Fragments

Scanning protocols, coordinate systems, and data formats mirror those used for the full scrolls. Every fragment dataset contains:

- **3D X-ray volumes** at several resolutions / beam energies.
- **Multispectral photographs** (RGB + IR).
- **Hand-labeled ink masks** for at least one surface volume, suitable for supervised ML.

We group the fragments by the facility where they were scanned:

1. **DLS Fragments (2023)** – Six fragments scanned at Diamond Light Source (UK).
2. **ESRF Fragments (2025)** – Three fragments scanned on beamline BM18 at the European Synchrotron Radiation Facility (Grenoble, FR).

Work-in-progress 👷 File formats, folder names, and alignment conventions may still shift. Expect additional volumes, surface volumes and meshes and ink labels to appear over time!

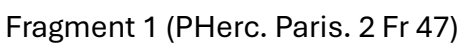
---

1 · DLS (Diamond Light Source) Fragments

The first six fragments to be released. They were scanned at Diamond Light Source. For more technical details, see EduceLab-Scrolls: Verifiable Recovery of Text from Herculaneum Papyri using X-ray CT and EduceLab Herculaneum Scroll Data (2023) Info Sheet.

⚠️ 3D x-ray scan volumes of Fragments 5-6 are aligned, but Fragments 1-4 are NOT aligned.

Fragment 1 (PHerc. Paris 2 Fr 47)

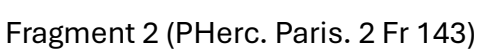


Fragment 1 (PHerc. Paris. 2 Fr 47)

Volume 20230205142449: 3.24µm, 54keV, 7219 x 20MB .tif files. Total size: 145 GB

Volume 20230213100222: 3.24µm, 88keV, 7229 x 24MB .tif files. Total size: 171 GB

Fragment 2 (PHerc. Paris 2 Fr 143)



Fragment 2 (PHerc. Paris. 2 Fr 143)

Volume 20230216174557: 3.24µm, 54keV, 14111 x 46MB .tif files. Total size: 645 GB

Volume 20230226143835: 3.24µm, 88keV, 14144 x 43MB .tif files. Total size: 599 GB

Fragment 3 (PHerc. Paris 1 Fr 34)

Fragment 3 (PHerc. Paris. 1 Fr 34)

Volume 20230212182547: 3.24µm, 88keV, 6650 x 20MB .tif files. Total size: 134 GB

Volume 20230215142309: 3.24µm, 54keV, 6656 x 18MB .tif files. Total size: 121 GB

Fragment 4 (PHerc. Paris 1 Fr 39)

Originally held back for automated scoring in the Kaggle competition, this fragment has since been released.

Fragment 4 (PHerc. Paris. 1 Fr 39)

Volume 20230215185642: 3.24µm, 54keV, 9231 x 23MB .tif files. Total size: 211 GB

Volume 20230222173037: 3.24µm, 88keV, 9209 x 24MB .tif files. Total size: 216 GB

Fragment 5 (PHerc. 1667 Cr 1 Fr 3)

From the same original scroll as Scroll 4 (PHerc. 1667), which was partially opened in 1987 using the Oslo method. Find this fragment on Chartes.it.

Fragment 5 (PHerc. 1667 Cr 1 Fr 3)

Volume 20231121133215: 3.24µm, 70keV, 7010 x 13MB .tif files. Total size: 87 GB

Volume 20231130111236: 7.91µm, 70keV, 3131 x 3MB .tif files. Total size: 8.5 GB

Fragment 6 (PHerc. 51 Cr 4 Fr 8)

Fragment 6 (PHerc. 51 Cr 4 Fr 48)

Volume 20231121152933: 3.24µm, 53keV, 8855 x 29MB .tif files. Total size: 253 GB

Volume 20231130112027: 7.91µm, 53keV, 3683 x 6MB .tif files. Total size: 21 GB

Volume 20231201112849: 3.24µm, 88keV, 8855 x 29MB .tif files. Total size: 253 GB

Volume 20231201120546: 3.24µm, 70keV, 8855 x 29MB .tif files. Total size: 253 GB

Show tiny-fragment context photo

---

2 · ESRF Fragments (BM18, Grenoble — May 2025)

Between **6 May 2025 and 12 May 2025** we scanned three additional fragments on the brand-new **BM18** 4th-generation beamline at the European Synchrotron Radiation Facility. Phase-contrast helical CT, ultrafine 2.2 µm voxels, and generous sample-to-detector distances were explored to squeeze out every bit of ink contrast.

👉 **Draft info-sheet**: ESRF Fragment Data (May 2025)

*All ESRF volumes are published as **OME-Zarr** (six-level multiscale) rather than loose TIFF stacks.

Fragment 500P2

PHerc. 500P2 – infrared

- **2.215 µm, 110 keV** · OME-Zarr
- **4.317 µm, 111 keV** · OME-Zarr
- Multispectral stack (16 bands, 420–1050 nm)
- Case + mesh STL (nylon 12 print-ready)

Fragment 343P

PHerc. 343P – infrared

- **2.215 µm, 111 keV** · OME-Zarr
- **4.320 µm, 116 keV (bin×2)** · OME-Zarr
- Multispectral stack (16 bands)
- Case + mesh STL

Fragment 9B

PHerc. 9B – infrared

- **4.320 µm, 116 keV (bin×2)** · OME-Zarr
- Multispectral stack (16 bands)
- Case + mesh STL

Preliminary ink detection – Our TimesFormer ML models already pick up discrete strokes in the 2.2 µm volumes of 343P and 500P2. For a sneak-peek see the blog post. Surface volumes & precise IR alignment are coming soon.

Data format at a glance

Work-in-progress 👷 File formats, folder names, and alignment conventions may still shift. Expect additional volumes, surface volumes and meshes and ink labels to appear over time!

```
/fragments/            # EduceLab classic datasets (TIFF stacks)
└─ Frag1/PHerc... .volpkg/
    ├─ config.json     # metadata
    ├─ volumes/        # multiple resolutions / energies
    │  └─ 202302.../0000.tif ...
    └─ working/
       └─ 54keV_exposed_surface/
          ├─ surface_volume/00.tif ...
          ├─ ir.png
          ├─ inklabels.png
          └─ alignment.psd

/fragments/        # ESRF 2025 datasets (OME-Zarr)
└─ PHerc0500P2/
    ├─ 2.215um_HEL_TA_0.4m_110keV.zarr/
    ├─ 4.317um_HA_... .zarr/
    ├─ paths/          # surface volumes (WIP)
    ├─ multispectral/
    └─ cases/          # STL meshes & 3D-printed holders
```

- **Fragments 1-6** use the original *TIFF-stack* .volpkg layout.

⚠️ **ESRF fragments** ship as **OME-Zarr multiscale** volumes for instant cloud streaming.

⚠️ All infrared & multispectral images are supplied *pre-aligned* where possible. Otherwise, check the alignment.psd layers.

Training of ML models for ink detection

The idea is to train ML models on the fragments, since we have the ground truth data of where the ink is (in addition to the "crackle" method). Then, those ML models can be applied to the scrolls.

At a high level, training on a fragment works like this:

From a fragment (a) we obtain a 3D volume (b), from which we segment a mesh (c), around which we sample a surface volume (d). We also take an infrared photo (e) of the fragment, which we align (f) with the surface volume, and then manually turn into a binary label image (g). For more details, see Tutorial 5.

License

Fragment datasets are released under different licenses. We encourage you to abide by the related dataset license before working on it.