what use is research when it goes down the memoryhole?

SciOp!

an online website
https://sciop.net

Resilient Data Infrastructure in an Era of Informational Crisis

Jonny Saunders[123*], Hazel Brenado[3], Raymond Chang[14], Ashley Gay[3], Irene Knapp[3], Ryan Ly[5], Audrey Mendez-Pratt[3], Takuya Sasatani[6], Henrik Schönemann[37], Wouter-Michiel Vierdag[9], William Waites[38], Daniel Aharoni[1]

1) UCLA, Department of Neurology; 2) Institute of Pirate Technology; 3) Safeguarding Research and Culture; 4) UCLA, Physics and Astronomy Dept., 5) Scientific Data Division, Lawrence Berkeley National Laboratory; 6) Dept. of Electrical Engin. and Information Systems, The Univ. of Tokyo, Tokyo, Japan; 7) Humbolt University, Berlin; 8) University of Southampton, UK; 9) European Molecular Biology Laboratory (EMBL)
*Produced with unionized labor, UAW 4811

this poster is online!

Climate data: **gone**.
CDC research: **gone**.
Any mention of transgender people: **gone**.
Our archives are under attack.

SciOp is a project to preserve our collective memory in survivable peer-to-peer archives. Merging Bittorrent as a p2p backbone with the social tooling of federation, we invite our colleagues to join us in building the archives we need - linked communities of practice combining volunteered resources that can scale from megabytes to petabytes, from hyperlocal to global, from now into the the indefinite future.

In 6 months, from April to November 2025, with $0 in funding and a $40/month VPS donated by FlokiNet, SciOp has gone from nothing to...

Indexing 269 datasets with 1,027 uploads.

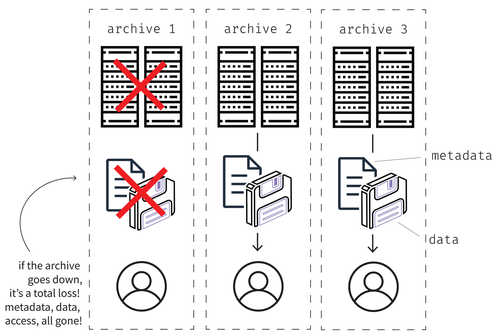
12,000 peers, 11,000 seeders sharing 260 Terabytes in 10 million files.

Total storage capacity in the swarm:

>1.5 Petabytes

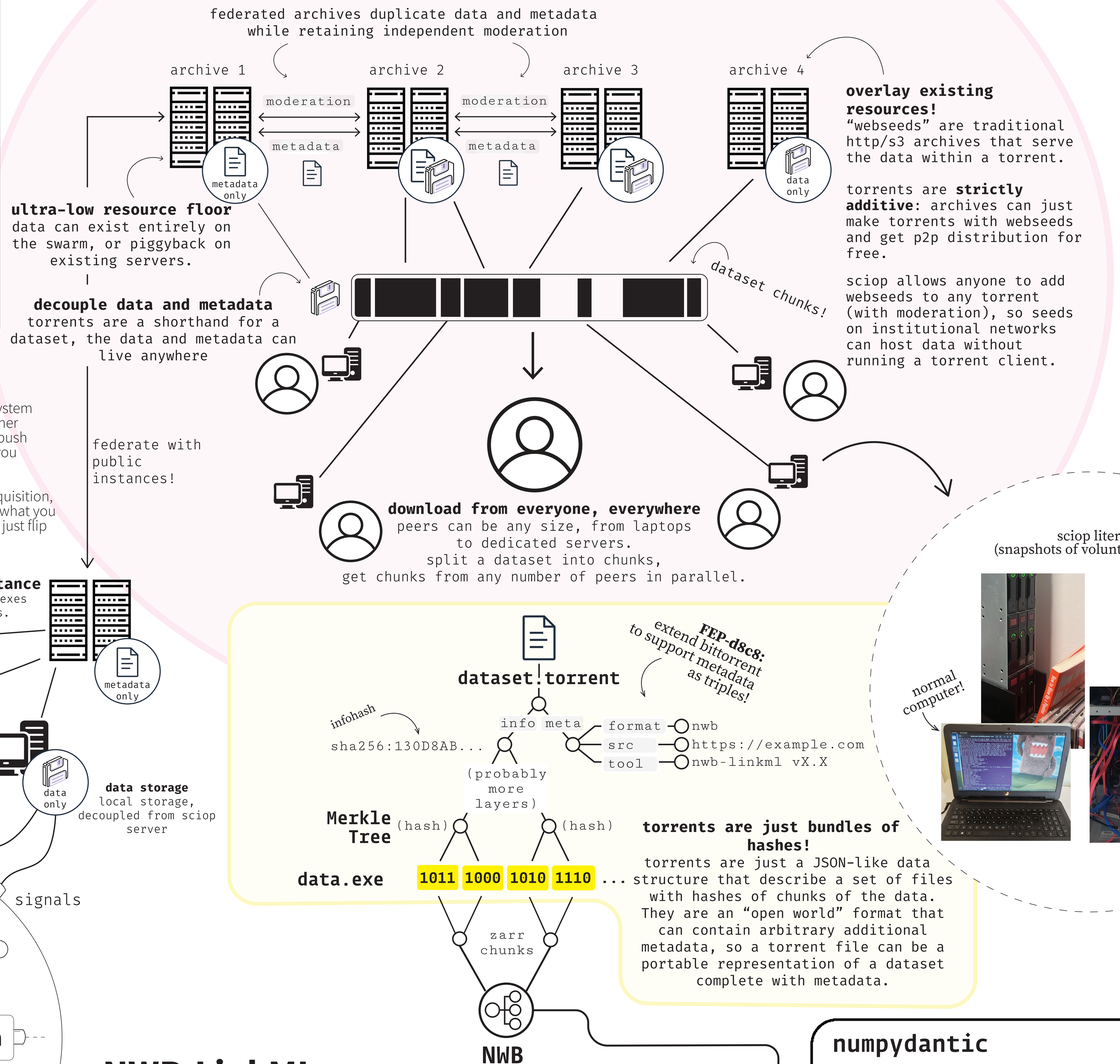
traditional archives

Vertically integrated silos: store and serve all the data and metadata, items only exist by virtue of the archive staying online.



Hybrid Federated/P2P Archives = activitypub + bittorrent

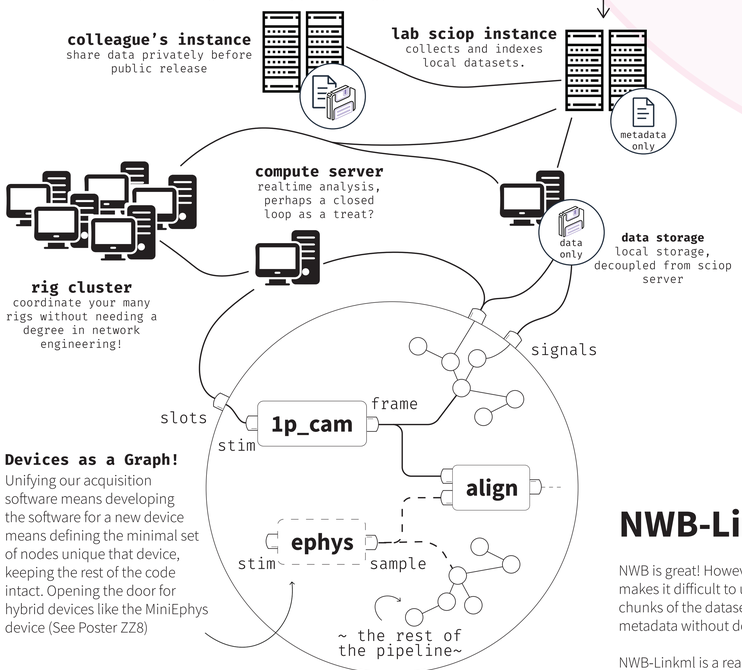
the idea is simple:
servers that federate & talk to each other for redundant, distributed metadata curation and moderation; coordinating scraping, storage, and transfer among p2p bittorrent swarms



Benchtop Federation

We have taken pains to design sciop to be "ridiculously deployable" - no external system services, pure python, just pip install and it goes - to make it simple for labs and other groups to run their own instance. One goal of sciop and its extended universe is to push against ceding infrastructure to "the cloud" by making it easy to use the hardware you already have.

