

IF OUR WILDEST DREAMS FOR "OPEN SCIENCE" ARE

Social-First P2P

Strategies for Joy, Agency, and Ethics in Digital Infrastructure

PAY AMAZON TO RENT OUR OWN DATA

PAY MICROSOFT TO GENERATE PAPERS NOBODY WROTE

PAY ELSEVIER TO PUBLISH PAPERS NOBODY READS

TO SURVIVE ANOTHER DAY ON THE PRESTIGE TREADMILL

THOSE ARE NIGHTMARES.

WE NEED BETTER DREAMS

BEYOND THE CLOUD

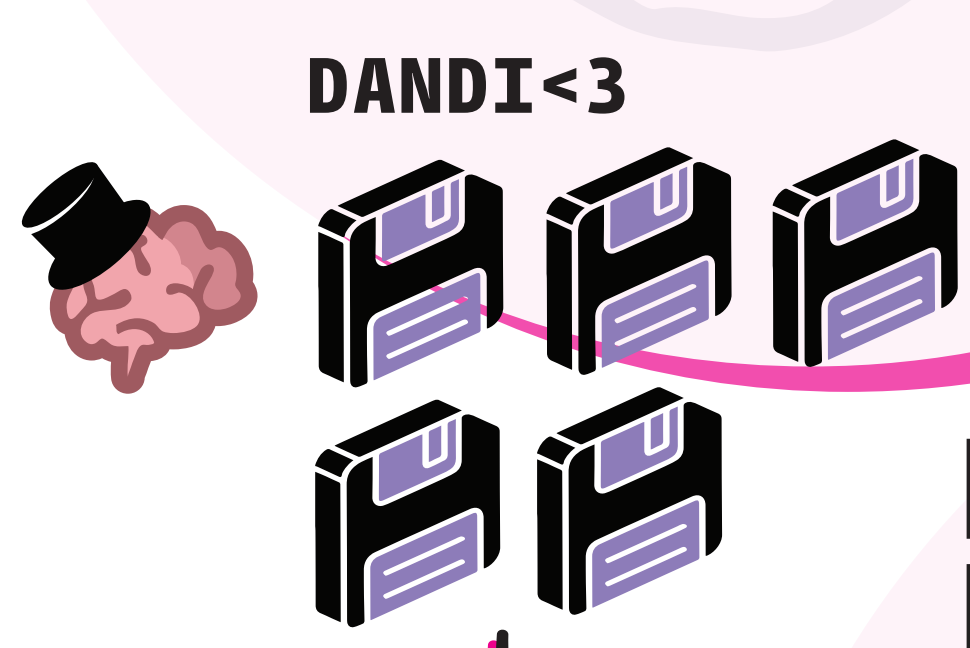
Open data is the next theater of enclosure for scientific infrastructure. In a race for an ill-specified "AI"-driven future thirsty for data, researchers and funding agencies alike are sleep walking into a trap. The cloud isn't optimal or inevitable: it's a business model intent on injecting ever-increasing storage and compute costs into the practice of science[1], profoundly limiting what our infrastructures could be. Peer to peer (p2p) infrastructures are the alternative. We present groundwork in data formats, sketching plans to build a new kind of social-first, graphical p2p system based on continuous, vernacular expression. Avoiding utopian pipe dreams, keeping close to actual practice, we invite you to join us bridging work from private to public; local to global; across domains like data, compute, and publishing. We can have better dreams!

