Research Data Management for Bioimage Data at the HHU

Metadata Curation: Key-Value Pairs

Tom Boissonnet

I3D:bio
Information Infrastructure for BioImage Data

Adapted from: Schmidt C., Bortolomeazzi M., Boissonnet T., Fortmann-Grote C. *et al.* (2023). I3D:bio's OMERO training material: Re-usable, adjustable, multi-purpose slides for local user training. Zenodo. DOI: 10.5281/zenodo.8323588

Heinrich Heine Universität Düsseldorf

Metadata details in form of Key-Value Pair annotation

Key-Value Pairs allow (standardized) annotation of detailed metadata

Consists of

- **Key:** Denotes a real-world object or an abstract concept that can be assigned a specific value (of several or many possible values)
- **Value:** Number or text string that specifies the object denoted under „Key“

Examples:

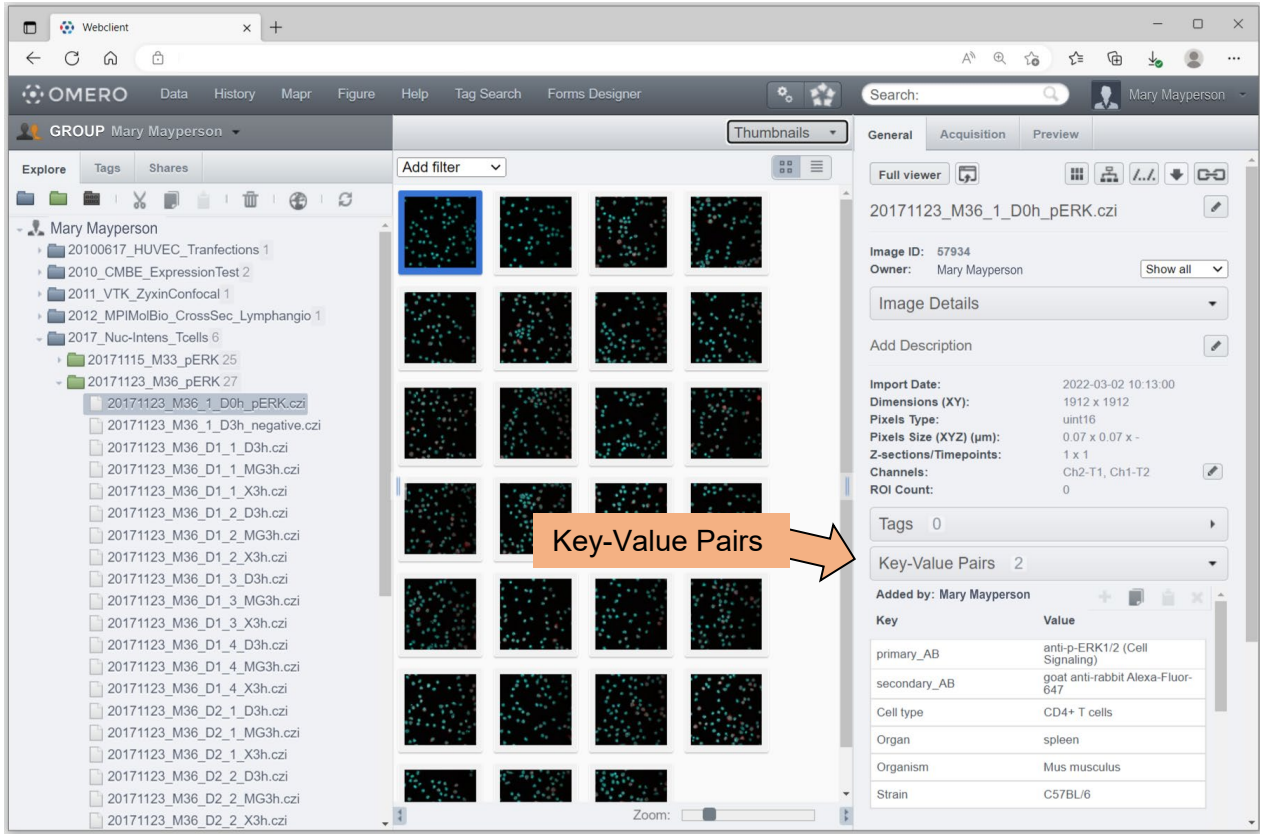
Key: „cell type“ **Value:** „CD4+ T cell“

Key: „disease model“ **Value:** „experimental autoimmune encephalomyelitis“

Key-Value Pairs are part of the metadata (here: in OMERO.web)

