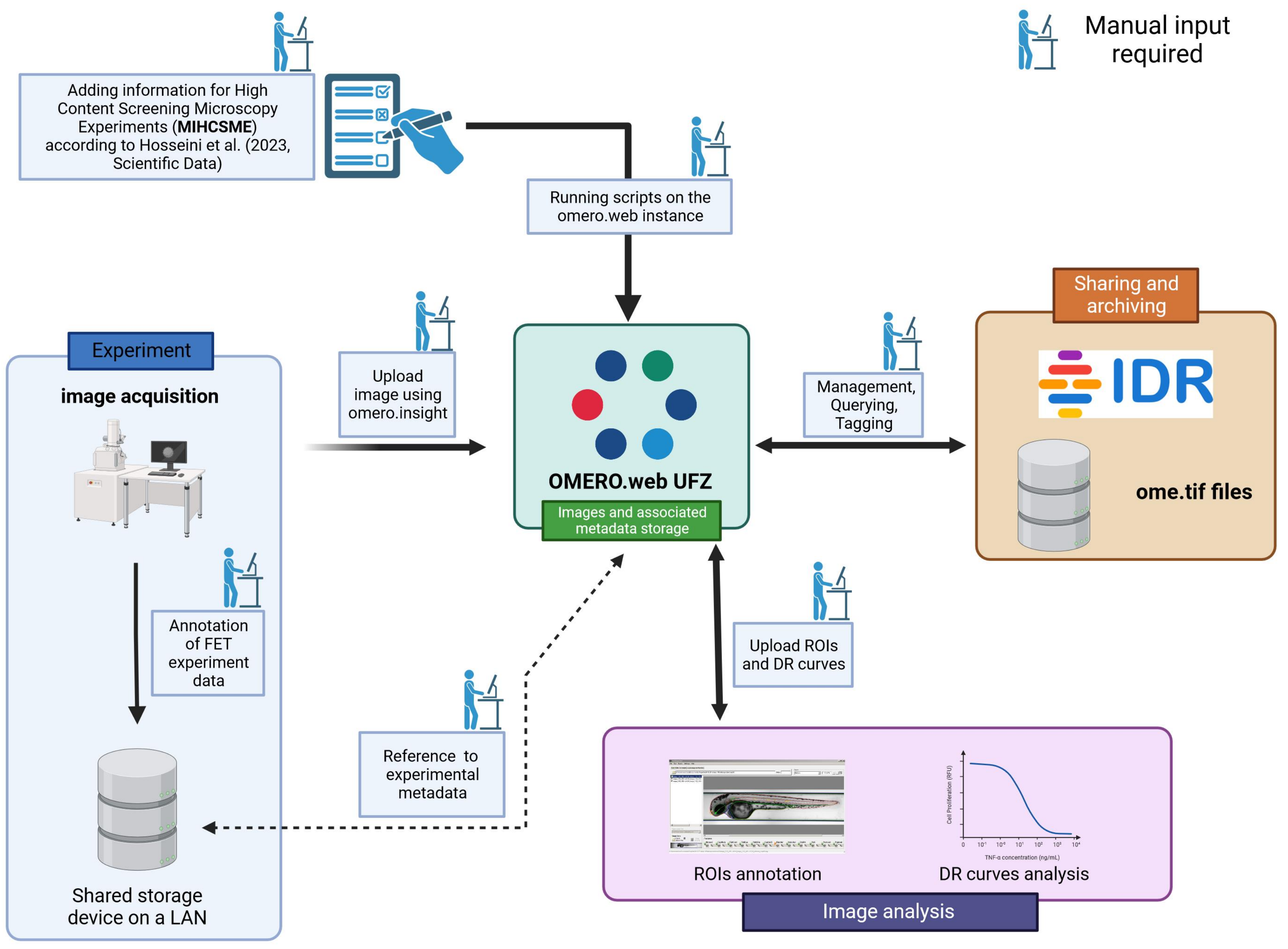
# Developing (semi)automatic analysis pipelines and technological solutions for metadata annotation and management in high-content screening (HCS) bioimaging

