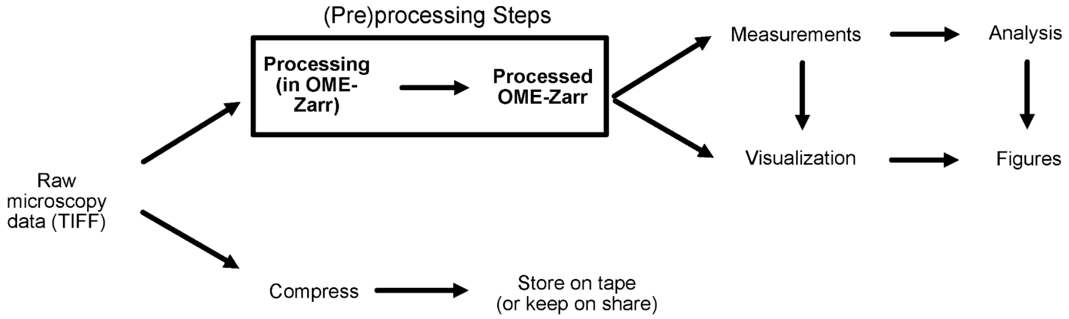
Fractal Data Life Cycle / Workflow Flexibility

Data life cycle



Core user cases:

DLC1. As a user, I can process the data that directly comes from a microscope acquisition using Fractal

DLC2. As a user, I can losslessly compress uncompressed raw microscope image files and resave them as new raw, allowing the deletion of the original uncompressed data

DLC3. As a user, I can convert my microscope input (raw or compressed, see DLC2) into OME-Zarr, and I am able to process the data to correct for microscopy related “artifacts” (e.g. uneven illumination, etc).

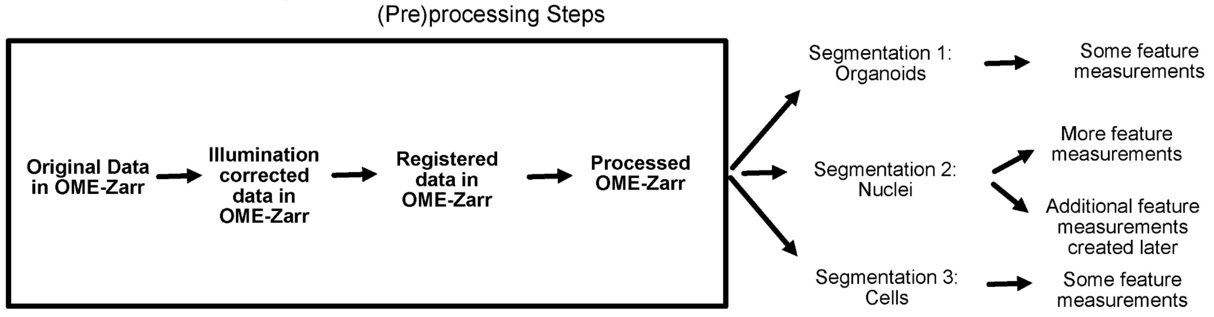
DLC4. As a user, I can use the processed OME-Zarr output structure in order to perform further measurements which will not affect the image content, only augment the OME-Zarr structure into having more information about the underlying biological data (label maps, feature extraction, etc)

DLC5. As a user, I can visualize the processed OME-Zarr in napari/other viewer, as the output structure is OME-Zarr standard compliant.

DLC6. As a user, I can perform e.g. statistical analysis on the extracted features and combine the extracted values together with the image data for e.g. Figure creation /presentations/science sharing.

DLC7. As a user, I am responsible for keeping the compressed microscope raw data on the server until it is taped. Taping must follow local institution/group policies

Workflow flexibility


(Pre)processing Steps
Original Data in OME-Zarr
Illumination corrected data in OME-Zarr
Registered data in OME-Zarr
Processed OME-Zarr
Segmentation 1: Organoids
Some feature measurements
Segmentation 2: Nuclei
More feature measurements
Additional feature measurements created later
Segmentation 3: Cells
Some feature measurements

(Pre)processing Steps: Users need to be able to test them on subsets of the data, but eventually run a pipeline with the relevant parameters for the full dataset. Users can optionally keep some intermediate data, but the default is to overwrite data when doing the preprocessing, just have the "Processed OME-Zarr" saved in the end

Core user cases:

WF1. As a user, I can process the data that directly comes from a microscope acquisition using Fractal

WF2. As a user, I can combine multiple tasks into a (pre)processing workflow which enables the correction/augmentation of the raw images to correct for microscopy artefacts (e.g. illumination correction, possible deconvolution, registration of channels, stitching, etc)

WF3. As a user, I can test particular tasks or whole workflows on a subset of the initially acquired images in order to explore the task's parameter space. Each new test will create a new, temporary .zarr structure reflecting this image subset.

WF4. As a user, I can retrieve all of the parameters for all of the tasks associated in creating a particular .zarr output.

WF4. As a user, I can retrieve and share the processing workflow I have used in order to produce the processed OME-Zarr structure, which contains all of the parameters used for processing the original data until this point.

WF5. As a user, I can retrieve the current workflow state at any time of the (Pre)processing, including the parameters used for redoing particular tasks.

WF6. Each .zarr output history is directly associated with its presence on disk, i.e. if a .zarr output is deleted, so is its history.

WF7. Segmentation of the processed OME-Zarr output can happen at different times, as it only creates new label layers into the .zarr structure. Segmentation can thus happen at virtually any order, as long as dedicated segmentation strategies exist (for now focused on neural network models) that can create such labels from the processed OME-Zarr input.

WF8. As a user, I can run further segmentation strategies and perform feature measurements based on already existing labels

WF9. As a user, I can test segmentation or measurement workflows on subsets of the data (equivalent to WF3 for segmentation and measurement workflows).

WF10. All workflows with their parameters that are run on the whole dataset, creating either segmentations or measurements, are saved. As a user, I can see and retrieve these parameters.