Key-Value Pairs can be annotated

- at the Image level
- at the Dataset level
- at the Project level



Options to annotate Key-Value Pairs in OMERO

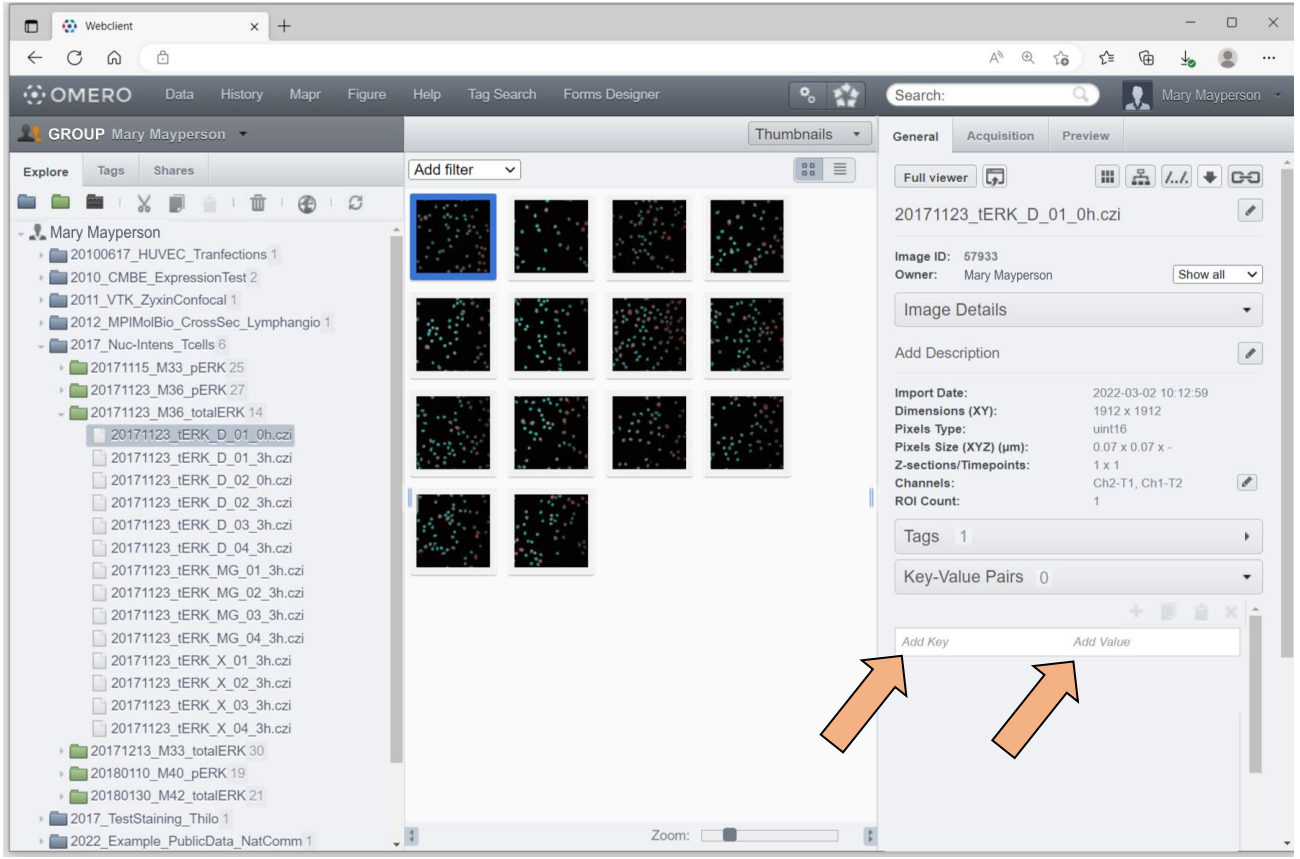
- Manual Key-Value Pair annotation
- Using Bulk Annotation Tools (scripts in OMERO.web, this option is only available if the respective scripts were installed for OMERO.web by the OMERO administrator)
- Using the Metadata Editor Tool OMERO.mde (only during data upload with the OMERO.insight client)

Manual annotation of Key-Value Pairs in OMERO

To add a Key-Value Pair manually, select

- an image
- a collection of images
- a Dataset
- a Project

and fill out the fields under the Key-Value Pairs toggle



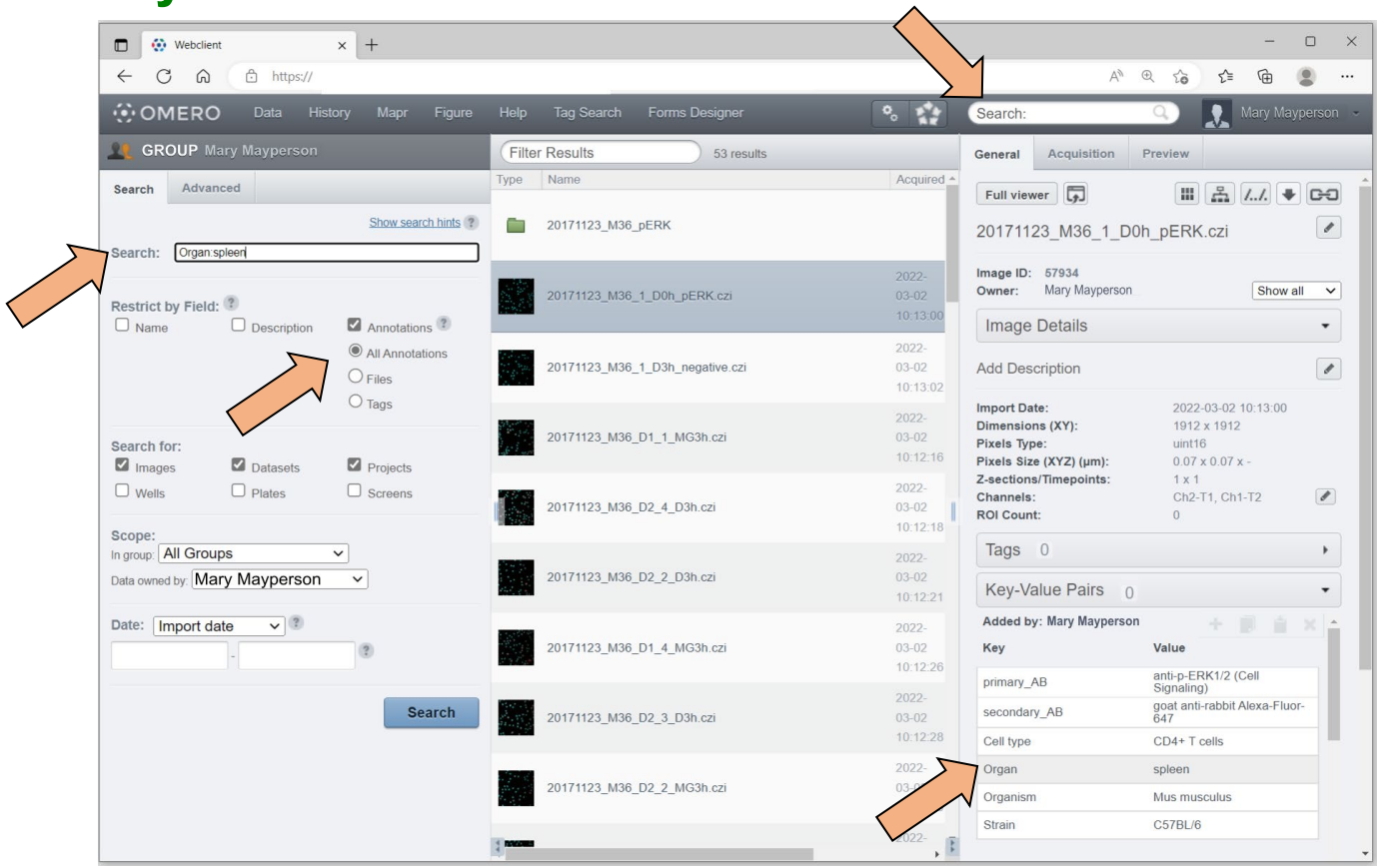
hhu.