In the cloud, you own nothing. Everything passes through a central point of control

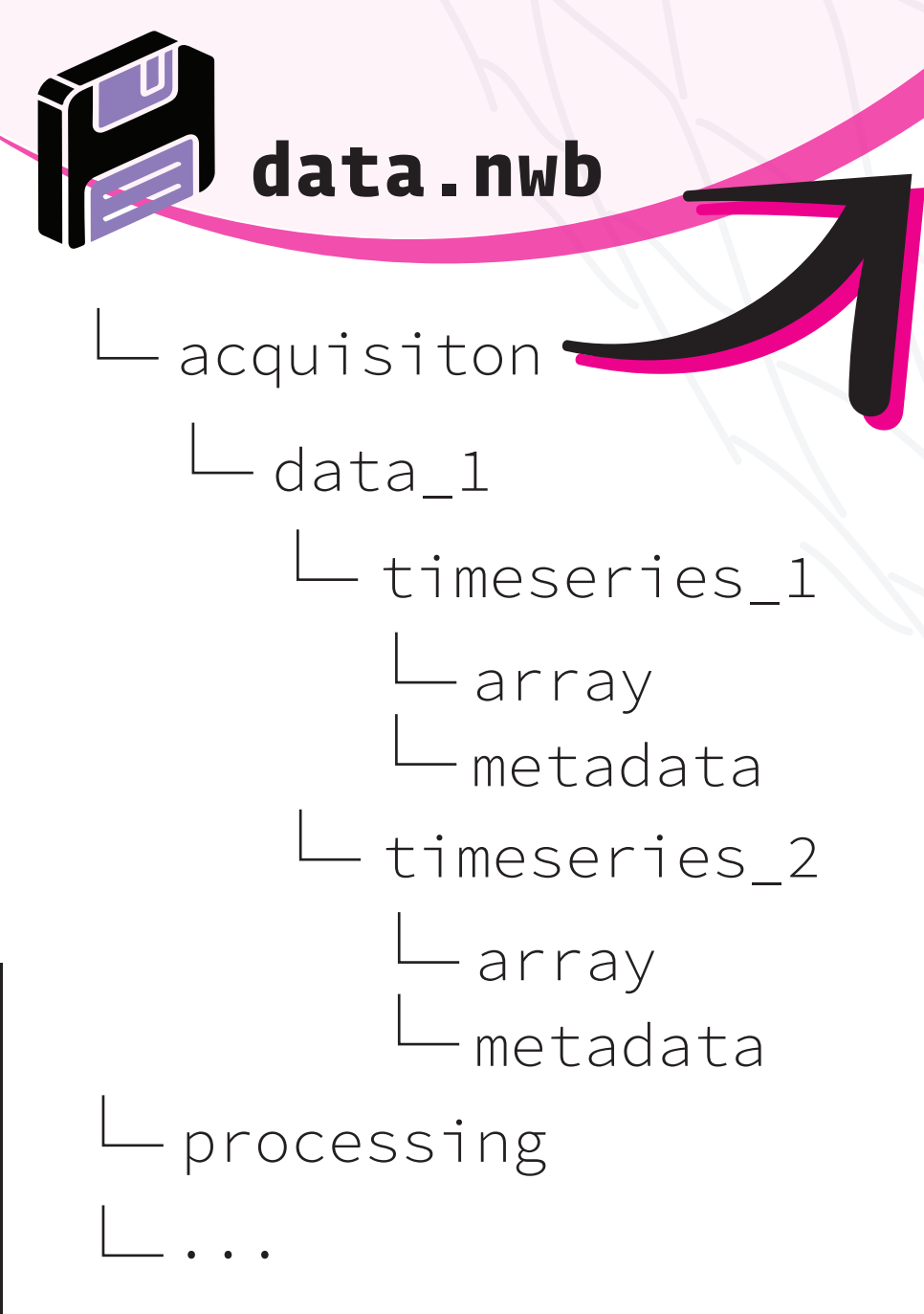
In a P2P system, you own everything. Peers connect directly to one another, from phones, laptops, to large server farms!

Delocalizing Archives

Archives are social! We love our archivists, but why should they have all the fun? We can value their hard work, contribute what we can with eg. schema curation and tool development, and help defray the complexity and cost of hosting with P2P archive mirrors. We scraped DANDI and are using it as a test case for social organization of complex data.

