NWB User Days 2023

Using the DANDI Archive

For Producers of Data

- Making a well-annotated dandiset

- How to ensure your files follow 'Best Practices'

For Consumers of Data

- How to find candidate data on DANDI for re-analysis

- How to perform quick quality checks on an entire DANDI set

- How to stream data directly from the DANDI Archive

For Both

- How to update analysis scripts to read from the NWB versions of the data

- How to use the **free** DANDI Hub compute resources

How to use the **free** DANDI Hub compute resources

- Step 1: Make an account on https://dandiarchive.org
  - 1a: Requires a GitHub account
  - 1b: Wait for account approval (~1 day)

- Step 2: Go to https://hub.dandiarchive.org/
  - 2a: Sign in
  - 2b: Select instance size (amount of resources)
  - 2c: Wait a couple of minutes

You're in!

Making a well-annotated Dandiset

Register a new dataset

Name*

Usually the name of my publication

Provide a title for this dataset

Description*

Usually the abstract of my publication; if pre-publication, basic description of the type of experiments.

Provide a description for this dataset

☐ Embargo this dataset   (What is this?) ⓘ

License*

spdx:CC-BY-4.0

A metabolic function of the hippocampal sharp wave-ripple

ID: **000233** | DRAFT

👤 Contact **Tingley, David**

📄 File Count **345**

🗐 File Size **12.3 TB**

📅 Created **April 14, 2022**

🕘 Last update **May 24, 2022**

⚡ Licenses: **spdx:CC-BY-4.0**

👤? Access Information: **dandi:OpenAccess**

The hippocampus has previously been implicated in both cognitive and endocrine functions. We simultaneously measured electrophysiological activity from the hippocampus and interstitial glucose concentrations in the body of freely behaving rats to identify an activity pattern that may link these disparate functions of the hippocampus. Here we report that clusters of sharp wave-ripples recorded... [+ see more]

Dandiset Actions

- ⬇ DOWNLOAD ⌄
- ❞ CITE AS ⌄
- 📁 FILES
- 📄 METADATA
- 📁 MANIFEST

SHARE

✅ 💾 ↶ ↷ ⬇

| GENERAL | DANDISET CONTRIBUTORS | SUBJECT MATTER OF THE DATASET | ETHICS APPROVALS | RELATED RESOURCE | ASSOCIATED PROJECTS |

Dandiset title

A metabolic function of the hippocampal sharp wave-ripple ℹ

Description

The hippocampus has previously been implicated in both cognitive and endocrine functions. We simultaneously measured electrophysiological activity from the hippocampus and interstitial glucose concentrations in the body of freely behaving rats to identify an activity pattern that may link these disparate functions of the hippocampus. Here we report that clusters of sharp wave-ripples recorded from the hippocampus reliably predicted a decrease in peripheral glucose concentrations within about 10 min. This correlation was not dependent on circadian, ultradian or meal-triggered fluctuations, could be mimicked with optogenetically induced ripples in the hippocampus (but not in the parietal cortex) and was attenuated to chance levels by pharmacogenetically suppressing activity of the lateral septum, which is the major conduit between the hippocampus and the hypothalamus. Our findings demonstrate that a function of the sharp wave-ripple is to modulate peripheral glucose homeostasis, and offer a mechanism for the link ℹ

Study Target ℹ

License
spdx:CC-BY-4.0 ⌄ ℹ

Protocol ℹ

Keywords
glucose ✕ ecephys ✕ pharmacology ✕ ℹ

Acknowledgement

Be generous with assigning contributors - these don't just have to be the main authors of the citation

The 'role' can take on many values…

Remember to link to ORCID and ROR affiliations for maximal findability!