Search for a specific Key-Value Pair

Using the combination of

<Key>:<Value>

in the search field allows you to directly search in your data for a specific Key-Value Pair annotation

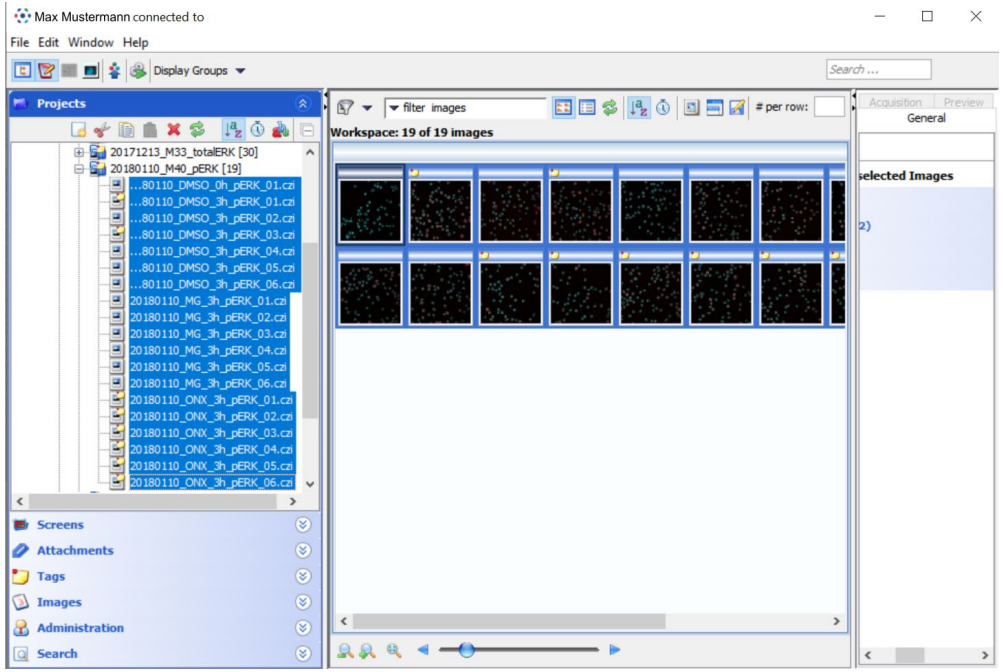


Key-Value Pair enrichment with OMERO Bulk Annotation Tools

Based on an original script by Christian Evenhuis
https://github.com/evenhuis/omero-user-scripts
(Modified by other users‘ contributions, see
https://github.com/ome/omero-scripts for latest version)

Key-Value Pair Annotation with the „KeyVal from csv“ script (1/7)

1) Prepare a table with Keys in row 1. The first Key should be „Image“
2) Select a group of images in OMERO.insight (*not* in OMERO.web)
3) Copy the image names and paste them under the Key „Image“ into a table sheet



	A	B	C	D
1	Image	Organism	Strain	Anatomical structure
2	20180110_DMSO_0h_pERK_01.czi			
3	20180110_DMSO_3h_pERK_01.czi			
4	20180110_DMSO_3h_pERK_02.czi			
5	20180110_DMSO_3h_pERK_03.czi			
6	20180110_DMSO_3h_pERK_04.czi			
7	20180110_DMSO_3h_pERK_05.czi			
8	20180110_DMSO_3h_pERK_06.czi			
9	20180110_MG_3h_pERK_01.czi			
10	20180110_MG_3h_pERK_02.czi			
11	20180110_MG_3h_pERK_03.czi			
12	20180110_MG_3h_pERK_04.czi			
13	20180110_MG_3h_pERK_05.czi			
14	20180110_MG_3h_pERK_06.czi			
15	20180110_ONX_3h_pERK_01.czi			
16	20180110_ONX_3h_pERK_02.czi			
17	20180110_ONX_3h_pERK_03.czi			
18	20180110_ONX_3h_pERK_04.czi			
19	20180110_ONX_3h_pERK_05.czi			
20	20180110_ONX_3h_pERK_06.czi			
21				
22				
23				
24				
25				
26				

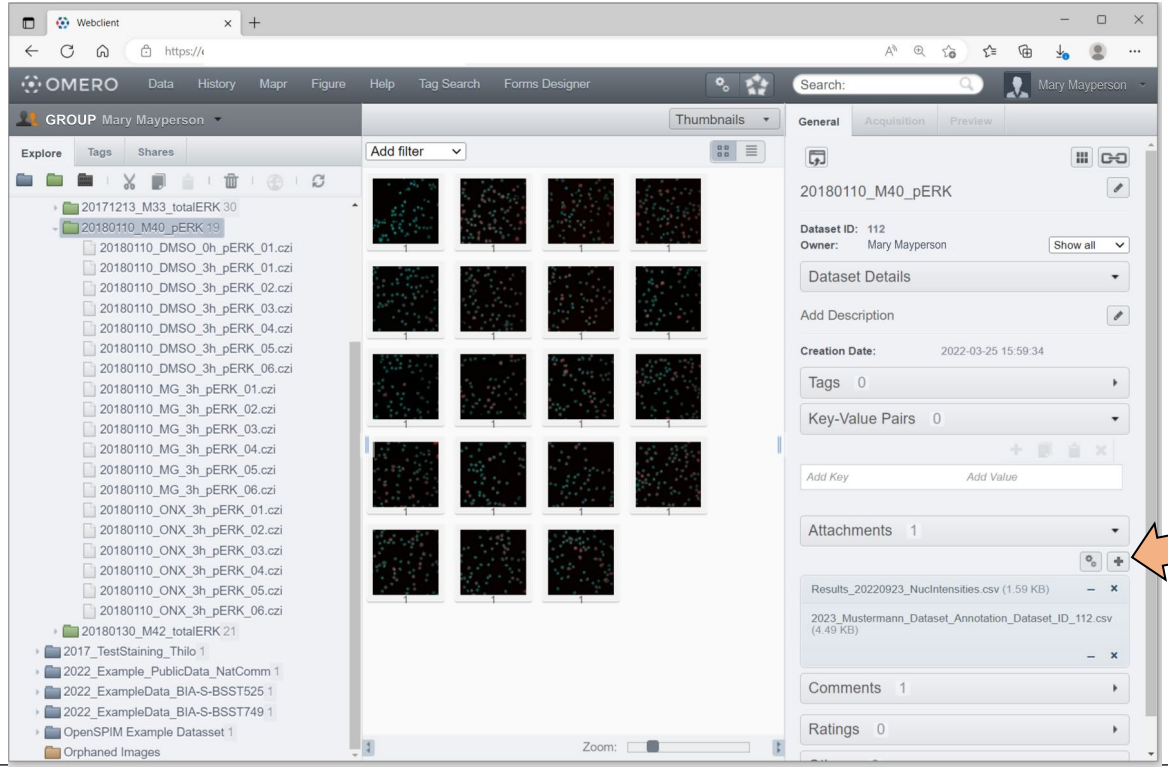
Key-Value Pair Annotation with the „KeyVal from csv“ script (2/7)