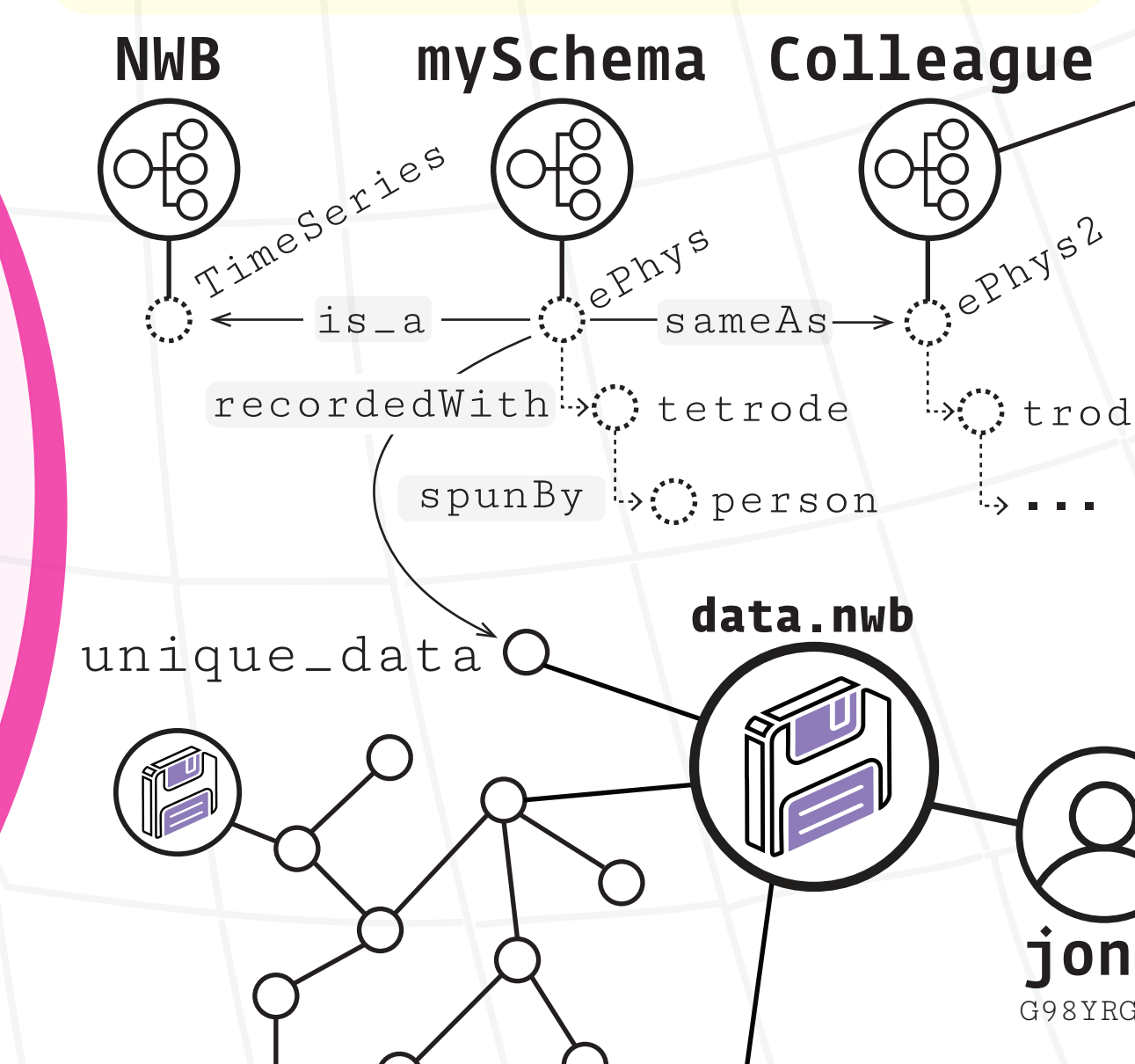


NWB -> LinkML
Formats should be vernacular! To facilitate fluid use of structured data, we developed **nwb-linkml**, an independent implementation of Neurodata Without Borders, translating it to LinkML in the process. This allows us to use NWB data with multiple backends (eg. relational databases, triple stores), write code-free schema extensions, and simplify its API with pydantic.



