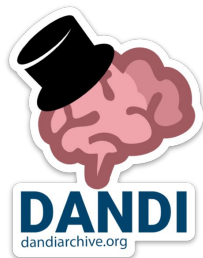


How standards and use-cases shape up the FAIR DANDI archive or How FAIR is DANDI?

<https://dandiarchive.org>



Yaroslav O. Halchenko

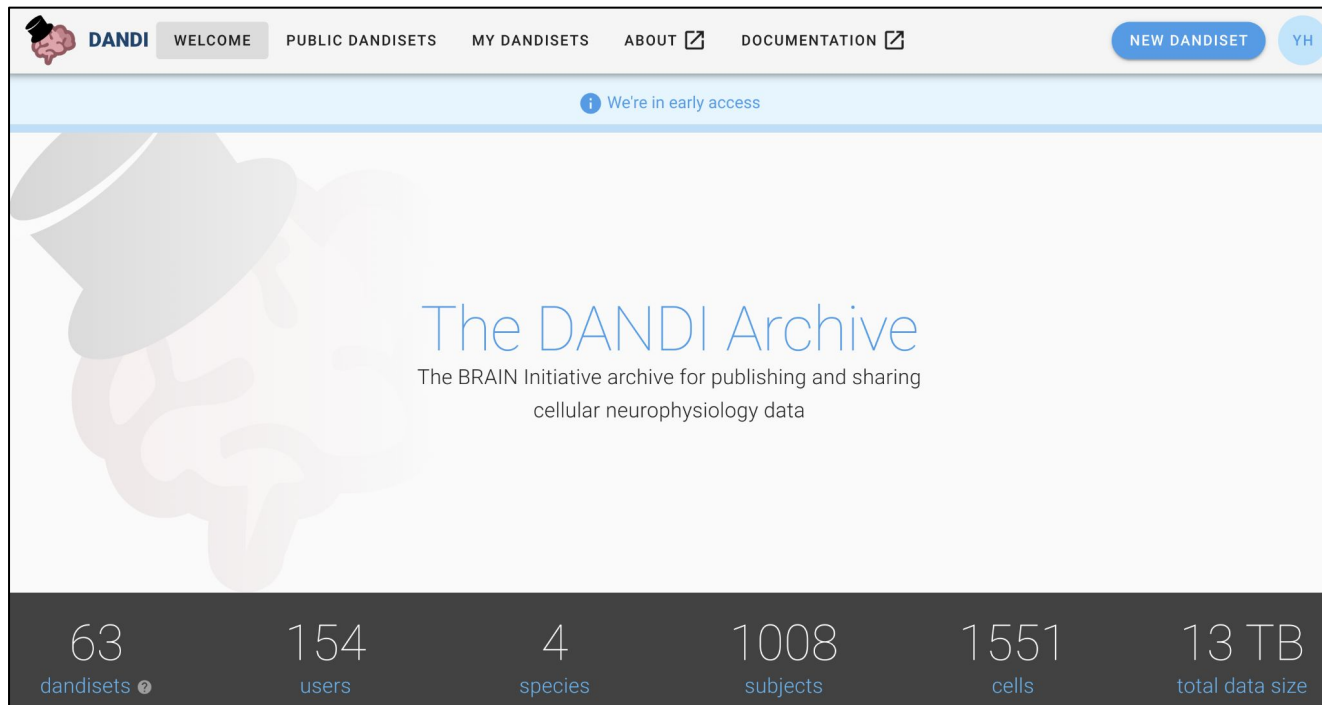
Satrajit Ghosh

Michael Grauer

Ben Dichter

...