4) Fill the Values for each Key and each image as necessary
5) Save the table as CSV (Comma delimited) (*csv)

	A	B	C	D	E	F	G	H	I	J
1	Image	Organism	Strain	Anatomical structure	Cell Type	Cell Activation	Concentrated - Cell Activation	Unit - Concentrated - Cell Activation	Time - Cell Activation	Compound Based Treatment
2	20180110_DMSO_0h_pERK_01.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	unstimulated	0	ug/mL	0 h	polar aprotic solvent
3	20180110_DMSO_3h_pERK_01.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	polar aprotic solvent
4	20180110_DMSO_3h_pERK_02.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	polar aprotic solvent
5	20180110_DMSO_3h_pERK_03.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	polar aprotic solvent
6	20180110_DMSO_3h_pERK_04.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	polar aprotic solvent
7	20180110_DMSO_3h_pERK_05.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	polar aprotic solvent
8	20180110_DMSO_3h_pERK_06.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	polar aprotic solvent
9	20180110_MG_3h_pERK_01.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
10	20180110_MG_3h_pERK_02.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
11	20180110_MG_3h_pERK_03.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
12	20180110_MG_3h_pERK_04.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
13	20180110_MG_3h_pERK_05.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
14	20180110_MG_3h_pERK_06.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
15	20180110_ONX_3h_pERK_01.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
16	20180110_ONX_3h_pERK_02.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
17	20180110_ONX_3h_pERK_03.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
18	20180110_ONX_3h_pERK_04.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
19	20180110_ONX_3h_pERK_05.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
20	20180110_ONX_3h_pERK_06.czi	Mus musculus	C57BL/6	Spleen	CD4-positive, alpha-beta T cell	antibody-mediated anti-CD3 anti-CD28	5, each	ug/mL	3 h	enzyme inhibitor
21										
22										
23										
24										
25										
26										
27										
28										

Key-Value Pair Annotation with the „KeyVal from csv“ script (3/7)

6) Go to the Dataset in OMERO.web

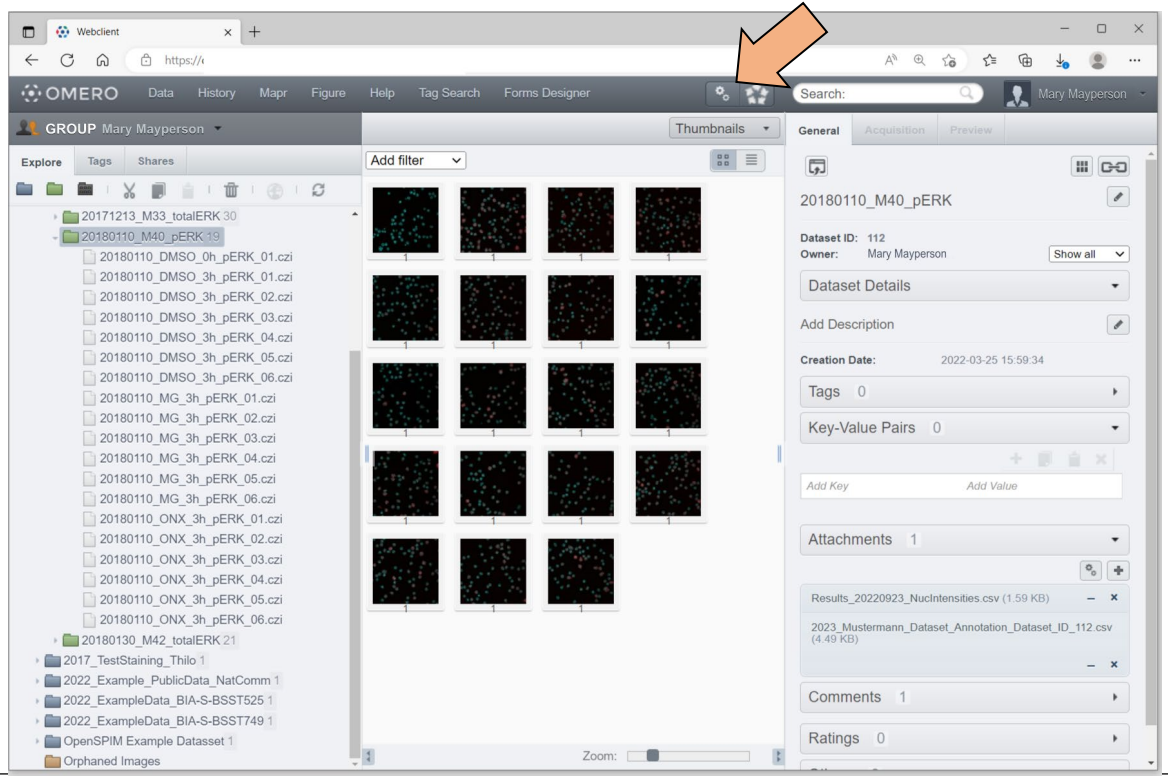


Select the Dataset that contains the images for annotation (do not select an individual image!)

