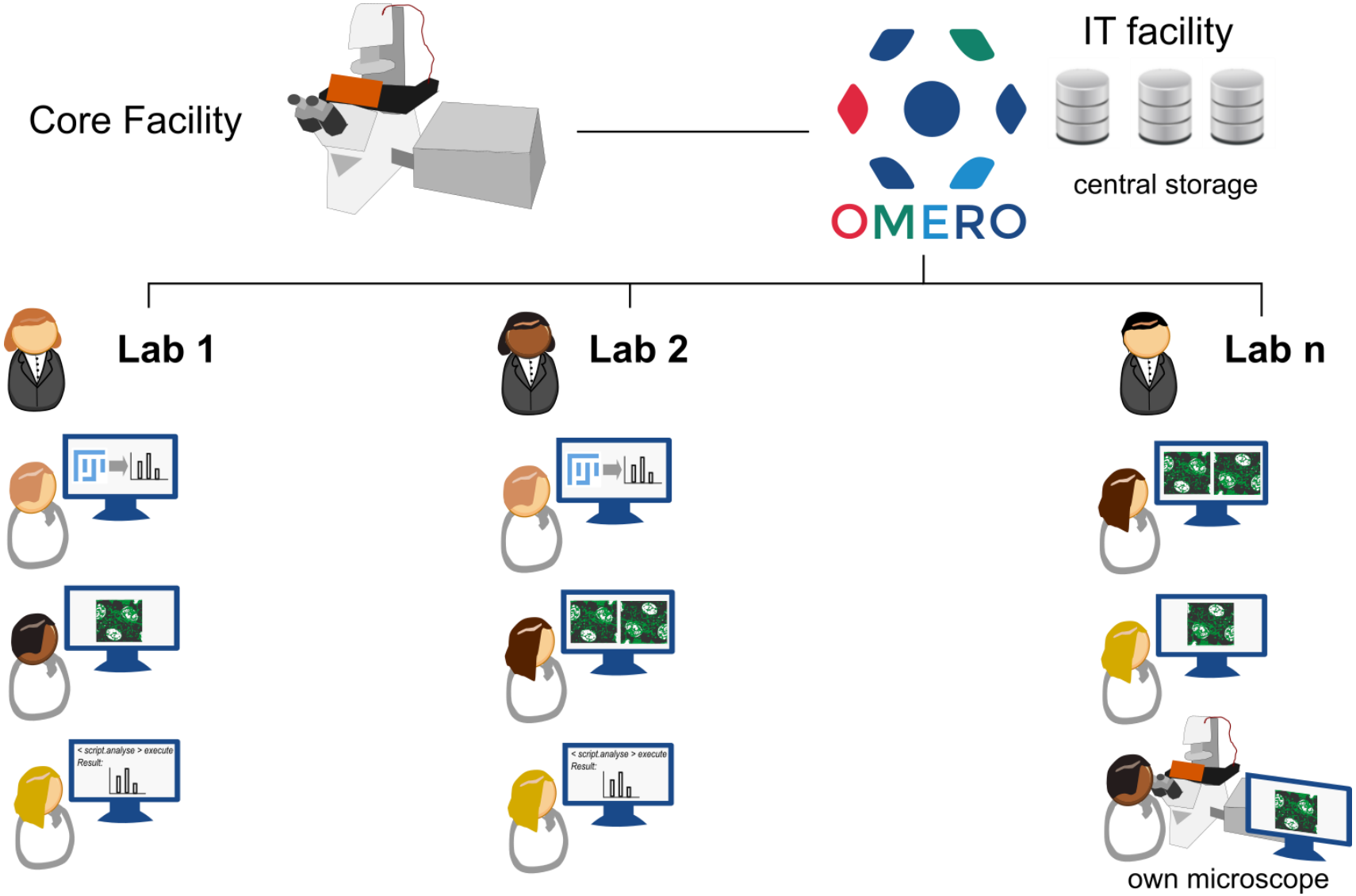


Courtesy of: Y. Stahl, S. Weidtkamp-Peters, HHU Düsseldorf



Courtesy of: Jan Huisken, University of Göttingen

OMERO – What is it?