CC-BY



MIT

- Satrajit Ghosh
- Nicole Lo
- Jakob Voigts
- Mark Harnett

Dartmouth

- Yaroslav O. Halchenko
- John T. Wodder II
- Matt Van der Meer

Kitware

- Michael Grauer
- Brian Helba
- Jake Nesbitt
- Daniel Chiquito
- Roni Choudhury

CatalystNeuro

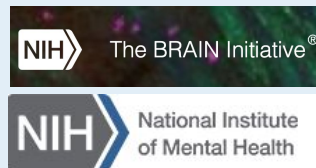
- Ben Dichter
- Luiz Tauffer
- Alessio Buccino
- Cody Baker
- Daniel Sotoude

Collaborators

- Neurodata Without Borders (nwb.org)
- Brain Cell Data Center (BCDC/BICCN)
- Allen Institute for Brain Science

- Gonçalo Lopes (Bonsai)
- Project Jupyter
- Ariel Rokem (U19 - U Washington)

Support



1R24MH117295

AWS Public
dataset program

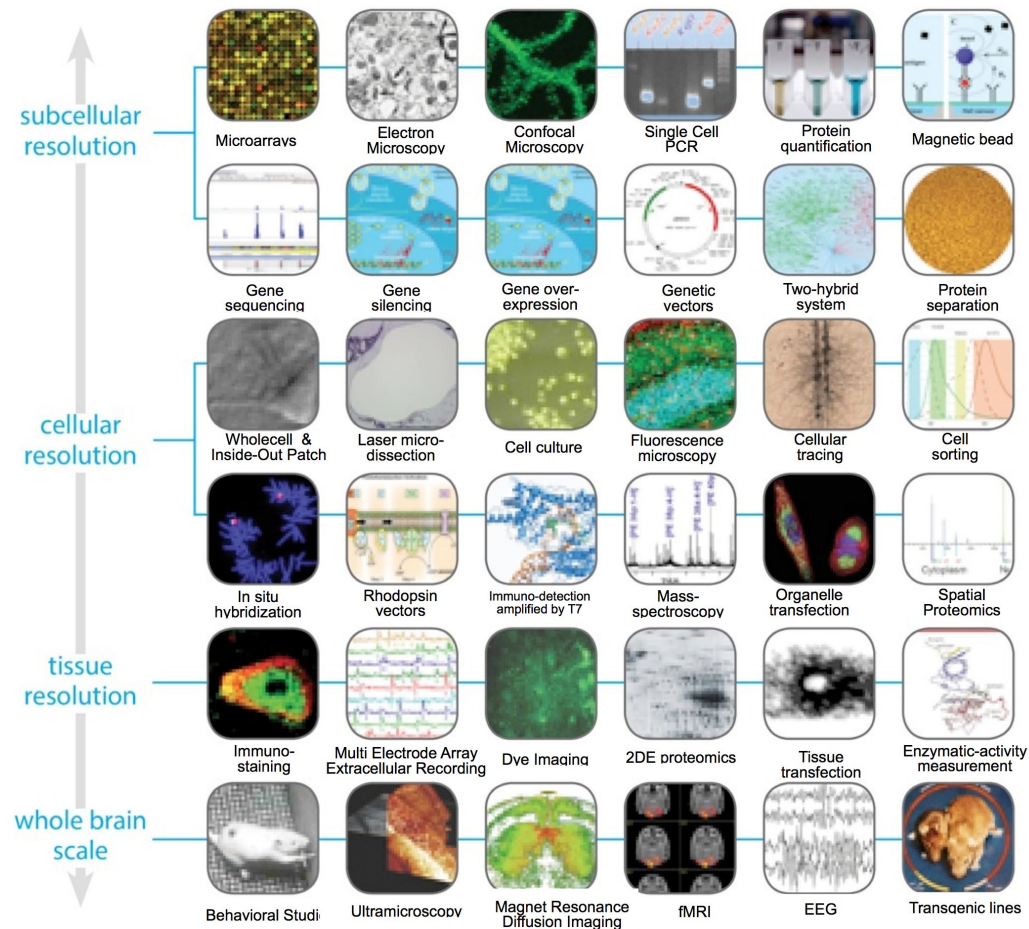
DANDI (<https://dandiarchive.org>)
is an NIH BRAIN Initiative archive
for scientists to share, collaborate, and process
data from cellular neurophysiology experiments.

