HELMHOLTZ
IMAGING

Contribution ID: **10**   Type: **not specified**

A Research Data Management (RDM) framework for High-Content Screening (HCS) bioimaging and video data of zebrafish (Danio rerio) embryos

Wednesday 25 June 2025 11:17 (13 minutes)

In recent decades, high content screening (HCS) assays using zebrafish (Danio rerio) embryos have become increasingly common in international research laboratories. Several of these assays involve the acquisition of large datasets in a few hours or the acquisition of video data using automated recording devices. After analysis, this data can aid in the understanding of cellular processes and support drug development testing. It is clear that such approaches produce a significant amount of information, ranging from raw data in various file formats (e.g., AVI, TIFF, JPG) to analysis results that require proper metadata annotation and storage. These files need to be linked to the fish embryo exposure assay (FET) results for proper context understanding. In addition to the importance of data storage for these files, which can be up to several terabytes in size, it is also crucial to consider the reusability of analysis pipelines according to FAIR (Findable, Accessible, Interoperable, and Reusable) standards. It is evident that research data management (RDM) for this specific data type presents a significant RDM challenge.

In this work, we showcase our in-house strategy for managing and storing zebrafish larvae images and videos. This framework comprises various data infrastructures and databases that can be interconnected using Workflow Management Systems (WMS). Unique sample ID links from the Integrated Effect Database for Toxicological Observations (INTOB) - a comprehensive data management tool for zebrafish effect data - can be linked to OMERO, which is used to store HCS bioimaging data and related metadata. Furthermore, we use the Helmholtz data infrastructure, including dCache/Infinite Space, to store video data and more space-intensive image formats. Finally, we developed KNIME and Galaxy pipelines to retrieve and analyze the data in a reproducible manner. This approach enables us to tackle the complexity of metadata, ensure HCS data integrity, and facilitate internal collaboration.

Presenter: MASSEI, Riccardo

Session Classification: Talks 6-10