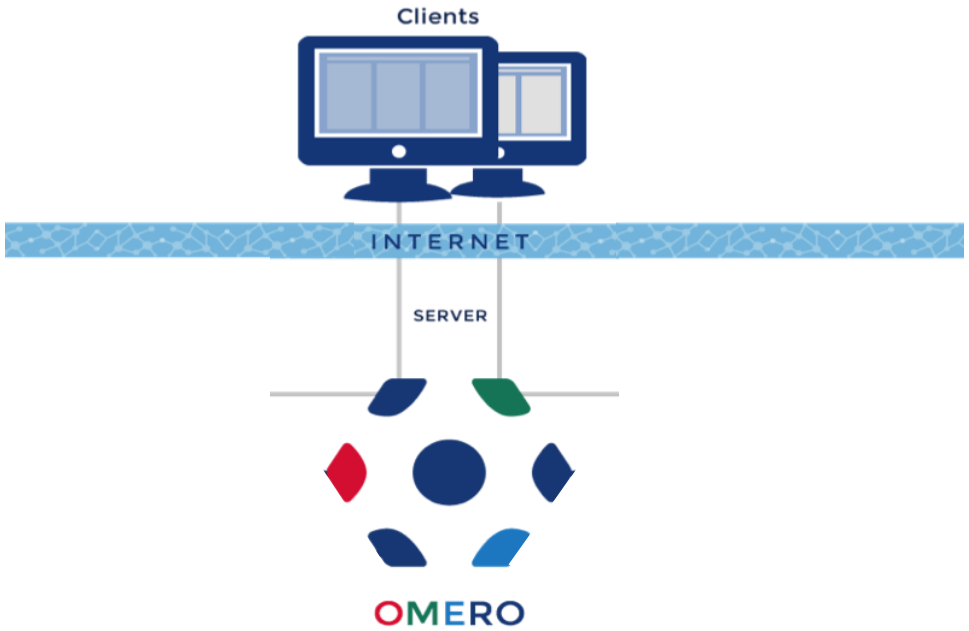


OMERO – What is it?

OMERO: An open-source software for image data management

OMERO = „OME Remote Objects“

Created by the **O**pen **M**icroscopy **E**nvironment Consortium (OME)



- User computer
- Microscope computer
- Processing / analysis server

- Installed on an institute or central IT server
- Storage and handling of imaging data
- Accessible from outside via „clients“

OMERO.web – the main user interface

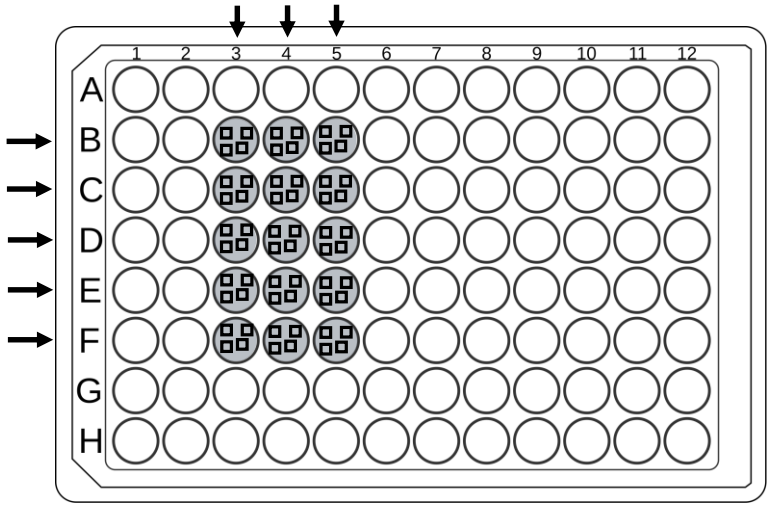
Taken from: Ksenia Krooß, Talk at GFPi Sommertagung, May 17th, 2024

Demonstration and Exercise

Make yourself familiar with OMERO.web

Example Data for this workshop

- A plate acquired at week intervals



- One compound per row
  - Compound B, Compound C, …
- Tested at different concentration (columns):
  - 2 µM, 5 µM, 10 µM
- Multiple images are taken from each well for an accurate sampling.
- Two independent samples are tested (with the sample IDs: 18746 and 22123)

How is this structured correctly?