Infrastructure

- Dissemination, Search, Visualization, Computation
- Collaboration and coordination
- FAIRness and efficiency

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:

Organize

```
$ dandi organize -d 000020 -f symlink src_files
2020-05-12 16:21:45,288 [ INFO] Loading metadata from 10 files
2020-05-12 16:21:54,756 [ INFO] Organized 10 paths.
Visit 000020/
```

```
$ ls 000020/
dandiset.yaml          sub-599387254          sub-601462915 sub-60146292
                        sub-601462951          sub-603513553
```

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

Validate

```
$ dandi validate 000020/
000020/dandiset.yaml: ok
000020/sub-603513553/sub-603513553_ses-20170726T202952.nwb: ok
000020/sub-603513553/sub-603513553_ses-20170726T211815.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T173925.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T200802.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T185554.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T183148.nwb: ok
000020/sub-601462951/sub-601462951_ses-20170710T192333.nwb: ok
000020/sub-601462926/sub-601462926_ses-20170712T215236.nwb: ok
000020/sub-599387254/sub-599387254_ses-20170706T204704.nwb: ok
000020/sub-601462915/sub-601462915_ses-20170713T174229.nwb: ok
Summary: No validation errors among 11 file(s)
```

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



NEURODATA
WITHOUT BORDERS



BIDS
BRAIN IMAGING DATA STRUCTURE



github.com/dandi/dandi-cli/blob/master/dandi/models.py

```
class CommonModel(DandibaseModel):
    schemaVersion: str = Field(
        default=DANDI_SCHEMA_VERSION, readOnly=True, nskey="schema"
    )
    name: Optional[str] = Field(
        None,
        title="Title",
        description="The name of the item.",
        max_length=150,
        nskey="schema",
    )
    description: Optional[str] = Field(
        None, description="A description of the item.", nskey="schema"
    )
    contributor: Optional[List[Union[Person, Organization]]] = Field(
        None,
        title="Contributors",
        description="Contributors to this item.",
        nskey="schema",
    )
    about: Optional[List[Union[Disorder, Anatomy, TypeModel]]] = Field(
        None,
```

gui-beta-dandiarchive.org/netlify/app/16/dandiset/000003

WELCOME PUBLIC DANDISETS MY DANDISETS ABOUT DOCUMENTATION NEW DANDISET

Layer-Specific Physiological Features and Interlaminar Interactions in the Primary Visual Cortex of the Mouse

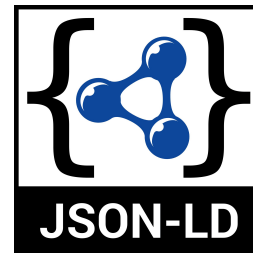
Click a field below to edit it.