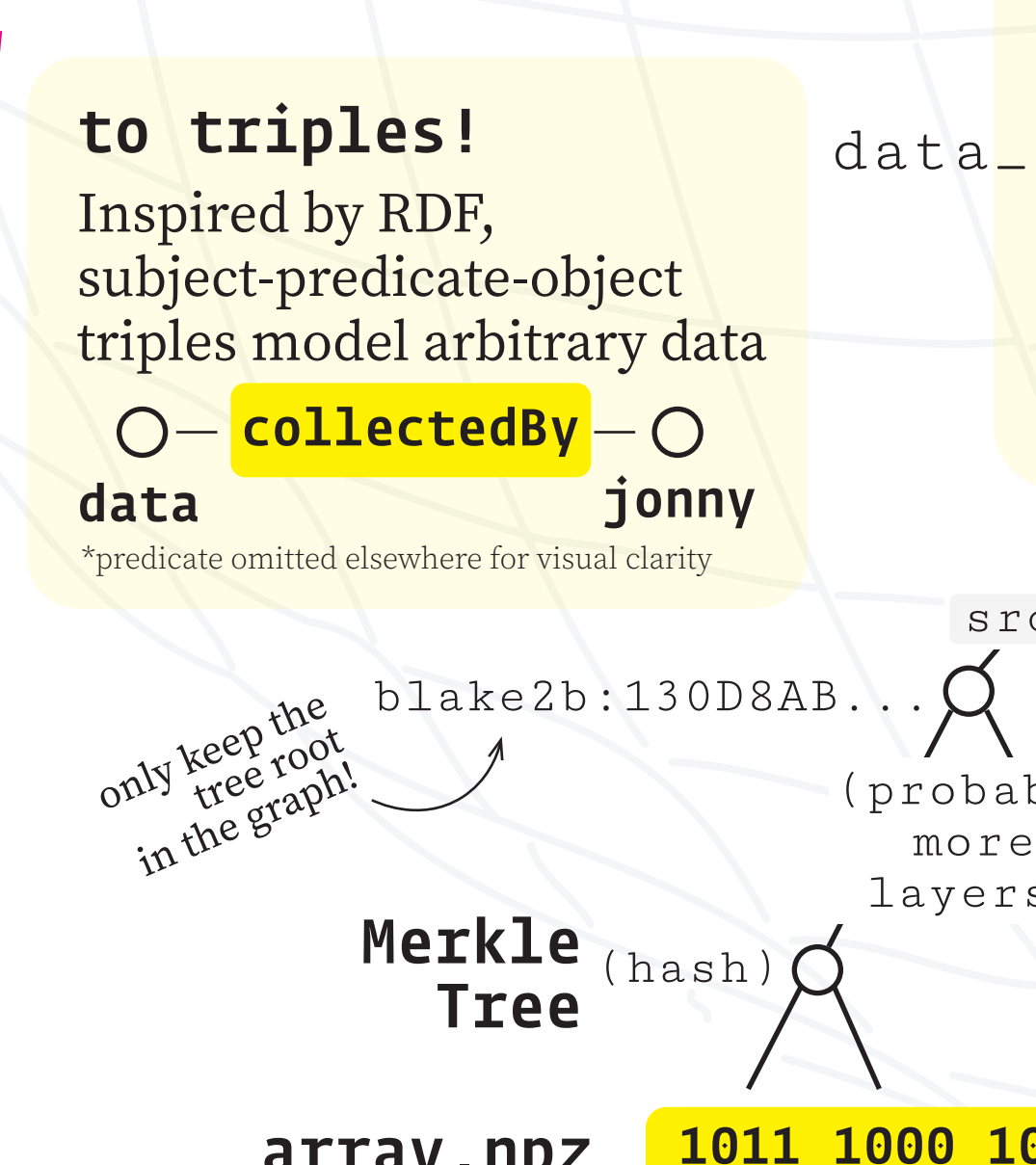
Vernacular Schema

Peers can make schemas as-needed, negotiating relationships between terms as a continuous, languagelike social process



