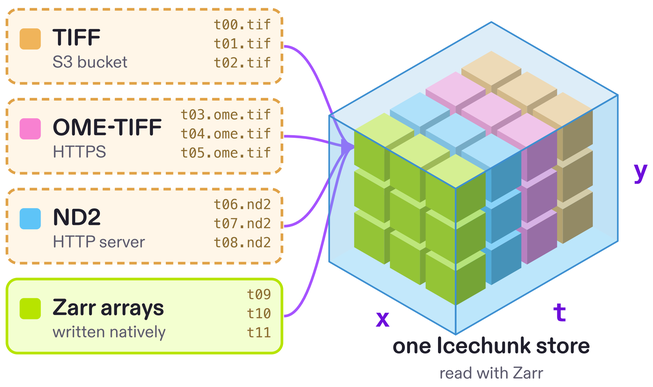
Unlocking Imaging Archives With Icechunk

zero-copy ingest of TIFFs and more into OME-Zarr

earthmover
Ian Hunt-Isaak, PhD · ian@earthmover.io

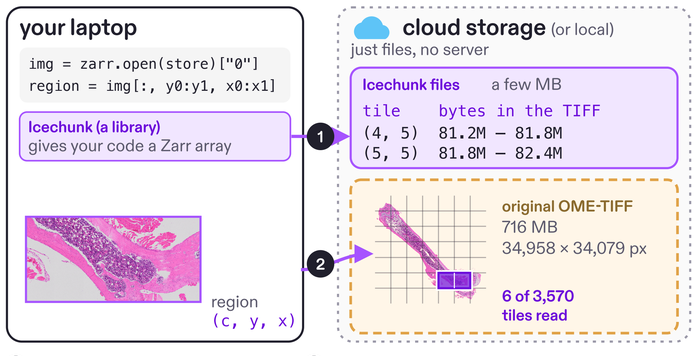
zero-copy ingest

combine sources



- files stay where they are
- parse once, at ingest
- Zarr access without doubling storage

fetch only what you need



1 look up which bytes hold the tiles. 2 read just those bytes from the TIFF.

one region of a 198 MB NDPI slide

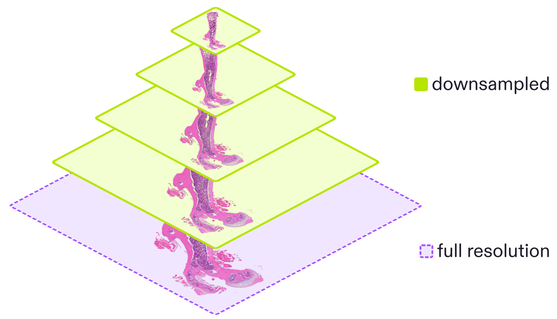
download whole file	198 MB · 1.6 s
Icechunk + coalesced reads	4.4 MB · 0.4 s

OpenSlide CMU-1 slide, 80 ms simulated latency

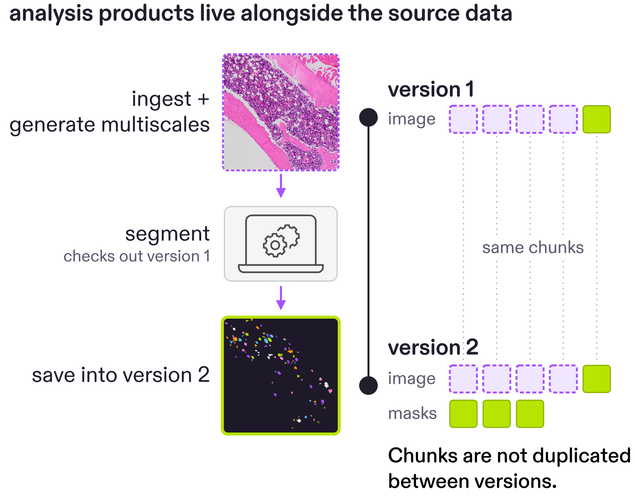
mix with Zarr chunks

raw + outputs, one store

multiscales built during ingest



analysis products live alongside the source data

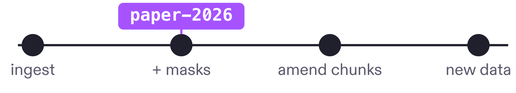


Chunks are not duplicated between versions.

reproducible

Tag the exact version a paper used. It stays readable as new data is added or chunks are amended.

```
repo.readonly_session(tag="paper-2026")
```



make it FAIR: reorganize hierarchy and metadata

files → one OME-Zarr plate

```
s3://cellpainting-gallery/…/BR00116991/
r01c01f01p01-ch1sk1fk1fl1.tiff      2.3M
r01c01f01p01-ch2sk1fk1fl1.tiff      2.3M
r01c01f01p01-ch3sk1fk1fl1.tiff      2.3M
r01c01f01p01-ch4sk1fk1fl1.tiff      2.3M
r01c01f01p01-ch5sk1fk1fl1.tiff      2.3M
r01c01f01p01-ch6sk1fk1fl1.tiff      2.3M
r01c01f01p01-ch7sk1fk1fl1.tiff      2.3M
r01c01f01p01-ch8sk1fk1fl1.tiff      2.3M
r01c01f02p01-ch1sk1fk1fl1.tiff      2.3M
r01c01f02p01-ch2sk1fk1fl1.tiff      2.3M
r01c01f02p01-ch3sk1fk1fl1.tiff      2.3M
r01c01f02p01-ch4sk1fk1fl1.tiff      2.3M
r01c01f02p01-ch5sk1fk1fl1.tiff      2.3M
r01c01f02p01-ch6sk1fk1fl1.tiff      2.3M
r01c01f02p01-ch7sk1fk1fl1.tiff      2.3M
```

…27,648 files, 62.7 GB

normalize layout →

```
BR00116991/          384 wells
├ A/1/               well
│  ├ 0/              field 1 of 9
│  │  ├ 0            full-res (8, 1080, 1080)
│  │  └ 1            16× downsampled (8, 67, 67)
│  └ …
├ A/2/ … P/24/
├ annotations        platemap per well
└ illum              illumination correction
```

virtual native

in the files		in the store
metadata in file names `r08c12f01p01-ch5sk1fk1fl1.tiff`	→	metadata in the hierarchy
inconsistent layouts `images/` `Images/`	→	one uniform layout
sidecar files `load_data.csv` `platemap.txt`	→	annotations in the store

Icechunk

many languages

Any tool that reads Zarr can be made to work with Icechunk. Every language wraps one Rust core.

Python · Rust · JavaScript · R via reticulate · Java via Appose

- streaming reads at hardware limits
- transactional, no race conditions
- version control: snapshots, branches, tags
- no server, just your bucket

see it animated
earth-mover.github.io/icechunk-illustrated

Data: IDR idr0096 (Tratwal et al., *Front. Endocrinol.* 2020; Williams et al., *Nat. Methods* 2017). NDPI: CMU-1, OpenSlide test data, CC0 (openslide.cs.cmu.edu/download/openslide-testdata). Cell Painting Gallery cpg0000 (Chandrasekaran et al., *Nat. Methods* 2024; Weisbart et al., *Nat. Methods* 2024).