<HMC> HELMHOLTZ Metadata Collaboration
HMC Conference
Virtual Event // 04.-06.11.2024

Contribution ID: **74** Type: **TALK**

Building FAIR image analysis pipelines for high-content-screening (HCS) data using Galaxy

Tuesday 5 November 2024 13:00 (20 minutes)

Imaging is crucial across various scientific disciplines, particularly in life sciences, where it plays a key role in studies ranging from single molecules to whole organisms. However, the complexity and sheer volume of image data, especially from high-content screening (HCS) experiments involving cell lines or other organisms, present significant challenges. Managing and analysing this data efficiently requires well-defined image processing tools and analysis pipelines that align with the FAIR principles—ensuring they are findable, accessible, interoperable, and reusable across different domains.

In the frame of NFDI4BioImaging (the National Research Data Infrastructure focusing on bioimaging in Germany), we want to find viable solutions 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. In scientific research, such pipelines are crucial for maintaining data integrity, supporting reproducibility, and enabling interdisciplinary collaboration. These tools can be used by different users to retrieve images based on specific attributes as well as support quality control by identifying appropriate metadata.

Galaxy, an open-source, web-based platform for data-intensive research, offers a solution by enabling the construction of reproducible pipelines for image analysis. By integrating popular analysis software like CellProfiler and connecting with cloud services such as OMERO and IDR, Galaxy facilitates the seamless access and management of image data. This capability is particularly valuable in bioimaging, where automated pipelines can streamline the handling of complex metadata, ensuring data integrity and fostering interdisciplinary collaboration. This approach not only increases the efficiency of HCS bioimaging but also contributes to the broader scientific community's efforts to embrace FAIR principles, ultimately advancing scientific discovery and innovation.

In the present study, we proposed an automated analysis pipeline for storing, processing, analysing, and sharing HCS bioimaging data. The (semi)automatic workflow was developed by taking as a case study a dataset of zebrafish larvae and cell lines images previously obtained from an automated imaging system generating data in an HCS fashion. In our workflows, images are automatically enriched with metadata (i.e. key-value pairs, tags, raw data, regions of interest) and uploaded to the UFZ-OME Remote Objects (OMERO) server using a novel OMERO tool suite developed with GALAXY. 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.

Please specify ”other”

In addition, please add 3 to 5 keywords.

OMERO, Galaxy, FAIR, workflow, HCS

Please specify ”other”

For whom will your contribution be of most interest?

Researchers

Please assign yourself (presenting author) to one of the following groups.

Data professionals and stewards

Primary author: MASSEI, Riccardo

Co-authors: Dr NYEFFLER, Jo (UFZ); Dr BERNDT, Matthias (UFZ); SCHOLZ, Stefan (UFZ - ETOX); Dr DUNKER, Susanne (iDiv); BUSCH, Wibke; BUMBERGER, Jan; BOHRING, Hannes

Presenter: MASSEI, Riccardo

Session Classification: Session E2

Track Classification: Connecting research data: 5. Technical solutions for findable and machine-readable metadata