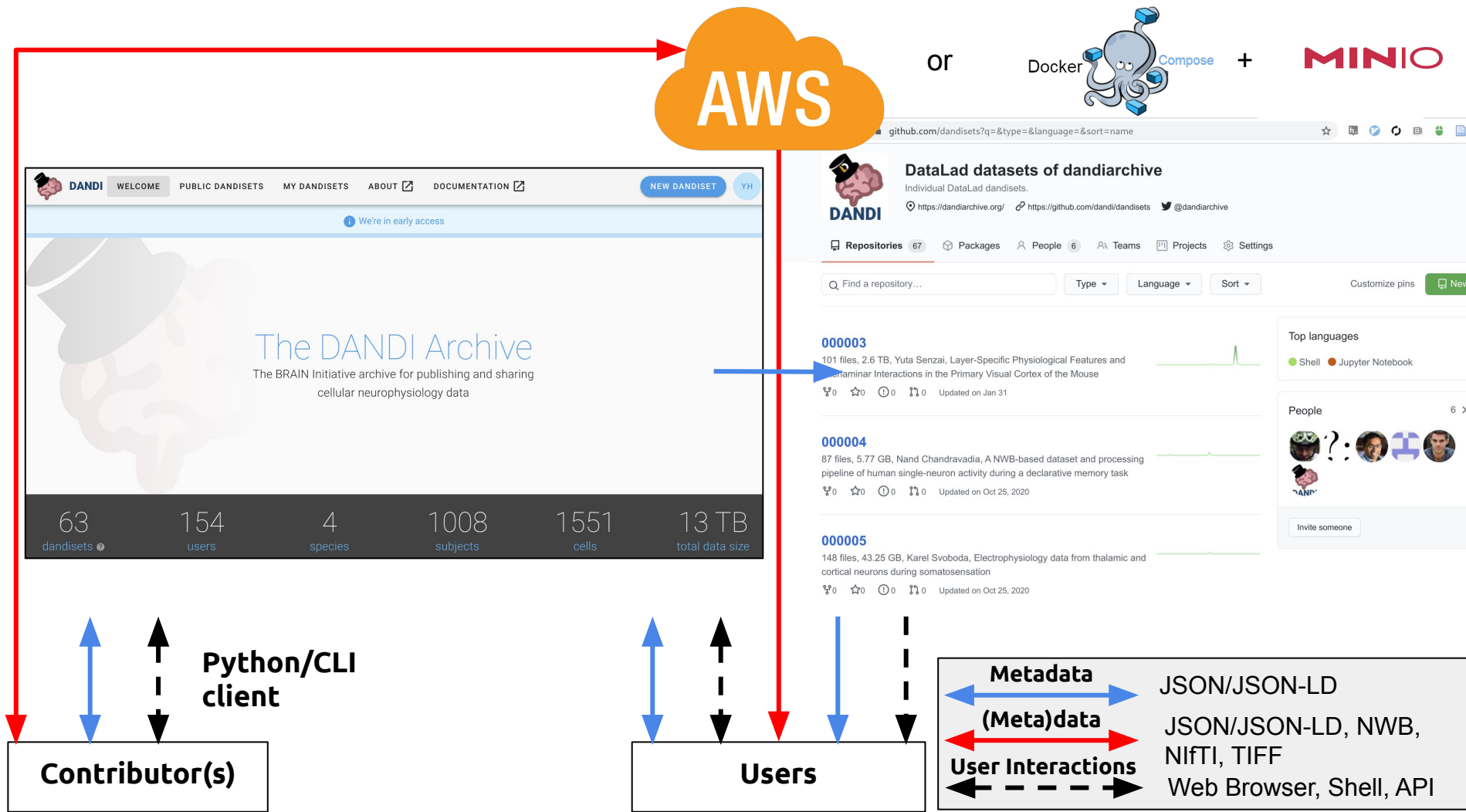
DANDISET CONTRIBUTORS SUBJECT MATTER ETHICS APPROVAL ACCESS TYPE

RELATED RESOURCE NAME OF THE PROJECT

Dandiset Title
Layer-Specific Physiological Features and Interlaminar Interactions in the Primary Visual Cortex of the Mouse