Riccardo Massei[1,2,3], Stefan Scholz[3], Wibke Busch[3], Thomas Schnicke[2], Hannes Bohring[2], Jan Bumberger[1,2]

[1]Department of Monitoring and Exploration Technologies, UFZ—Helmholtz Centre for Environmental Research , Leipzig, Germany

[2] Research Data Management (RDM) and Scientific and Commercial Data Processing (WKDV)

[3]Department of Bioanalytical Ecotoxicology, UFZ—Helmholtz Centre for Environmental Research, Leipzig, Germany



Manual input required

Adding information for High Content Screening Microscopy Experiments (**MIHCSME**) according to Hosseini et al. (2023, Scientific Data)

Running scripts on the omero.web instance

Experiment

**image acquisition**

Upload image using omero.insight

**OMERO.web UFZ**

Images and associated metadata storage

Management, Querying, Tagging

Sharing and archiving

IDR

**ome.tif files**

Annotation of FET experiment data

Upload ROIs and DR curves

Reference to experimental metadata

Shared storage device on a LAN

ROIs annotation

DR curves analysis

Image analysis

**Fig.1:** A general non-automatic workflow for storing, processing, analysing, and sharing zebrafish HCS pictures and associated metadata.

## Scientific background and motivation

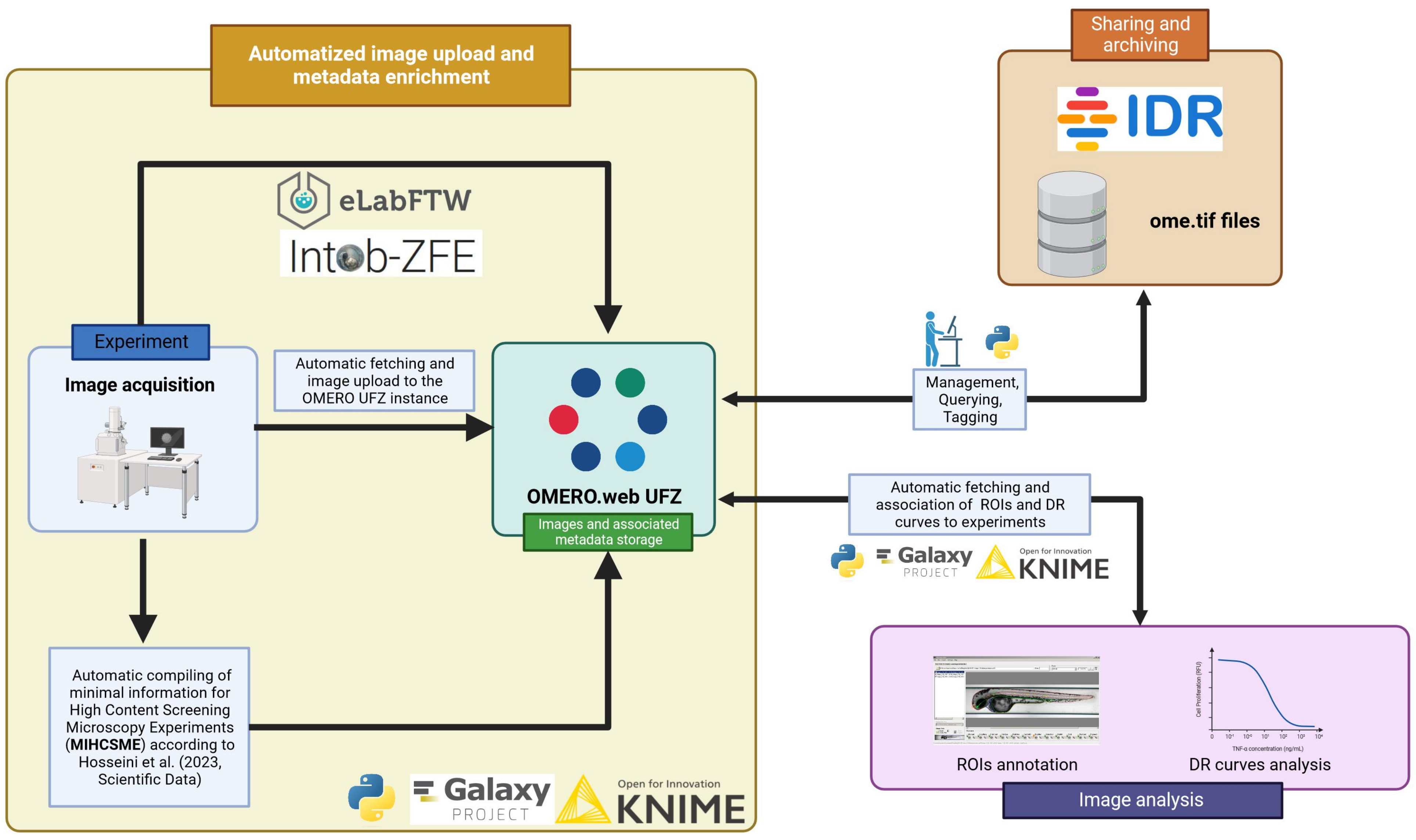
**High-content screening (HCS)** bioimaging approaches are powerful techniques consisting of the automated imaging and analysis of large numbers of biological samples, to extract quantitative and qualitative information from the images. HCS still presents **several bottlenecks** restraining these approaches from exerting their full potential for scientific discoveries. As major example, a **huge amount of metadata is generated in each experiment**, capturing critical information about the images. The efficient and accurate treatment of image metadata is of great importance, as it provides insights that are essential for effective image management, search, organisation, interpretation, and sharing. It is vital to find ways to properly deal with the huge amount of complex and unstructured data for implementing **Findable, Accessible, Interoperable and Reusable (FAIR)** concepts in bioimaging.

