Data preparation for SEM/FIB Tomography

The data are contained in multiple files. In the zip folder for SEM/FIB Tomography acquisition data (eg: - NFDI Tomographiedaten/20200818\_AlSi13 XRM tomo2.zip), there will be a file called EMproject.emxml. This file contains general project information. And is needed for the acquisition\_main and acquisition\_dataset schemas.

Then there is a folder called Images which multiple folders containing the images of the slices cut during the tomography like “SEM Image”, “SEM Image 2”, etc

Each such folder is a dataset. The acquisition\_dataset schema uses information from the EMproject.emxml file corresponding to the Dataset corresponding to the folder-name.

In the snippet below, the dataset corresponding to SEM Image, should be mapped with the highlighted values (rows, columns, TileColumn and TileRow), after first checking if the “Name” of the dataset matches to that of the folder name which is “SEM Image”. In a similar way, the process should be done also for Dataset contained in the folder “SEM Image 2”.

A screenshot of a computer program

Description automatically generated with medium confidence



At present we are only extracting metadata from “SEM Image” and “SEM Image 2”. The EDS folder can be ignored for the moment as the metadata contained in it hidden in proprietary format.

There are a lot of metadata contained in the images that are repeated and common to the dataset. To read such metadata which belong to the dataset schema, it is enough to just read out from one image out of the entire folder.

Finally for the individual image schema, the metadata needs to read from each image and combined into an array of objects.

To reflect this nested nature of the schema, the map is divided into three sections, namely “acquisition”, “dataset” and “image”, each corresponding to the three different schemas which are nested in the order they appear.

This map is in the form “destination path”:”origin path”, contrary to the one for the SEM schema, to adapt to the nested structure of the schema. This format is also in agreement with the MRI JSON map which also contains such a nested structure.

Following are the keys which need some extra data preparation.

|  |  |
| --- | --- |
| acquisition.dataset.entry.definition | To be assigned the value “acquisition\_dataset” |
| acquisition.dataset.entry.datasetType | To be read from the folder name, such as “SEM Image”, SEM Image2”, “EDS”, “EBSD”, “BSE”, etc. |
| acquisition.dataset.entry.numberOfItems | The number of images in the folder, eg. If 719 images are there in SEM Image, the value should be 719 |
| acquisition.dataset.entry.instrument.spot | Need to be deleted from the metadata if acquisition.dataset.entry.instrument.eBeam.beamCurrent.value is present  (The mapping "acquisition.dataset.entry.instrument.spot":"Images.SEM Image.SliceImage.Beam.spot" not included in current map for ease) |
| acquisition.dataset.entry.images.entry.definition | To be assigned the value “acquisition\_image” |

Further Important Points:

* Check if all Booleans are marked properly. Instead of True and False, some values will be On, Off, yes or no. These need to mapped to True or False accordingly.
* The SEM/FIB Tomography metadata uses SI units, so there are no unit mappings for values.