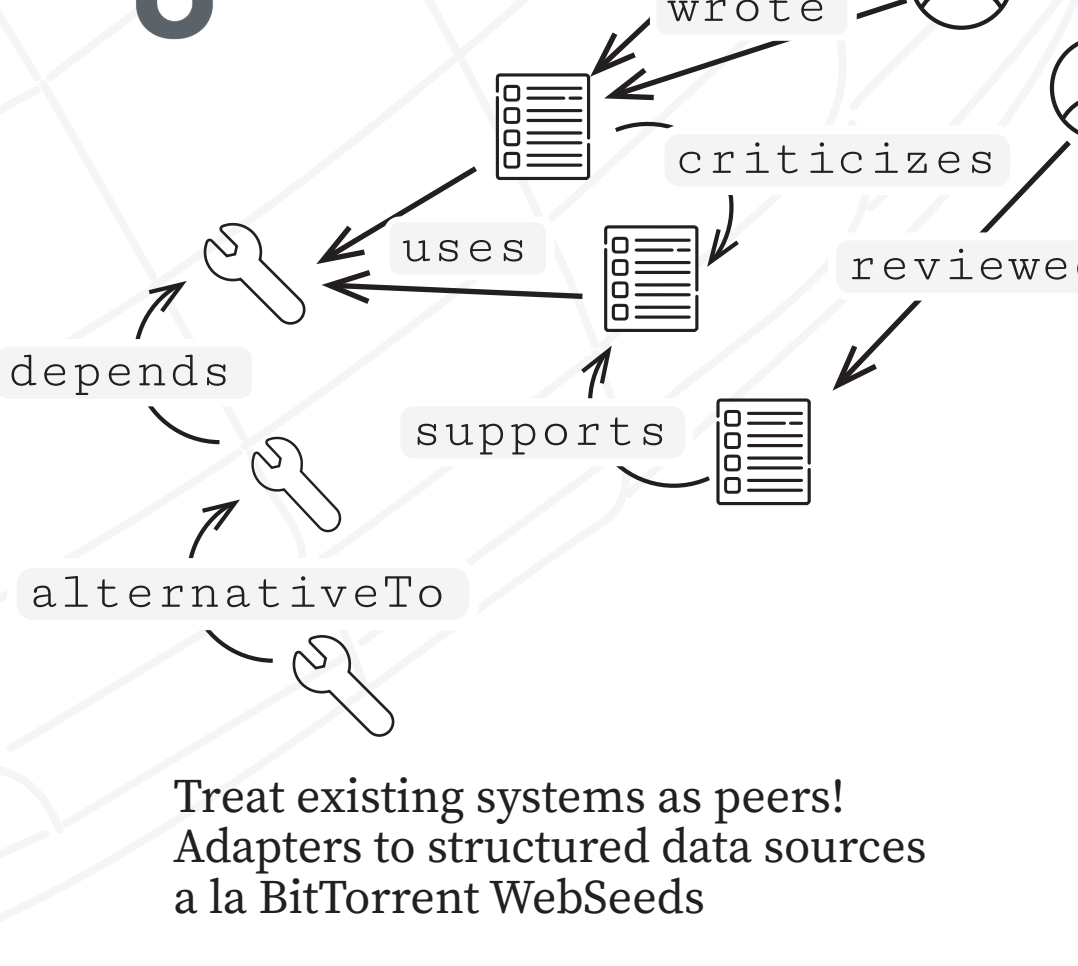
Keep Your Files!