## Our goals

In the frame of **NFDI4BIoImaging**, we want to **develop new solution and pipeline based on open software as open software products for storing, processing, analysing, and sharing HCS data.** In particular, we want to develop solutions to make findable and machine-readable metadata using (semi)automatic analysis pipelines.

## Proposed (semi)automatic workflow using KNIME and Galaxy

We proposed an **automated analysis pipeline for storing, processing, analysing, and sharing HCE bioimaging data**. The (semi)automatic workflow was developed by taking as a case study a dataset of zebrafish larvae images previously obtained from an automated imaging system generating data in an HCS fashion. Images are automatically enriched with standardize HCS metadata according to Hosseini et al., (2023), unique identifiers (i.e. **key-value pairs, tags**), additional experimental metadata (i.e. **raw data, regions of interest**) and uploaded to the **UFZ-OME Remote Objects (OMERO) server** using python scripts embedded in workflows developed with **KNIME or GALAXY**. The workflows give the possibility to the user to intuitively fetch images from the local server and perform image analysis (i.e. **annotation**) or even more complex toxicological analyses (dose response modelling). Furthermore, we want to improve the FAIRness of the protocol by adding a direct upload link to the **Image Data Resource (IDR)** repository to automatically prepare the data for publication and sharing.



Automatized image upload and metadata enrichment

eLabFTW

Intob-ZFE

Experiment

**Image acquisition**

Automatic fetching and image upload to the OMERO UFZ instance

**OMERO.web UFZ**

Images and associated metadata storage

Automatic compiling of minimal information for High Content Screening Microscopy Experiments (**MIHCSME**) according to Hosseini et al. (2023, Scientific Data)

Galaxy PROJECT

Open for Innovation KNIME

Sharing and archiving

IDR

**ome.tif files**

Management, Querying, Tagging

Automatic fetching and association of ROIs and DR curves to experiments

ROIs annotation

DR curves analysis

Image analysis

NFDI4BIOIMAGE

**RDM homepage**

### References

- Hosseini, Rohola, et al. "FAIR High Content Screening in Bioimaging." Scientific Data 10.1 (2023): 462