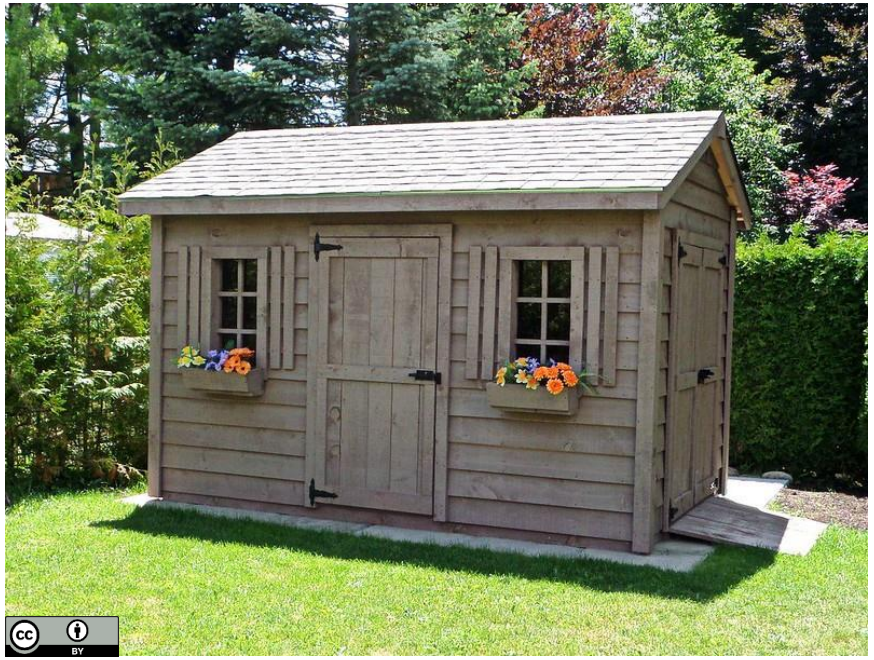
- What should our "dataset unit" be ?
  - the whole week, both plates with all images?
  - One whole plate, one dataset per week ?
  - each compound of every week ?
  - each compound of a single week ?
  - …
- What's a convenient way to describe the data content ?

A little analogy (not so serious)

datasets

images

annotations



By Tom Rossini, CC-BY

By Gordon Jolly, https://flickr.com/photos/loopzilla/8132242517/, CC-BY-SA, https://creativecommons.org/licenses/by-sa/2.0/

A little analogy (not so serious)