Git is great because it slips invisibly underneath files. Bittorrent is great because .torrents are just files. Separating the metadata graph from binary files gives us content addressability + friendly UX



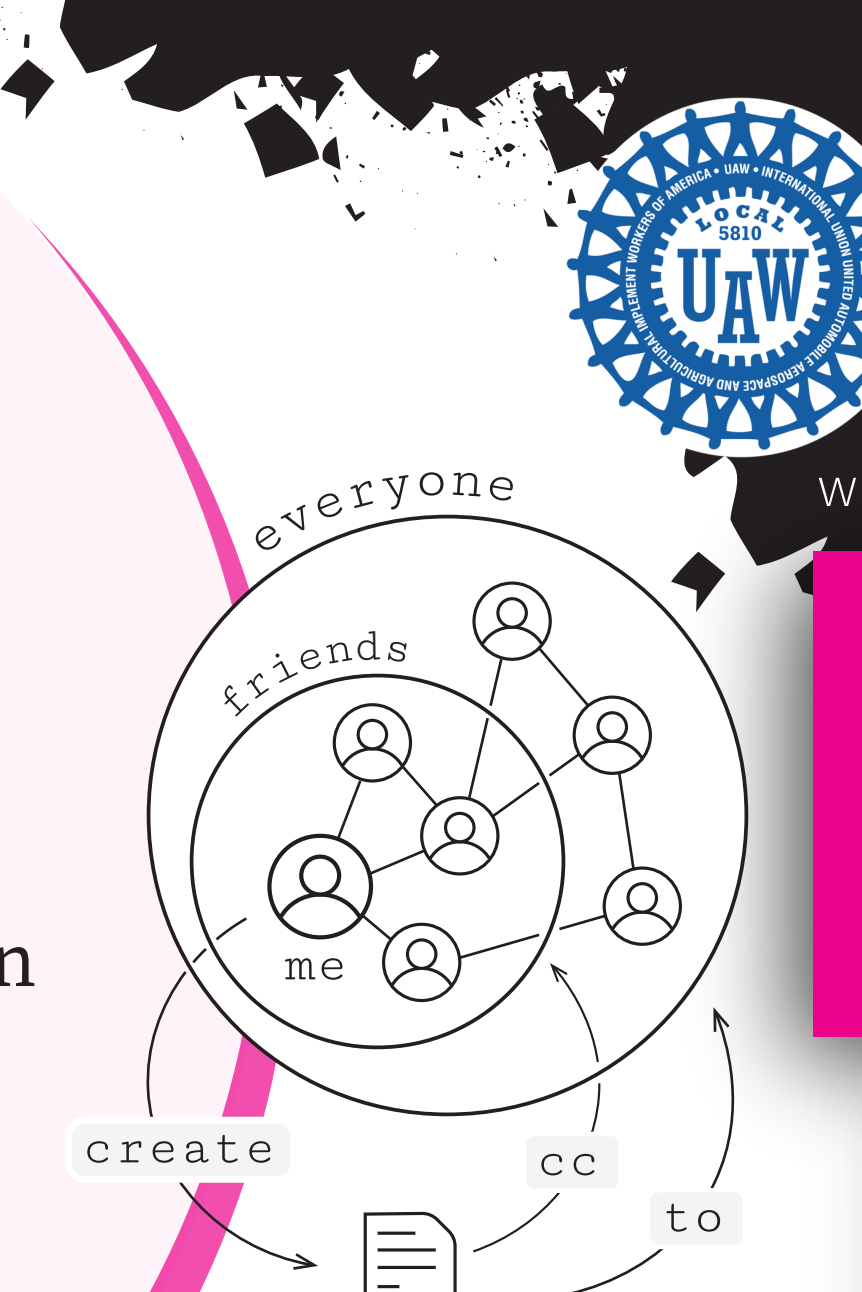
Radical Interoperability!

We don't need a new platform or journal, we need to glue them together. We'll be exploring...
- Direct export data from and import config to experimental tooling
- Crawlers to treat existing HTTP, RDBMS, S3, etc. resources as peers
- Delegated identity and discovery from existing eg. social media.



