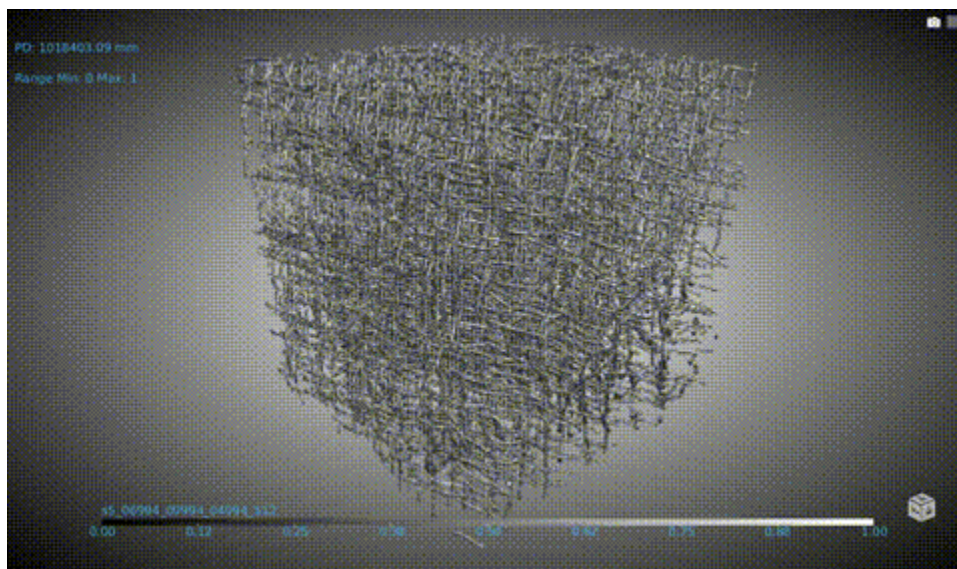**Curated Datasets**

The data available through Vesuvius Challenge is large, frequently updated, and can be overwhelming to navigate.

Here are some organized datasets suited for particular tasks or subproblems. Largely, these curate the segmentation efforts of our team and community. Click one of the datasets to find a download along with more information.

fiber-skeletons

A dataset of manually annotated papyrus fibers - the individual strands that make up a papyrus sheet.

Fiber labels inside a scroll cube.

Another view of fiber skeletons.

Cubes of size 256^3 or 512^3 were selected from within the scroll, and inside each cube, every papyrus fiber was traced and labeled. The fibers have been converted to a volumetric/voxelized representation to be used as inputs to machine learning or other methods that expect 3D image data.

- README
- .zip download (422 MB)

volumetric-instance-labels

Volumetric instance segmentation labels.

Two annotated cubes, with volumetric labels representing papyrus sheet instances.

This dataset contains a subset of Scroll 1, chunked into 256x256x256 cubes. For each cube, the original scroll volume data and the instance segmentation data are provided (each in .nrrd format).

- [README](#)
- [.zip download](#) (1.4 GB)

grand-prize-banner-region

Data related to the 2023 Grand Prize (GP) region from Scroll 1.

Surface mesh.

ML predictions for medial surface segmentation.

The dataset includes the scan volume and segmented surface meshes created by our segmentation team. We also provide predictions from machine learning [models](#) that aim to segment the medial surface of the papyrus sheet.

- [README](#)
- [gp_meshes.7z](#) (288 MB)
- [gp_volume.zarr/](#) (77 GB)
- [gp_tifstack.7z](#) (389.9 GB)
- [gp_legendary-medial-surfaces.7z](#) (5.8 GB)
- [gp_legendary-medial-surfaces-softmax.7z](#) (146.8 GB)