datasets

images

Tags

10 µM − week 1 − compound A −

Key-value

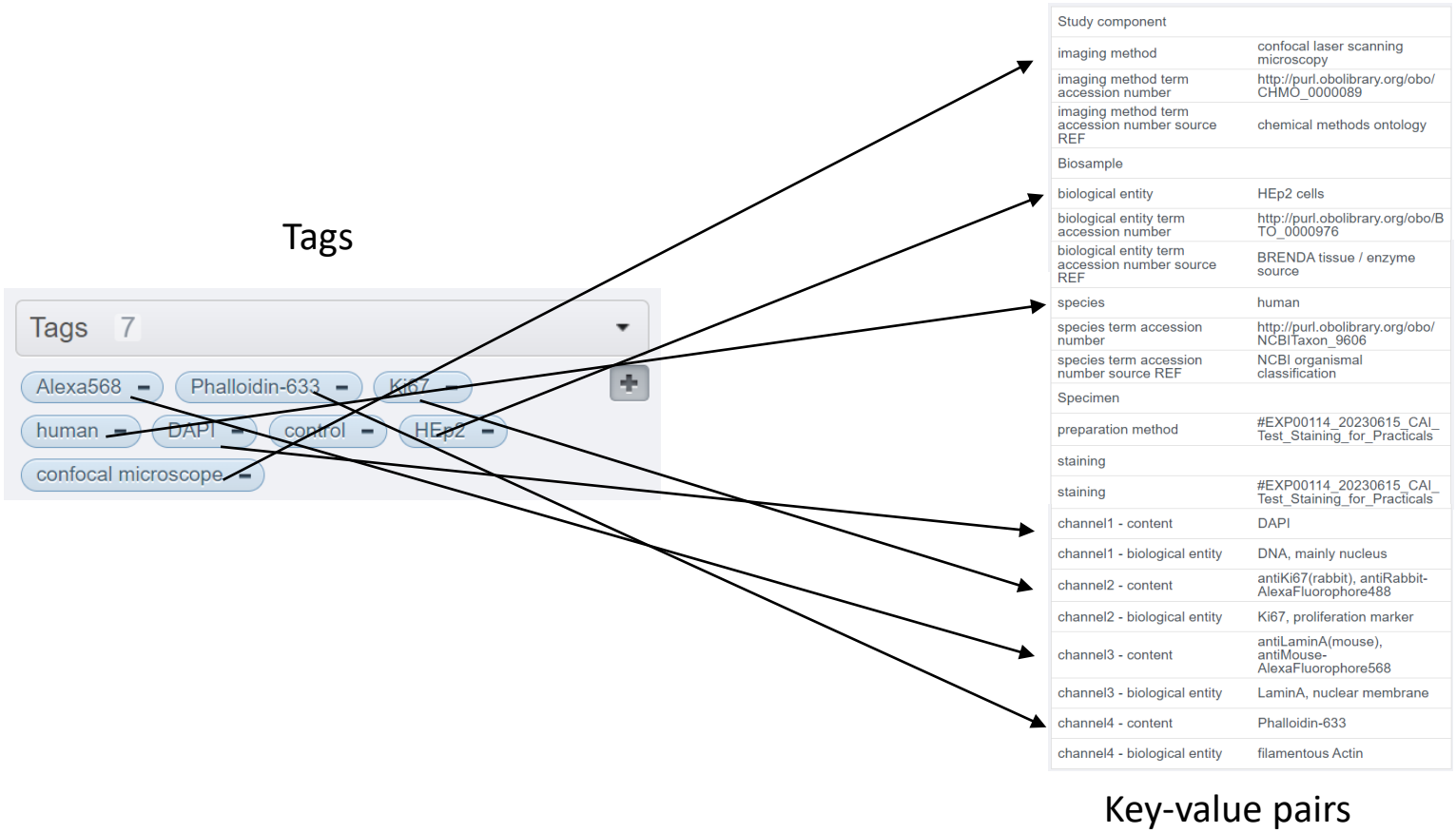
compound	A
concentration	10 µM
week	1

Week1_150607_B02_s1_c1-DAPI_compoundA_10µM.tif

By Tom Rossini, CC-BY

By Gordon Jolly, https://flickr.com/photos/loopzilla/8132242517/, CC-BY-SA, https://creativecommons.org/licenses/by-sa/2.0/

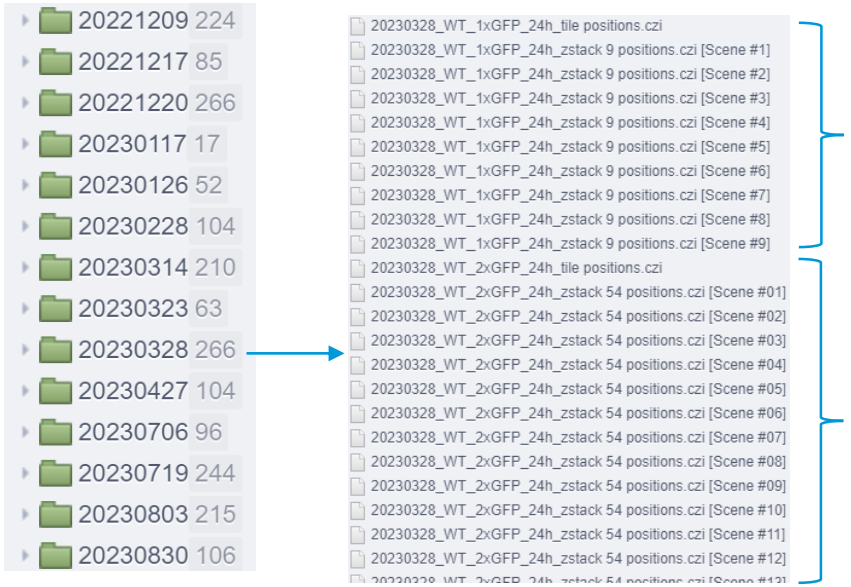
Redundant concepts?



(REMBI checklist)
→ Module 4

Folder hierarchies are what we are used to

- Folder habits are intuitively matched to projects and datasets
- Thinking dataset as folders leads to a confusing structure in OMERO:



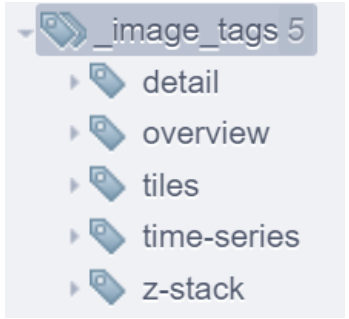
- How should I annotate this?
  - Annotate dataset: unclear to know which images it applies to
  - Annotate first image of every set: other images need reference to the annotated image
  - Annotate every image: tedious and annotation duplication
- What if someone wants to look at the data differently?

Tags are like folders, but better

- We can use tags like folders (not possible with datasets)
- Don't overdo it: less tag categories -> less work

  (and less oversight)
- Tags are flexible and should be used for individual preferences

Need to group images across datasets?
Tags can do that



Are datasets even needed?