Public Spaces: Neuromatch.social

Twitter showed us how direly we need social infrastructure - what if it was built to support us rather than mine our attention? Neuromatch.social is an experiment in cooperatively governed scientific infrastructure - and a lovely, blossoming community. We run a fork of Mastodon, a federated social media server, and invite all members to hack on the software, propose policies, and take ownership as co-equal members. It will be the social heart of our next phase, federation to p2p.



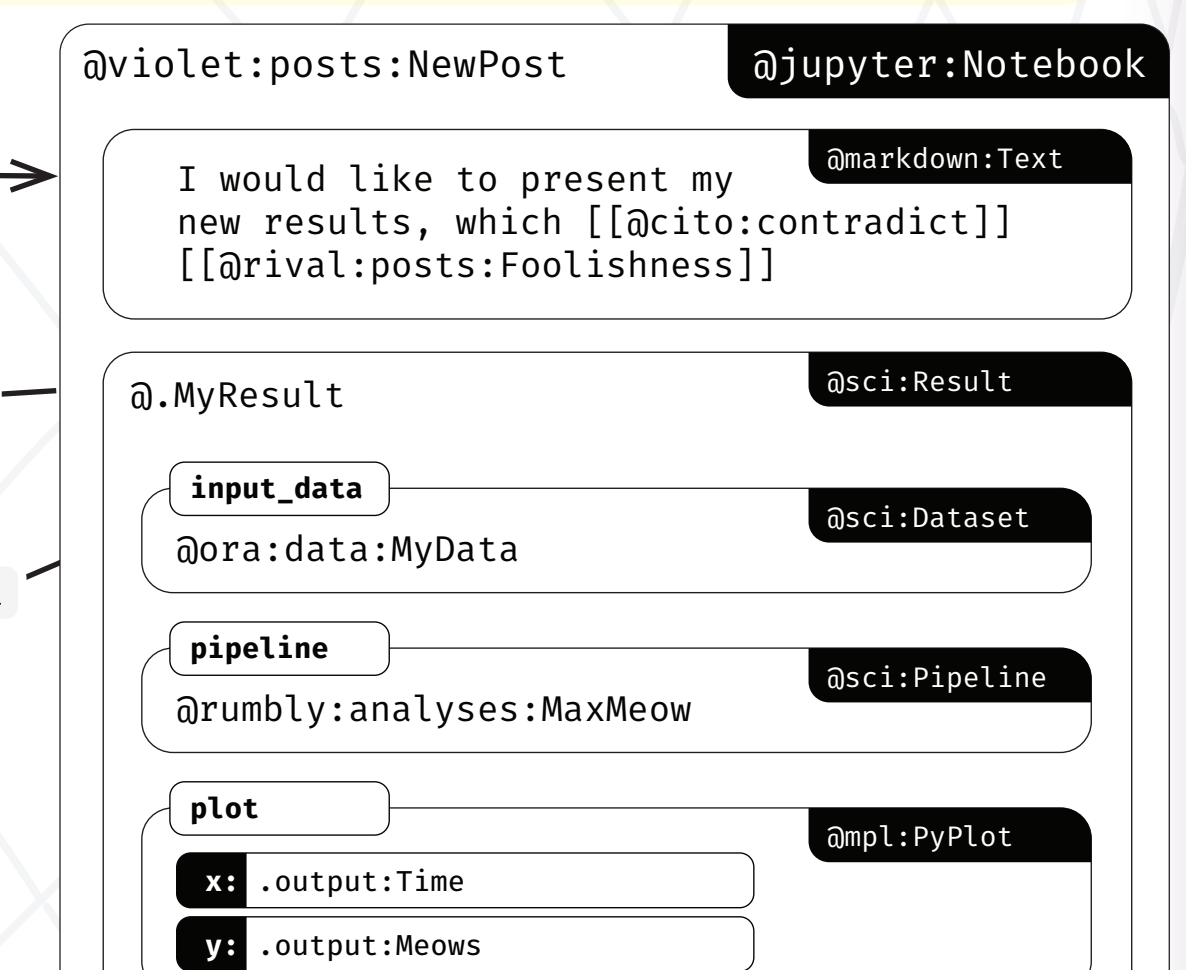
ActivityPub is already a triplet graph made of JSON-LD!

