How standards and use-cases shape up the FAIR DANDI archive
or How FAIR is DANDI?

https://dandiarchive.org

Yaroslav O. Halchenko
Satrajit Ghosh
Michael Grauer
Ben Dichter
...

CC-BY

The DANDI Archive
The BRAIN Initiative archive for publishing and sharing cellular neurophysiology data

63 dandisets
154 users
4 species
1008 subjects
1551 cells
13 TB total data size
<table>
<thead>
<tr>
<th>MIT</th>
<th>Dartmouth</th>
<th>Kitware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satrajit Ghosh</td>
<td>Yaroslav O. Halchenko</td>
<td>Michael Grauer</td>
</tr>
<tr>
<td>Nicole Lo</td>
<td>John T. Wodder II</td>
<td>Brian Helba</td>
</tr>
<tr>
<td>Jakob Voigts</td>
<td>Matt Van der Meer</td>
<td>Jake Nesbitt</td>
</tr>
<tr>
<td>Mark Harnett</td>
<td></td>
<td>Daniel Chiquito</td>
</tr>
<tr>
<td>CatalystNeuro</td>
<td>Collaborators</td>
<td>Roni Choudhury</td>
</tr>
<tr>
<td>Ben Dichter</td>
<td>- Neurodata Without Borders</td>
<td></td>
</tr>
<tr>
<td>Luiz Tauffer</td>
<td>(nwb.org)</td>
<td></td>
</tr>
<tr>
<td>Alessio Buccino</td>
<td>- Brain Cell Data Center</td>
<td></td>
</tr>
<tr>
<td>Cody Baker</td>
<td>(BCDC/BICCN)</td>
<td></td>
</tr>
<tr>
<td>Daniel Sotoude</td>
<td>- Allen Institute for Brain</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science</td>
<td></td>
</tr>
<tr>
<td>Collaborators</td>
<td></td>
<td>- Gonçalo Lopes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Bonsai)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Project Jupyter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Ariel Rokem</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(U19 - U Washington)</td>
</tr>
</tbody>
</table>

AWS Public dataset program

Support

1R24MH117295

NIH The BRAIN Initiative

NIH National Institute of Mental Health
DANDI (https://dandiarchive.org) is an NIH BRAIN Initiative archive for scientists to share, collaborate, and process data from cellular neurophysiology experiments.
Multiscale, multispecies neuroscience

- Theories
  - Models
  - Experimental designs
- Driving biological applications
  - Clinical
  - Behavioral
  - Knowledge generation
- Tools
  - Instruments
  - Protocols
  - Analytics
- Data
  - Large (TBs) and Wide
  - Rich descriptors
  - Provenance

International Neuroinformatics Coordinating Facility (INCF)
DANDI works with BRAIN Initiative community data **standards:**
- Neurodata Without Borders (Ruebel, PI) (NWB; https://nwb.org)
- Support for BIDS, NIDM coming in 2021
  - Brain Imaging Data Structure (Poldrack, PI) (BIDS; https://bids.neuroimaging.io)
  - BEP031 (microscopy)
  - BEP032 (animal electrophysiology)
- Neuroimaging Data Model (Keator, PI) (NIDM; https://nidm.nidash.org)

... and aims to make harmonization easy:

**Validate**

```bash
$ dandi validate 000020/
000020/000020/dandiset.yaml: ok
000020/000020/sub-603513553/sub-603513553_ses-20170726T202952.nwb: ok
000020/000020/sub-603513553/sub-603513553_ses-20170726T211815.nwb: ok
000020/000020/sub-603513553/sub-603513553_ses-20170710T183148.nwb: ok
000020/000020/sub-603513553/sub-603513553_ses-20170710T192333.nwb: ok
000020/000020/sub-601462951/sub-601462951_ses-20170712T215236.nwb: ok
000020/000020/sub-601462926/sub-601462926_ses-20170712T173925.nwb: ok
000020/000020/sub-601462951/sub-601462951_ses-20170710T173925.nwb: ok
000020/000020/sub-601462915/sub-601462915_ses-20170713T174229.nwb: ok
000020/000020/sub-599387254/sub-599387254_ses-20170706T204704.nwb: ok
000020/000020/sub-601462926/sub-601462926_ses-20170712T152036.nwb: ok
000020/000020/sub-599387254/sub-599387254_ses-20170706T204704.nwb: ok
000020/000020/sub-601462915/sub-601462915_ses-20170713T174229.nwb: ok
000020/sub-601462915/sub-601462915_ses-20170713T174229.nwb: ok
000020/sub-599387254/sub-599387254_ses-20170706T204704.nwb: ok
000020/sub-601462915/sub-601462915_ses-20170713T174229.nwb: ok
000020/sub-601462926/sub-601462926_ses-20170712T152036.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T185554.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T173925.nwb: ok
000020/sub-601462915/sub-601462915_ses-20170713T192333.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T185554.nwb: ok
000020/sub-603513553/sub-603513553_ses-20170726T202952.nwb: ok
000020/sub-603513553/sub-603513553_ses-20170726T211815.nwb: ok
000020/sub-603513553/sub-603513553_ses-20170710T183148.nwb: ok
000020/sub-603513553/sub-603513553_ses-20170710T192333.nwb: ok
000020/sub-603513553/sub-603513553_ses-20170726T211815.nwb: ok
000020/sub-603513553/sub-603513553_ses-20170710T185554.nwb: ok
Summary: No validation errors among 11 file(s)
```

**Organize**

```bash
$ dandi organize -d 000020 -f symlink src_files
2020-05-12 16:21:45,288 [ INFO] Loading metadata from 10 files
Visit 000020/

$ ls 000020/
dandiset.yaml sub-599387254 sub-601462915 sub-601462926 sub-601462951 sub-603513553

$ ls 000020/sub-599387254/sub-599387254_ses-20170706T204704.nwb
```

... and aims to make harmonization easy:
DANDI establishes harmonized metadata schema to
- ensure all required metadata is provided in NWB and BIDS
- provide metadata serialization as
  - JSON-LD linked data
  - JSON records with json-schema for
    convenient User Interfaces
The DANDI Archive
The BRAIN Initiative archive for publishing and sharing cellular neurophysiology data.
More integrations: NWB Explorer

You can right click to get a direct download link to a file and use it directly with:

NWB Explorer from MetaCell/Open Source Brain
DANDI is built on open source technologies

Web portal
https://dandiarchive.org

Jupyterhub
https://hub.dandiarchive.org

* need to register with DANDI

Github organization
https://github.com/dandi

Contact
Register on DANDI and we will invite you to the DANDI Slack workspace or email: info@dandiarchive.org

Questions
https://neurostars.org or Slack
So how FAIR is DANDI?

- **Findable**
  - Rich schema with use of standard ontologies
  - All elements (dandiset and assets) carry unique identifiers
  - Searchable on dandiarchive.org and LD will be exposed

- **Accessible**
  - Standard containers (NWB, NIfTI, TIFF, etc)
  - Metadata is exported as JSON/JSON-LD alongside with data
  - Standard protocols (HTTPS), integration with external services
  - API, Python, CLI, DataLad, ROS3 HDF5

- **Interoperable**
  - JSON + json-schema, JSON-LD
  - schema.org, spdx.org (licenses), PROV
    - wasGeneratedBy, wasDerivedFrom, ...

- **Reusable**
  - License is to be specified
  - Rich metadata schema for cellular neurophysiology data
  - *Everything DANDI is released under Free and Open Source licenses*
Upcoming features

- Interact via a stable versioned API
- Publish with DANDI and mint DOIs for your dandisets
  - With Datacite metadata support
- UI/UX improvements
- Hub integration of more NWB applications
- Openephys/Bonsai integration with DANDI
- Better metadata management and search
- Support for timed embargo upto 1 year

Technical roadmap:

2021:
- API/Search
- Embargo

2022:
- Raw + Processed
- Provenance

2023:
- Local deployment
- DANDI Federation

2024:
- X-archive
- Interoperability
The end of 10,000 ft overview of the FAIR DANDI

Thank you