Link the computers on your rigs together, collect data that's clean at the time of acquisition, pipe it to an analysis machine, a raspberry pi with a stack of hard drives, and index what you have with sciop. When it's time to publish, all your data is already locally federated: just flip the switch to make it public and get crawled by participating public archives.



Devices as a Graph!
Unifying our acquisition software means developing the software for a new device means defining the minimal set of nodes unique that device, keeping the rest of the code intact. Opening the door for hybrid devices like the MiniEphys device (See Poster ZZ8)

mio & noob

We are unifying all miniscope acquisition software with a shared library of modular components, decoupling acquisition code from GUIs, minimizing the code needed for new devices and standardizing their programming interface.

We are in the process of refactoring mio to be built on top of noob (see poster ZZ13) for a unified software backbone from acquisition through analysis. Noob was designed to separate the abstract structure of a processing graph from how it's run, and we have nearly completed a p2p networked runner that can seamlessly run complex compositions of pipelines across machines.

After several years of hiatus, autopilot 2 is coming into focus on the horizon...

NWB-LinkML

NWB is great! However its structure as HDF5 files makes it difficult to use at runtime, access partial chunks of the dataset remotely, and introspect its metadata without downloading the whole thing.

NWB-Linkml is a rearchitecting of the NWB format stack, cleanly isolating the abstract schema layer from the concrete serialization layer. This makes all the metadata easily accessible from a YAML file, and makes HDF5 just one of many equivalent serializations: chunked zarr arrays sit comfortably next to csvs, reasonably encoded videos, etc.

After federating sciop, we will be aligning the chunking in torrents with the chunking in zarr arrays to support streaming array access over bittorrent.

```yaml
meta:
  id: my_dataset
  imports:
  - name: a_schema
    from: https://example.com/schema
  - name: nwb
    version: "2.7.0"
    from:
      pypi: nwb-models

my_dataset:
  is_a: nwb:NWBFile
  description: "its data my dude"
  acquisition:
    ephys_recording: "#ephys_recording"

ephys_recording:
  is_a: nwb:ElectricalSeries
  data:
  - type: hdf5
    from: data.h5:/acquisition/ephys
  - type: zarr
    from: blake2s: ...
```

numpydantic

github.com/p2p-ld/numpydantic
numpydantic.readthedocs.io

An interface between abstract array specifications and arbitrary array backends. Use with pydantic models or standalone.
Rather than building tools for a format, build formats for tools.

```python
from pydantic import BaseModel
from numpydantic import NDArray, Shape

class MyModel(BaseModel):
    array: NDArray[Shape["3 x, 4 y, * z"], int]

# numpy
MyModel(array=np.zeros((3, 4, 5), dtype=int))
# dask
MyModel(array=da.zeros((3, 4, 5), dtype=int))
# hdf5 datasets
MyModel(array=("data.h5", "/nested/dataset"))
# zarr arrays
MyModel(array=zarr.zeros((3,4,5), dtype=int))
# video files
MyModel(array="data.mp4")
```

metadata.yaml · video.mp4 · array.zarr · array.npz

Links

this poster	- poster!	https://aharoni-lab.com/sfn2025/sciop
sciop	- source:	https://codeberg.org/safeguarding/sciop
	web:	https://sciop.net
	docs:	https://sciop.net/docs
sciop-cli	- source:	https://codeberg.org/safeguarding/sciop-cli
sciop-scraping	- source:	https://codeberg.org/safeguarding/sciop-scraping
torrent-models	- source:	https://github.com/p2p-ld/torrent-models
	docs:	https://torrent-models.readthedocs.io/
fastapi-activitypub	- source:	https://github.com/p2p-ld/fastapi-activitypub
nwb-linkml	- source:	https://github.com/p2p-ld/nwb-linkml/
mio	- source:	https://github.com/miniscope/mio
noob	- source:	https://github.com/miniscope/noob
FEP-d8c8	- spec:	https://fediverse.codeberg.page/fep/fep/d8c8/
FEP-1580	- spec:	https://fediverse.codeberg.page/fep/fep/1580/

acknowledgements: the shrimps and their shrimp tools