Local Organization: Wikis, Chat

Organizing labor as a million google docs and slack messages is not our best work. We wrote **chatbridge** to bridge proprietary chat silos with a few clicks, and **wikibot** to gradually enrich chat streams with wiki gardens. We've seen fluidly structured semantic wikis change patterns of work across organizations, and will be working on p2p wikis for labs, unions, any group organizing people and information.

Peer Review, Reuse, and Governance in Co-Op Mode

What if the alternative to journals was simply the existing social reality of research made visible? What's more impressive, an h-index, or the extended graph of your contributions to the commons? P2P can help realign incentives towards cooperation - the very act of reusing someone's work is to credit and rehost it. A continual process of contextualization isn't the end of peer review, quite the opposite: it's an opportunity for an unimaginably richer means of collaboratively evaluating work.



Jonny Saunders*, Daniel Aharoni*
*UCLA, Department of Neurology
*Institute of Pirate Technology
web: jon-e.net/sfn23
contact: j@nny.fyi

come talk to me about...

Distributed Identity

P2P systems are usually very lonely since peers typically have ephemeral or anonymous identities and don't form a persistent social graph. The culture of BitTorrent trackers shows us that social organization is determinative of successful distributed archives, but how will identity work in cleanest p2p? Can we bootstrap identity from existing systems? Can we reverse the anarcho-capitalist plague of zero-trust "web 3" with explicit networks of trust?

Vernacular Schema

A recurring failure mode of linked data information systems is a drift towards few, authoritative schema warded by a closed ontological priesthood[2]. Language doesn't work like that, though, and language rocks. Can we do that instead? Can we create an evolving system of computer-readable meaning? How shall we negotiate, fork, and merge infinite variations of expression? Can we use social graphs as context to infer what is meant without needing to get a complete URI involved every time?

Adversarial Interoperability

The only thing scientists love more than pining for better infrastructure is getting absolutely steamrolled by billion-dollar industries when we try. Surveillance publishers and cloud giants are not our friends. Friends don't turn calls for "open science" into new, more extractive, more inequitable models. How can anti-profitable technologies use the walled gardens without being captured by them? What openings might give us the upper hand? Can we build bridges for our colleagues stuck in information industry traps?

Cooperative Governance

Let's not get distracted amidst the graphs and technology: the problems of scholarly infrastructure are primarily social organizing problems. Organizational modalities carry their own logic. Traditional board-run nonprofits will hold listening sessions and mean well, but power always flows from the top of a hierarchy. P2P systems require us to cooperate, but can academics cooperate if it means genuinely ceding power? Can we move past shopping for Infra as a Service and believe in our collective power to make it ourselves?

Software	
nwb-linkml	- source: https://github.com/p2p-ld/nwb-linkml/ docs: https://nwb-linkml.readthedocs.io
chatbridge	- source: https://git.jon-e.net/jonny/chatbridge
neuromatch.social	- source: https://github.com/NeuromatchAcademy/mastodon docs: https://wiki.neuromatch.social
miniscope-io	- source: https://github.com/Aharoni-Lab/miniscope-io docs: https://miniscope-io.readthedocs.io
p2p-ld	- docs: https://piracy.solutions/docs/
autopilot	- source: https://github.com/auto-pi-lot/autopilot docs: https://docs.auto-pi-lot.com
wikibot	- source: https://git.jon-e.net/jonny/wiki-postbot/

References	
[1]	Saunders, J. (2022) "Decentralized Infrastructure for (Neuro-)Science" doi:10.48550/arXiv.2209.07493, https://jon-e.net/infrastructure
[2]	Saunders, J. (2023) "Surveillance Graphs" doi:10.17613/syv8-cp10, https://jon-e.net/surveillance-graphs