New Name:

FISH Omics Exchange Data Formats

New Sub-Title:

FISH Omics Exchange Data Format: format conceptualization, metadata specifications and pipelines for repository deposition and data sharing

New Mission Statement:

A key output of the 4DN project is the open exchange of datasets related to the structure of the human cell nucleus and the genome, within. Recent years have seen a rapid expansion of FISH-omics methods, which quantify the spatial organization of DNA, RNA and protein in the cell and provide expanded understanding of how higher-order chromosome structure relates to transcriptional activity and cell development. Despite this progress, FISH-based image-data are not yet routinely made public upon scientific publication because of the lack of common specifications for data exchange. This challenge is experienced across the bioimaging community, as a result a solution built, tested and proven in 4DN can have a wide impact all over the world. Central to addressing this hurdle is the work carried out by the 4DN Imaging Working Group (IWG) in collaboration with the 4DN-DCIC, the Imaging and Omics Working Group (IOWG) and the Open Microscopy Environment initiative (https://www.openmicroscopy.org/) to support the accessibility and exchange of FISH Omics data. This will include the development of community specifications for image-data formats, quality control and documentation metadata and shared data management and analysis pipelines.

The mission of this sub-group of the IWG will focus on establishing standardized data formats of typical multiplexed FISH Omics data, and metadata specifications to promote optimized analysis and data-exchange.

Specific attention will be devoted to make sure specifications will be flexible enough to include other in-situ imaging-based omics approaches, and will be extensible enough to permit the description of additional multi-omic imaging data to be mapped to these structural data, such as RNA expression, protein distribution, cell morphology, or tissue coordinates.

Furthermore, we will establish pipelines exemplifying the conversion of SMLM image and localization data from the proprietary Bruker binary data format to a cloud-optimized OME-Zarr format [https://github.com/ome/ome-zarr-py] and showcase the analysis from the zarr-file format with software tools like IMGR [Walking along Chromosomes].

It should be noted that this effort will support super-resolved and widefield data being acquired with Bruker Vutara or VXL microscopes, but welcomes data produced from other imaging platforms!

We will specifically focus on two initial types of experiments:

1) Three-dimensional (3D) Chromatin Tracing (CT) experiments in which polymer tracing algorithms are used to string together the localization of individual DNA bright Spots to reconstruct the three-dimensional (3D) path of chromatin fibers (e.g., ORCA, https://doi.org/10.1038/s41596-020-00478-x; MINA, https://doi.org/10.1038/s41596-021-00518-0; Hi-M, https://doi.org/10.1016/j.molcel.2019.01.011; seqFISH+, https://doi.org/10.1038/s41586-019-1049-y; OligoFISSEQ, https://doi.org/10.1038/s41592-020-0890-0; DNA-MERFISH, https://doi.org/10.1016/j.cell.2020.07.032; and IGS).

2) Single-Molecule Localization Microscopy (SMLM) for volumetric imaging, such as OligoSTORM and OligoDNA-PAINT.