Be as precise as possible by linking subject matter with specific ontologies

Identifier
https://doi.org/10.1038/s41586-021-03811-w

A title of the resource
A metabolic function of the hippocampal sharp wave-ripple

URL of the resource
https://www.nature.com/articles/s41586-021-03811-w

Name of the repository
Nature

Resource relation
dcite:IsDescribedBy

━ A metabolic function of the hippocampal sharp wave-ripple   ⊖ REMOVE   ✎ EDIT

And, of course, link to the publication…

1. Make NWB files
2. Inspect NWB files
```
nwbinspector path/to/your/folder
```
3. Fix/improve as much as you are able

4. Download empty dandiset
```
dandi download DANDI:<six-digit-id>
```
5. Organize
```
cd <six-digit-id>
dandi organize ../path/to/your/folder
```
6. Validate
```
dandi validate path/to/your/folder
```
7. Upload
```
cd <six-digit-id>
dandi upload
```

How to ensure your files follow 'Best Practices'

How to perform quick quality checks on an entire DANDI set

NWB Inspector: A command-line tool for scanning NWB files

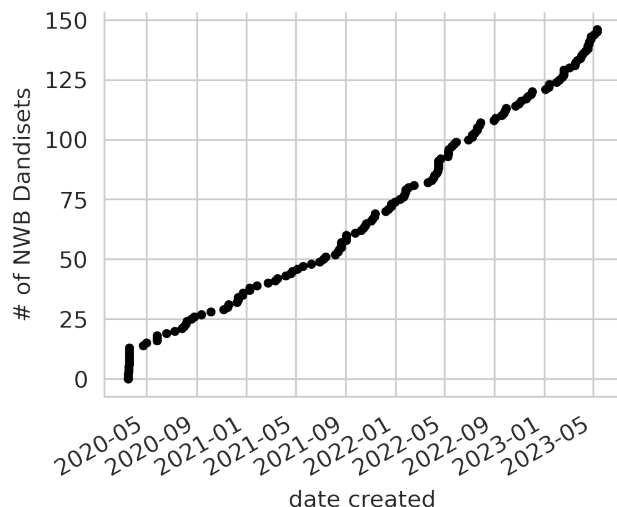Cody Baker
Ryan Ly
Oliver Ruebel
Ben Dichter

NEURODATA
WITHOUT BORDERS

THE
KAVLI
FOUNDATION

CATALYST
NEURO

Inspecting NWB files submitted to DANDI

NWB DANDI submissions continue to rise

- Passing schema validation does not mean the data is easily re-usable

- Some files have had data engineering mistakes *e.g.*, incorrect data orientation for any `TimeSeries` subtype

- Many datasets would benefit from more complete metadata

- Users can read and follow NWB Best Practices…

  ▸ …a lot of work

  ▸ …constantly updated using DANDI recommendations

Checking NWBFiles for Best Practices - Manual

As the [NWB Best Practices](#) list grows in size and complexity…

- It becomes more difficult for individuals converting to NWB to read it all and manually ensure compliance

- More often, professional reviewers of NWB files and Dandisets (*i.e.*, Ben Dichter, Satrajit Ghosh, *et al.*)
  - manually inspect each dataset
  - identify areas of improvement
  - communicate these to the original creators

- This can be a very time consuming process, especially as the upload rate of NWB files is growing every week

NWB Inspector

- **NWB Inspector** automates the task and provides immediate, autonomous feedback

- Implements a collection of "checks" that map to the NWB Best Practices document

- Checks are categorized by importance:

  ▸ CRITICAL: potentially incorrect data (9 checks)

    ❌ will prevent DANDI upload and publication

  ▸ BEST PRACTICE VIOLATION: very suboptimal data representation (20 checks)

  ▸ BEST PRACTICE SUGGESTION: improvable data representation (15 checks)

CLI Usage

- Installation in any Python 3.7-3.10 environment - current release is **v0.4.29**

  ```
  pip install nwbinspector
  ```

- Simple command line usage

  ```
  nwbinspector path/to/your/data.nwb

  nwbinspector path/to/your/folder
  ```

CLI: Human-readable Output

```
*****************************************************
NWBInspector Report Summary

Timestamp: 2022-04-14 18:08:40.949499+00:00
Platform: Linux-5.11.0-1019-aws-x86_64-with-glibc2.31
NWBInspector version: 0.3.11

Found 307 issues over 39 files:
      43 - CRITICAL
      55 - BEST_PRACTICE_VIOLATION
     209 - BEST_PRACTICE_SUGGESTION
*****************************************************


0   CRITICAL
===========

0.0   check_timestamps_match_first_dimension
----------------------------------------------

0.0.0   sub-Cori/sub-Cori_ses-20161214T120000.nwb and 38 other files: 'IntervalSeries' object at location '/processing/behavior/BehavioralEpochs/wheel_moves'
        Message: The length of the first dimension of data does not match the length of timestamps.

1   BEST_PRACTICE_VIOLATION
==========================

1.0   check_negative_spike_times
--------------------------------

1.0.3   sub-Cori/sub-Cori_ses-20161218T120000.nwb and 9 other files: 'Units' object with name 'units'
        Message: This Units table contains negative spike times. Time should generally be aligned to the earliest time reference in the NWBFile.
```