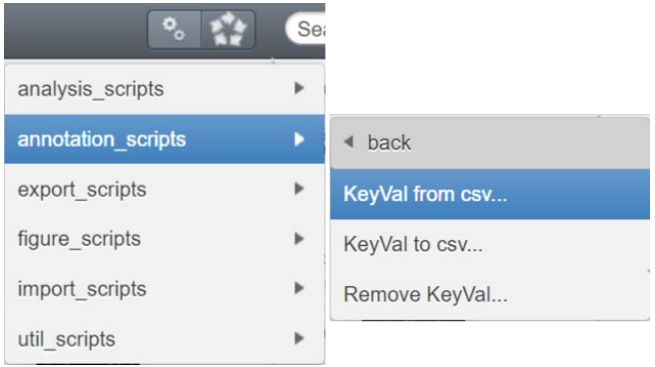
Upload the CSV-table as an attachment to the Dataset

Key-Value Pair Annotation with the „KeyVal from csv“ script (5/7)

8) Go to the Dataset in OMERO.web



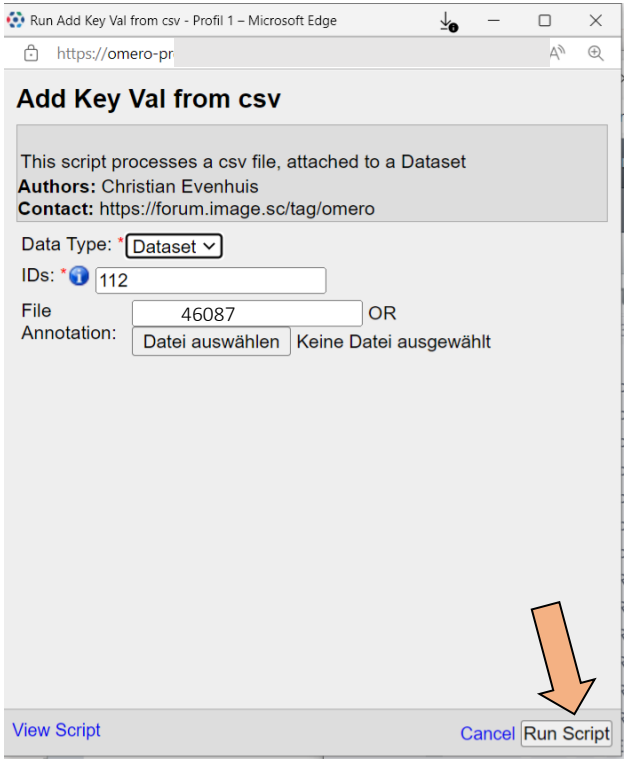
9) Go to the scripts ( ) go to **annotation_scripts** go to **KeyVal from csv…**



Optional:
Mark the table using the icon followed by before step 9

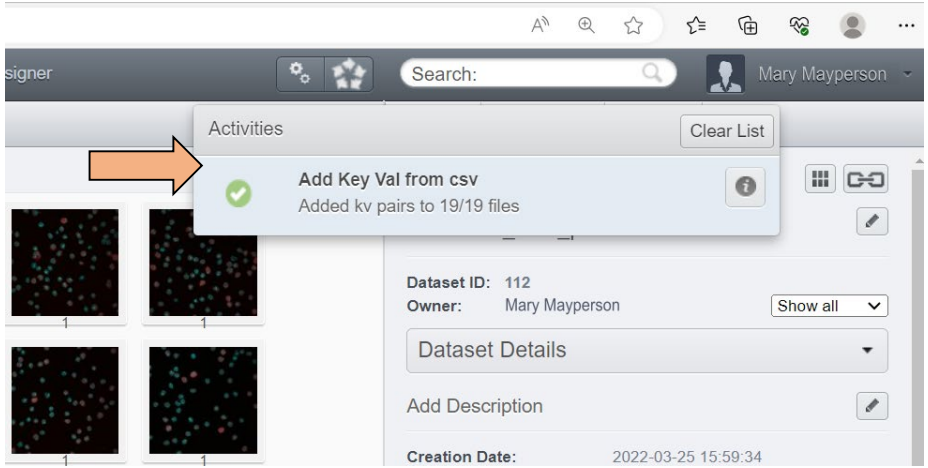
Key-Value Pair Annotation with the „KeyVal from csv“ script (6/7)

10) Enter the File Annotation (**Annotation ID:**) if you have not marked the table (☑)



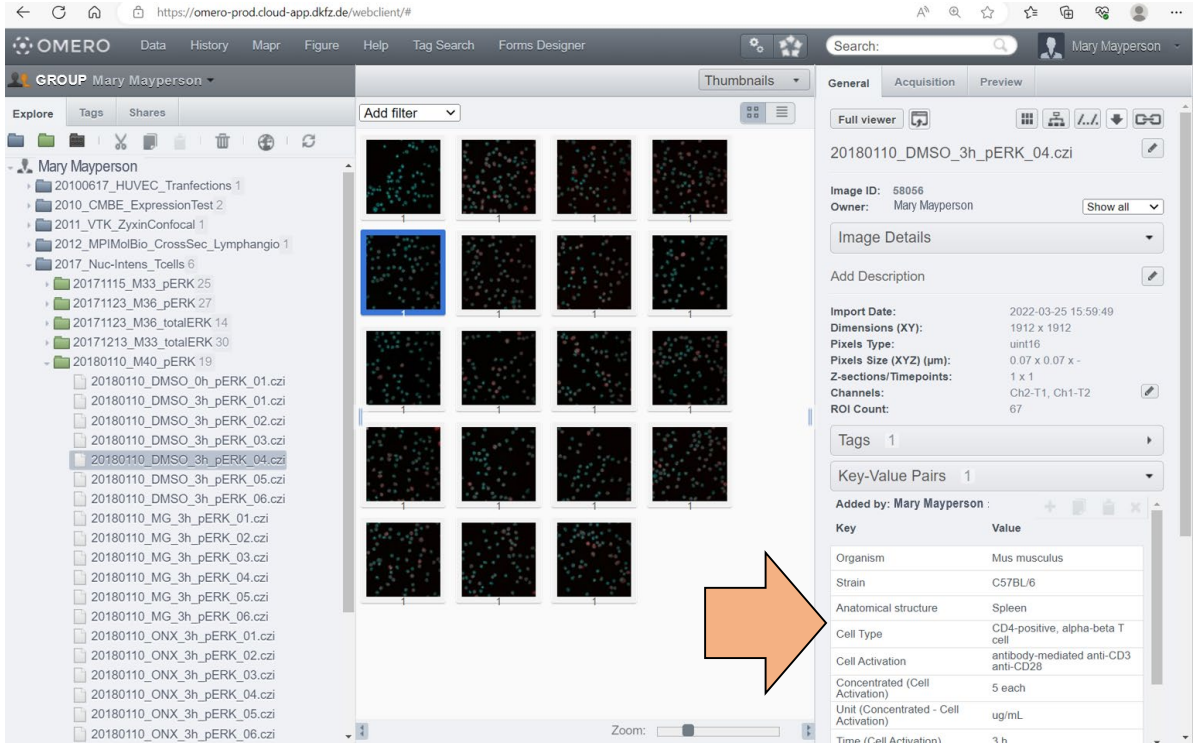
9) Run the script to upload the Annotations