- KV-pairs takes care of describing the data, so no need to redo it with tags
- Tags are like folders, but better (multiple tags per object, descriptions)
- Tags can be categorized with Tagsets

-> Good for organizing and filtering

- Tags could emulate my datasets (don't try)

- So why should we even bother with datasets ?

Advantage of metadata annotation

- Datasets can be annotated and used as structure in analysis!
  - Tags
  - Key-value pairs
  - Attachements
  - Ratings
  - Comments

} These will be captured in the Structured Annotation metadata in the OME.XML header of an OME-TIFF

- By annotating a dataset, I can implicitly annotate the images it contains
  - group images of a same experiment in the dataset

  -> less duplication of annotations
  - images can be annotated to give more details

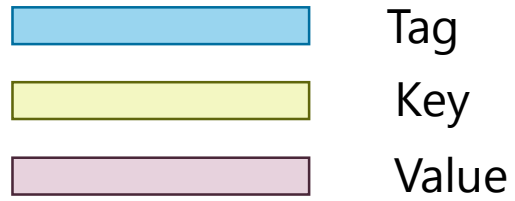
Structured metadata

```
<StructuredAnnotations>
        <TagAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_47706:2998580">
                <Description>tim of 5 µg/ml anti-CD3/anti-CD28 stimulation before staining (in h)</Description>
                <Value>0 h anti-CD3/CD38</Value>
        </TagAnnotation>

        <TagAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_19802:622773">
                <Description>Experiments with staining for phosphorylated ERK protein</Description>
                <Value>phospho-ERK</Value>
        </TagAnnotation>

        <TagAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_19545:561233">
                 <Value>DMSO control</Value>
        </TagAnnotation>

        <MapAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_46089:1280547"
Namespace="openmicroscopy.org/omero/client/mapAnnotation">
                <Value>
                        <M K="Organism">Mus musculus</M>
                        <M K="Strain">C57BL/6</M>
                        <M K="Organ">Spleen</M>
                        <M K="Cell Type">CD4-positive, alpha-beta T cell</M>
                        <M K="Manipulation 1">unstimulated</M>
                        <M K="Treatment 1">0.3% DMSO (control)</M>
                        <M K="Treatment 1 time">0 h</M>
                        <M K="Target 1">phospho-ERK</M>
                        <M K="Staining Method">immunofluorescence indirect</M>
                        <M K="Imaging Method">CLSM</M>
                        <M K="Lab Notebook Ref">CSchm#4 page 112</M>
                </Value>
        </MapAnnotation>
</StructuredAnnotations></OME>
```



Visualizing the OME.XML metadata in Fiji

NFDI 4 BIOIMAGE

```
<StructuredAnnotations>
  <TagAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_47706:2998580">
    <Description>tim of 5 µg/ml anti-CD3/anti-CD28 stimulation before staining (in h)</Description>
    <Value>0 h anti-CD3/CD38</Value>
  <TagAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_19802:622773">
    <Description>Experiments with staining for phosphorylated ERK protein</Description>
    <Value>phospho-ERK</Value>
  <TagAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_19545:561233">
    <Value>DMSO control</Value>
  <MapAnnotation ID="urn:lsid:export.openmicroscopy.org:Annotation:a99ddfbd-769b-474a-8e7b-93bcd175df65_46089:1280547" Namespace="openmicroscopy.org/omero/client/mapAnnotation">
    <Value>
      <M K="Organism">Mus musculus</M>
      <M K="Strain">C57BL/6</M>
      <M K="Organ">Spleen</M>
      <M K="Cell Type">CD4-positive, alpha-beta T cell</M>
      <M K="Manipulation 1">unstimulated</M>
      <M K="Treatment 1">0.3% DMSO (control)</M>
      <M K="Treatment 1 time">0 h</M>
      <M K="Target 1">phospho-ERK</M>
      <M K="Staining Method">immunofluorescence indirect</M>
      <M K="Imaging Method">CLSM</M>
      <M K="Lab Notebook Ref">CSchm#4 page 112</M>
```

Exercise – Bring structure to the data

Enrich the training dataset (structured according to your preference) with Tags (and KV pairs)

Acknowledgments

In cooperation with

Information **I**nfrastructure for Bio**I**mage **D**ata (I3D:bio)

https://www.i3dbio.de/

German Cancer Research Center (DKFZ), Heidelberg
Department Enabling Technology

Dr. Christian Schmidt, Project Coordinator

office@nfdi4bioimage.de

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

In cooperation with

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.