Streaming on DANDI

- No need to download an entire DANDI set ahead of time
  - Stream directly from the archive with the flag
    `--stream`
  - Easiest way to run this is to log into the DANDI Hub
    - ➢ Default terminal is already installed and configured
- When running over a large DANDI set…
  - Output can be dense (even with aggregation), so you can save the report with
    `--report-file-path ./save_my_report.txt`
  - Scan can be 'relatively' slow, so you can speed it up with
    `--n-jobs x` where `x` is an integer (`-1` will use all available processes)
    - ➢ Spawn the Hub with a 'large' instance for turbo speeds!
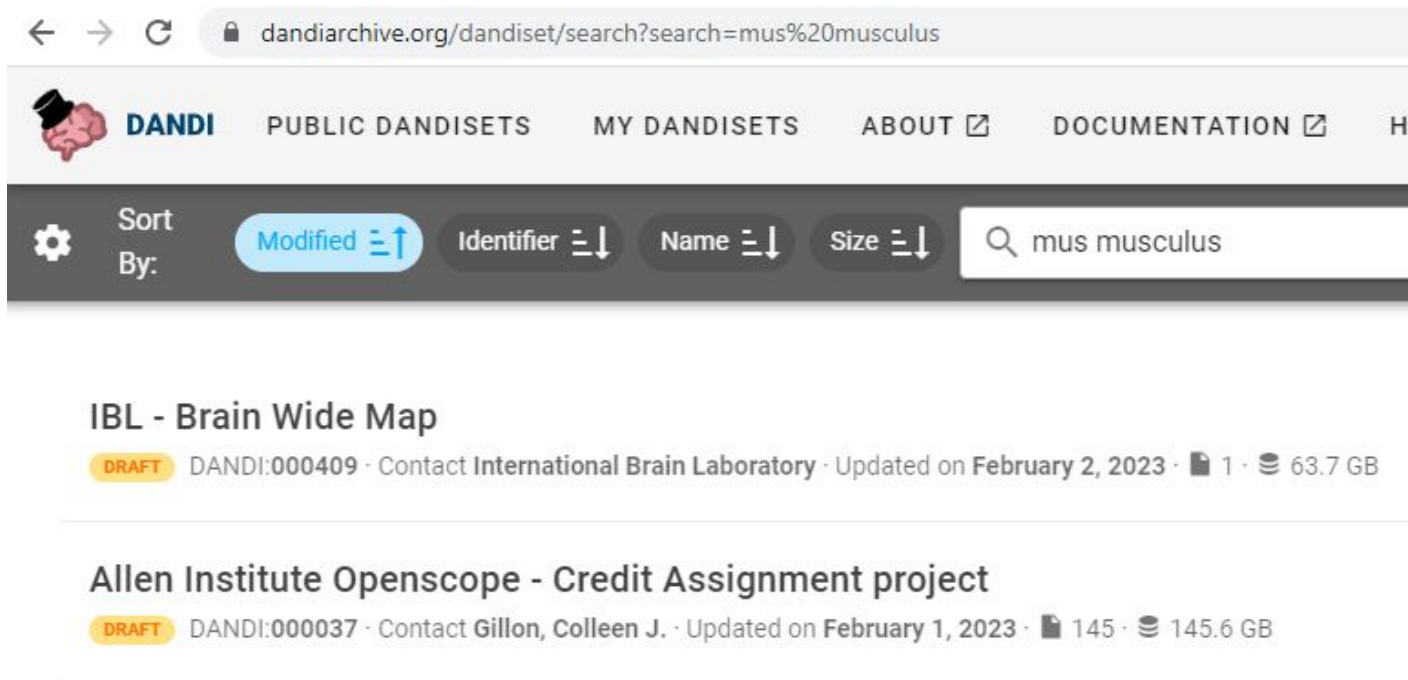
How to find candidate data on DANDI for re-analysis
How to stream data directly from the DANDI Archive