Review the script result:

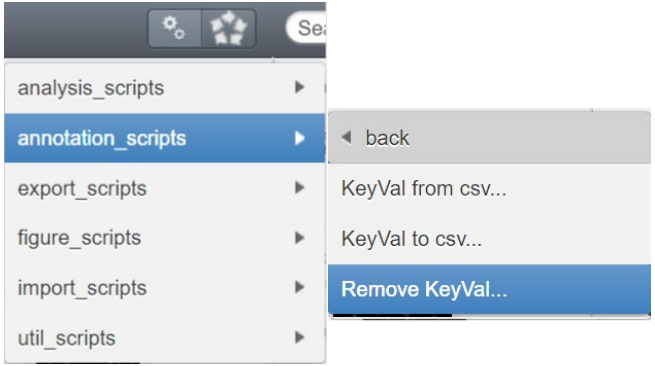


Key-Value Pair Annotation with the „KeyVal from csv“ script (7/7)

11) Check the images for successful Key-Value Pair population



Optional:
You can also remove all the Key-Value Pairs. Mark the files for which KV-Pairs should be deleted and go to:
Scripts
→ annotation_scripts
→ Remove KeyVal…

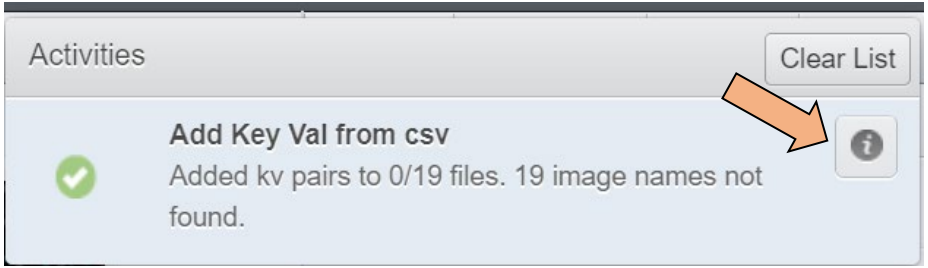


Key-Value Pair Annotation with the „KeyVal from csv“ script NOTE!

The script for Key-Value-Pair Annotation from csv must be installed for your OMERO instance. Please consult with your OMERO administrator if missing

Important:

Sometimes, the „KeyVal from csv“ script may fail:

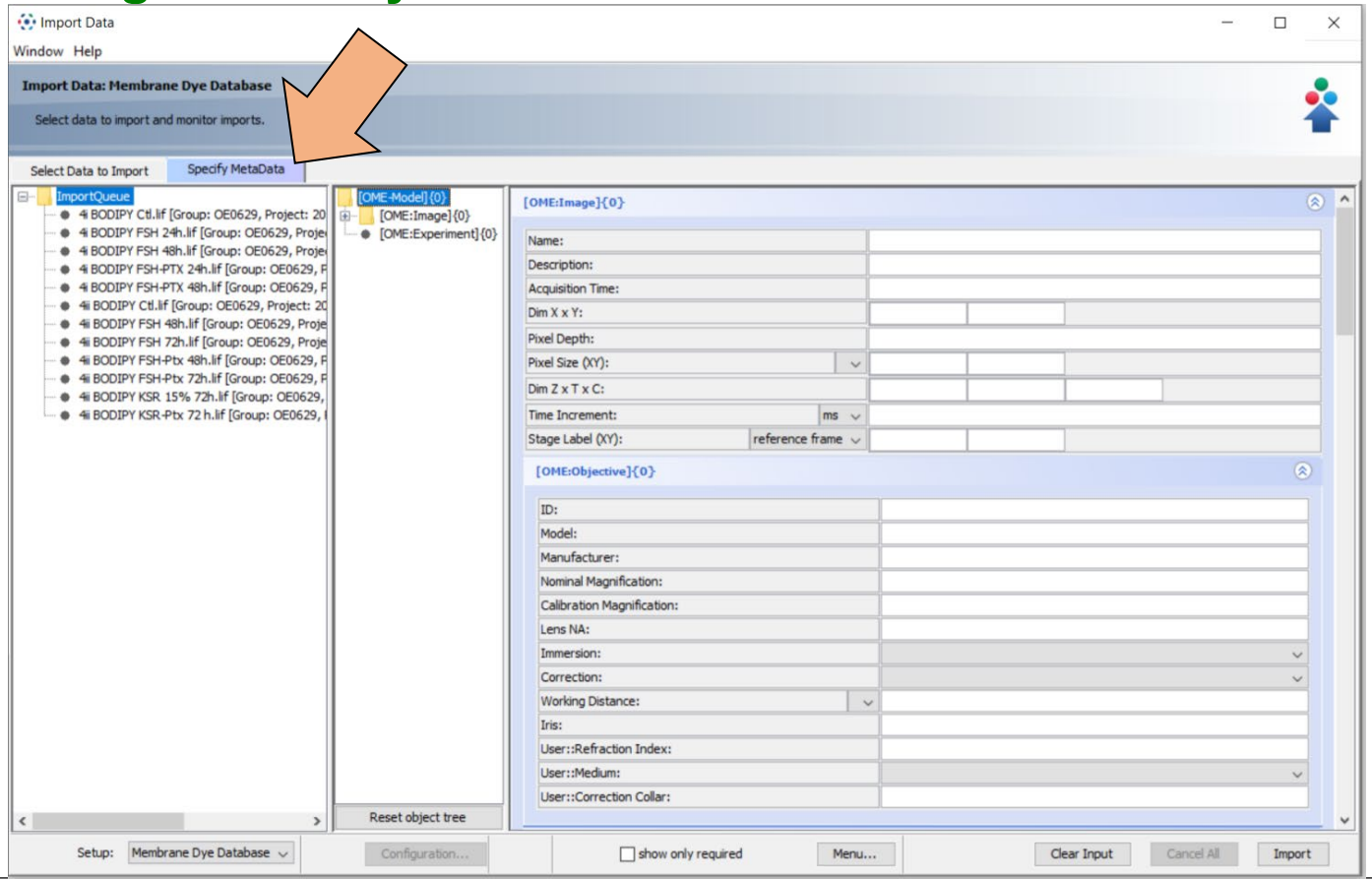



```
script params
File_Annotation 46978
IDs [112]
Data_Type Dataset
set ann id 46978
Original File 162024 2023_Test13-5_Mustermann_Dataset_ID_112.csv
Failed to sniff delimiter, using ','
header ['Image;Organism;Strain;Anatomical structure;Cell Type;Cell Activation;Concentrated
(Cell Activation);Unit (Concentrated - Cell Activation);Time (Cell Activation);Compound
Based Treatment;Treatment protocol;Test']
image_index: 0 well_index: -1 plate_index: -1
Image not found: 20180110_DMSO_0h_pERK_01.czi;Mus musculus;C57BL/6;Spleen;CD4-positive
Can't find object by image, well or plate name
Image not found: 20180110_DMSO_3h_pERK_01.czi;Mus musculus;C57BL/6;Spleen;CD4-positive
Can't find object by image, well or plate name
Image not found: 20180110_DMSO_3h_pERK_02.czi;Mus musculus;C57BL/6;Spleen;CD4-positive
Can't find object by image, well or plate name
Image not found: 20180110_DMSO_3h_pERK_03.czi;Mus musculus;C57BL/6;Spleen;CD4-positive
Can't find object by image, well or plate name
Image not found: 20180110_DMSO_3h_pERK_04.czi;Mus musculus;C57BL/6;Spleen;CD4-positive
Can't find object by image, well or plate name
Image not found: 20180110_DMSO_3h_pERK_05.czi;Mus musculus;C57BL/6;Spleen;CD4-positive
```