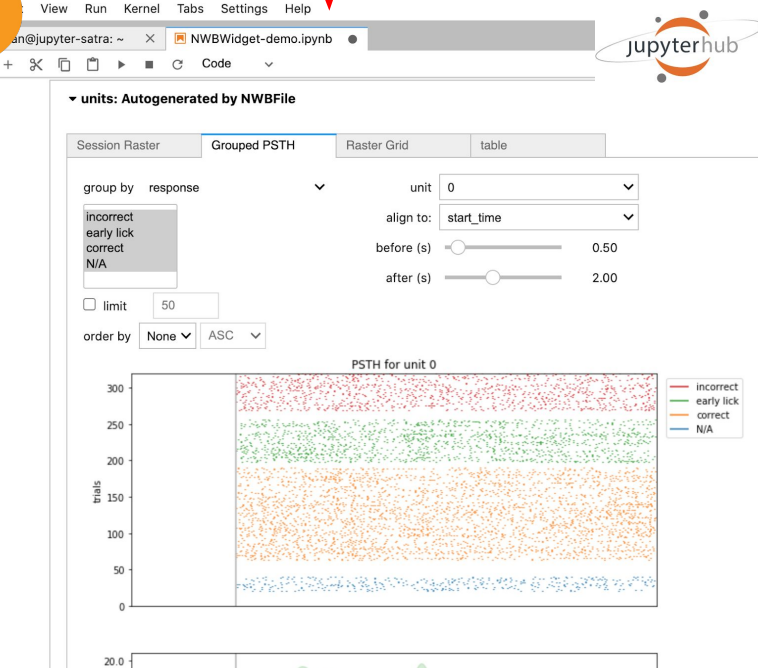
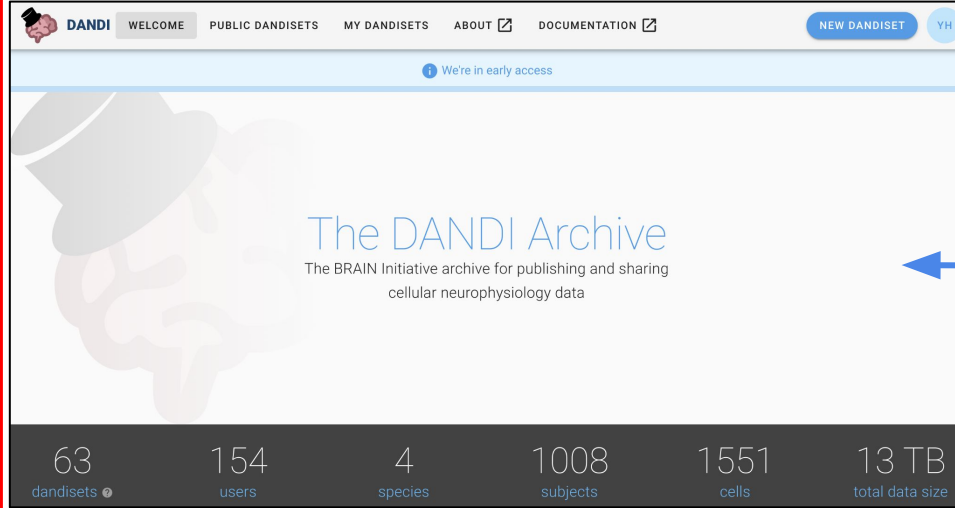
Description
Data from "Layer-Specific Physiological Features and Interlaminar Interactions in the Primary Visual Cortex of the Mouse" Senzai, Fernandez-Ruiz, Buzsaki, Neuron 2019. Electrophysiology recordings of hippocampus during theta maze exploration.





<https://hub.dandiarchive.org/>

AWS



Python/CLI
client

Contributor(s)

Users

Metadata

(Meta)data

User Interactions

JSON/JSON-LD

JSON/JSON-LD, NWB,
NifTI, TIFF

Web Browser, Shell, API

More integrations: NWB Explorer

NWB Explorer_{beta}

The screenshot shows the NWB Explorer beta web application. The interface is divided into several sections:

- General** (selected):
 - Session Description**: New/Old recognition task for ID: 7.
 - Identifier**: H10_7
 - Electrodes**: Description: metadata about extracellular electrodes
 - Subject**: Age: 37, Description: Left Lateral Frontal, Sex: M, Species: Human, Subject id: P10HMH
 - Units**: Description: Autogenerated by NWBFile, Waveform unit: volts
 - Experiment Description**: The data contained within this file describes a new/old recognition task performed in patients with intractable epilepsy implanted with depth electrodes and Behnke-Fried microwires in the human Medical Temporal Lobe (MTL).
 - Lab**: Rutishauser
 - Institution**: Hunigton Memorial Hospital
 - Data Collection**: learning: 80, recognition: 81
 - Related Publications**: Faraut et al. 2018, Scientific Data; Rutishauser et al. 2015, Nat Neurosci;
- Stimulus** (selected):
 - Filter**: (empty search bar)
 - acquisition.events**: AnnotationSeries. The events correspond to the TTL markers for each trial. For the learning trials, the TTL markers are the following: 55 = start of the experiment, 1 = stimulus ON, 2 = stimulus OFF, 3 = Question Screen Onset ("Is this an animal?"), 20 = Yes (21 = NO) during learning, 6 = End of Delay after Response, 66 = End of Experiment. For the recognition trials, the TTL markers are the following: 55 = start of experiment, 1 = stimulus ON, 2 = stimulus OFF, 3 = Question Screen Onset ("Have you seen this image before?"), 31:36 = Confidence (Yes vs. No) response [31 (new, confident), 32 (new, probably), 33 (new, guess), 34 (old, guess), 35 (old, probably), 36 (old, confident), 66 = End of Experiment]
 - 2 Matching Results**
- Python**:
 - units: Autogenerated by NWBFile**
 - Session Raster** (selected):
 - group by**: None
 - units**: (vertical slider from 0 to 31)
 - limit**: 50
 - order by**: None
 - Grouped PSTH**: (selected)
 - Raster Grid**: (selected)
 - table**: (selected)
 - window (s)**: 6706.88
 - duration (s)**: 5
 - units**: (raster plot showing spikes over time for units 0-31)

	drafts / 000004 / sub-P14HMH	
☒	sub-P14HMH_ses-20070601_obj-118wrds_ecephys+image.nwb	69.80 MB
☐	sub-P14HMH_ses-20070601_obj-7studd_ecephys+image.nwb	52.70 MB

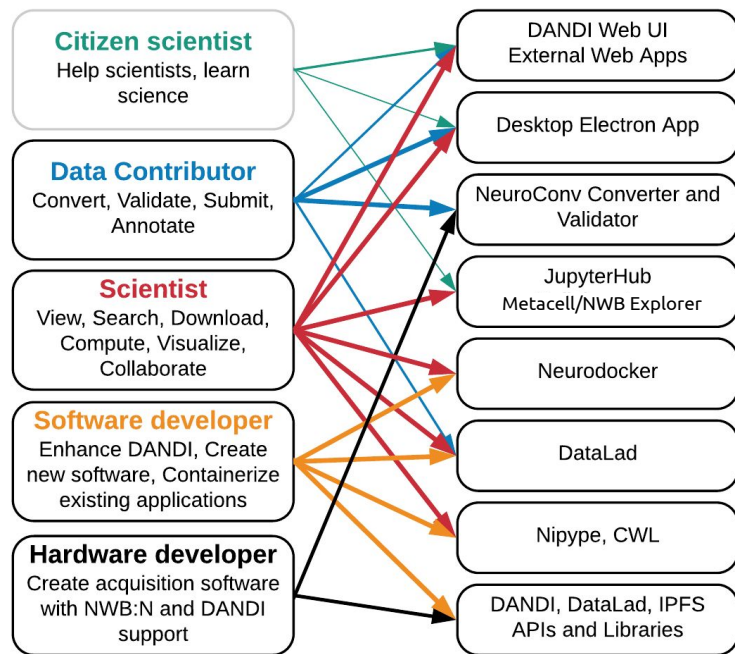
Rows per page: 25 1-2 of 2 < >

You can right click to the 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](https://dandiarchive.org) 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**

Dandiset. n.

An organized collection of assets (files) with both file level and dataset level metadata generated from an experiment or a project.

Technical roadmap:

2021:

- API/Search
- Embargo

2022:

- Raw + Processed
- Provenance

2023:

- Local deployment
- DANDI Federation

2024:

- X-archive
interoperability

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

The end of 10,000 ft overview of the FAIR DANDI

Thank you