DANDI Search and Access

From the Web Interface

■ From the main page https://dandiarchive.org/dandiset, type a query in the search bar

From the Web Interface

- These top-level queries look at the highest level of Dandiset metadata
  - Titles, keywords, dandiset description
  - Automatically extracted subject species, modality, and techniques
  - You can see these fields listed at the bottom of the main page of any dandiset

Coming soon - advanced search

■ Early unreleased version of advanced search criteria:
https://gui-staging.dandiarchive.org/search

From the Web Interface

- ■ Good for…
  - A quick glance at general content

  - Casual browsing for any data of relevance to your hypotheses

  - Finding the dandiset for a particular publication
    - ➢ usually the title, or linked as a Related Resource ('IsDescribedBy')

  - Finding all the dandisets belonging to a particular lab
    - ➢ search by name of a 'Contributor'

  - Finding all the dandisets that use a particular species
    - ➢ Latin binomial, *e.g.*: *Mus musculus*, *Rattus norvegicus*, etc.

From the Web Interface

- ■ Doesn't yet help with…
    - Presence or absence of identified brain regions
        - ➢ unless the creator of the dandiset set these as keywords or 'subject matter'

    - Presence or absence of exact brain coordinates (like Allen CCFv3)

    - Identification of vast variety of behavioral techniques
        - ➢ open exploration vs. maze task
        - ➢ virtual reality vs. simple stimulus presentation
        - ➢ trialized tasks or spontaneous events
        - ➢ and many, many more…

Using the DANDI API in Python

- For finer-grain searchability, we can use the dandi-api to scan the metadata.yml files to programmatically obtain information

  On a local installation of Python (preferably in a fresh environment…)

  ```
  pip install dandi jupyter
  jupyter notebook
  ```

  Or just open up a notebook on the DANDI Hub

- [DANDI search tutorials](#)

Reading directly from an identified file

- Once you have concluded your investigation and found some NWB files of interest, you can either…

  - ➢ download them locally via command-line

    ```
    dandi download DANDI:<six-digit-ID>  # Will download all files
    ```

Or
    ```
    dandi download <copy and paste individual file URL>
    ```

    Then in Python (script or notebook)...

    ```
    from pynwb import NWBHDF5IO


    io = NWBHDF5IO(path=".../path_to_single_file.nwb", load_namespaces=True)

    nwbfile = io.read()
    ```

Reading directly from an identified file

■ Once you have concluded your investigation and found some NWB files of interest, you can either…

➢ download them locally via command-line

```
dandi download DANDI:<six-digit-ID>  # Will download all files
```

Or
```
dandi download <copy and paste individual file URL>
```

Then in MATLAB…

```
%% With MatNWB downloaded and added to your MATLAB session path…

nwbfile = nwbRead('intro_tutorial.nwb')
```

Reading directly from an identified file

■ Once you have concluded your investigation and found some NWB files of interest, you can either…

➢ Or stream directly (most recommended for one-off analyses or quick calculations)

In Python…

```python
import h5py
import fsspec
from fsspec.implementations.cached import CachingFileSystem
from pynwb import NWBHDF5IO

s3_url = "<copy and paste S3 URL from DANDI>" # Or ask the nwbinsector API for it
cfs = CachingFileSystem(
    fs=fsspec.filesystem("http"), cache_storage="some/temporary/folder",
)
file_system = cfs.open(s3_url, "rb")
file = h5py.File(file_system)
io = NWBHDF5IO(file=file, load_namespaces=True)
nwbfile = io.read()
```

Reading directly from an identified file

■ Once you have concluded your investigation and found some NWB files of interest, you can either…

  ➢ Or stream directly (most recommended for one-off analyses or quick calculations)

    [In MATLAB](#)*…

```
%% The S3 path must be copy/pasted manually

s3 = 's3://dandiarchive/blobs/7ee/415/7ee41580-9b0b-44ca-8675-6959ddd8dc33'

nwbfile = nwbRead(s3)
```

* streaming speeds are usually much slower than in Python