Reason:

German and English Excel versions use different delimiters in the CSV file. While the script tries to identify the correct delimiter, it may occur that the delimiter is not correctly determined. This might result in an upload failure if the default delimiter is not the delimiter used in your csv for field separation. Please consult with your OMERO administrator if this issue occurs!

At a glance: Key-Value Pair Annotation with OMERO.mde (or MDEmic)



In **OMERO.insight**, go to the

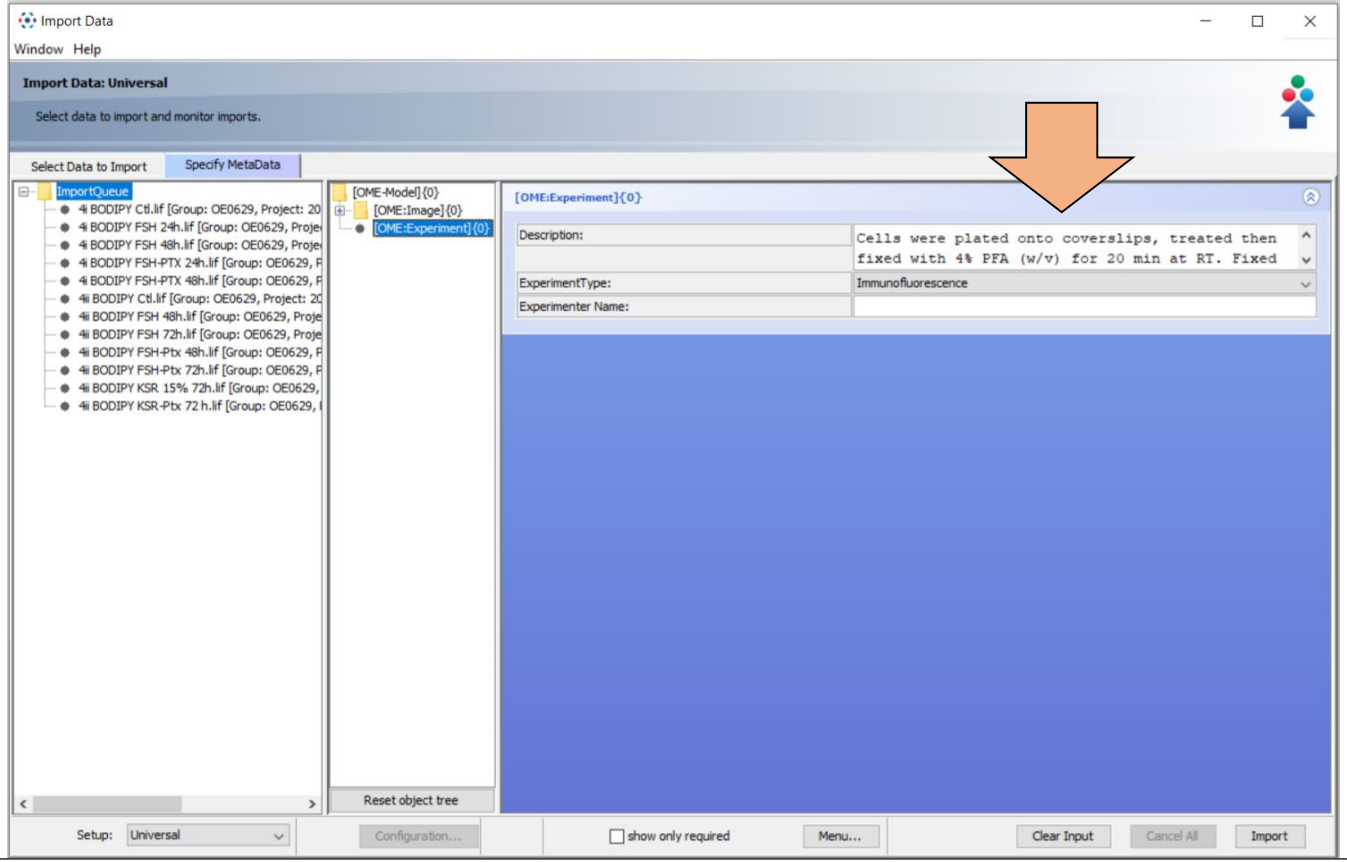
Specify MetaData

tab before importing the selected import queue.

(OMERO.mde is integrated into the OMERO.insight client)

For details, see: Kunis et al., 2021, Nature Methods. https://doi.org/10.1038/s41592-021-01288-z

At a glance: Key-Value Pair Annotation with OMERO.mde (or MDEmic)

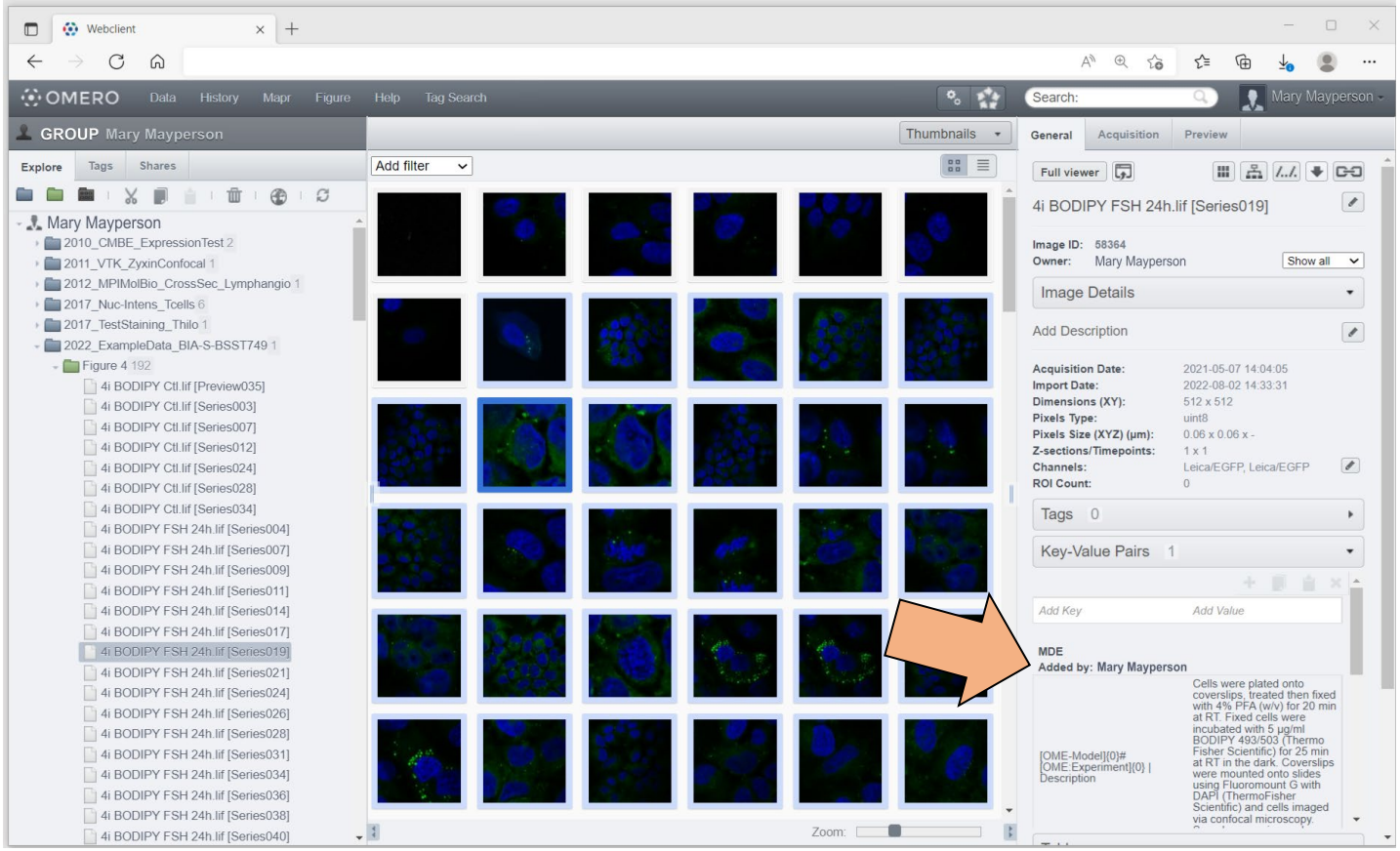


Use the entry masks of OMERO.mde's user interface to review the automatically extracted metadata from the files and to annotate metadata before import.

(OMERO.mde is fully configurable and complies with the OME Data Model)

For details, see: Kunis et al., 2021, Nature Methods. https://doi.org/10.1038/s41592-021-01288-z

Review the Key-Value Pairs after upload (here: OMERO.web)



Note:
MDE-generated Key-Value Pairs cannot be edited manually after import!

Example Data: Hanyaloglu et al. (2021). *BioStudies*, S-BSST749. Retrieved from https://www.ebi.ac.uk/biostudies/BioImages/studies/S-BSST749

Using ontologies with Key-Value Pairs in OMERO

For humans, natural language terms are good to understand the data.

For computers, natural language terms can be ambiguous.

- Unique identifiers are optimal for **machine readability** but are hard for humans to read. (e.g., a URI or URL)

OMERO does not provide a direct connection between the Key-Value Pair terms and ontologies so far.

→ What are the current recommendations?

Ontology-based Annotation in OMERO - recommendation

To create machine-actionable metadata for your data, make use of **ontology terms** and **ontology term source references**:

- Use the ontology-derived term for a specific Key as the Value
- Add the ontology term URL as the Value for a second Key using the same <Key> + „Term Accession Number“

KEY	VALUE
Biological entity	T cell receptor complex
Biological entity Term Accession Number	http://purl.obolibrary.org/obo/GO_0042101

Adopted from the REMBI guidelines: 10.1038/s41592-021-01166-8, in combination with ISA-TAB: https://isa-specs.readthedocs.io/en/latest/isatab.html (last access: 2023-08-16)

How should data be annotated to be sufficiently enriched?

The specific content of your annotation depends on your

- research field
- experimental setup
- analysis strategy
- intended reuse potential for your data.

Bioimaging-specific recommendations:

- Sarkans et al. (2021) REMBI: Recommended Metadata for Biological Images – enabling reuse of microscopy images in biology. *Nat Methods*, Dec;18(12):1418-1422. doi: 10.1038/s41592-021-01166-8.
- Hammer et al. (2021) Towards community-driven metadata standards for light microscopy: tiered specifications extending the OME model. *Nat Methods*, Dec;18(12):1427-1440. doi: 10.1038/s41592-021-01327-9.