
NWB Storage

Release v1.0.0

Aug 03, 2017

Table of Contents

1	NWB-N Storage	1
1.1	What is the role of data storage?	1
1.2	How are NWB-N files stored?	1
1.3	Are backends other than HDF5 supported?	1
2	HDF5	3
2.1	Format Mapping	3
2.2	Key Mapping	4
2.2.1	Groups	4
2.2.2	Datasets	4
2.2.3	Attributes	4
2.2.4	Links	5
3	Release Notes	7
3.1	Version: 1.0.0 and later	7
3.2	Version: Before 1.0.0	7
4	Credits	9
4.1	Authors	9
4.1.1	NWB-N: Version 1.1.x and later	9
4.1.2	NWB-N: Version 1.0.x and earlier	10
4.2	Acknowledgments	10
4.2.1	2016 - 2017	10
4.2.2	2015 - 2016	10
4.3	Copyright	10
4.4	Licence	10
5	Indices and tables	13

What is the role of data storage?

The NWB-N format specification describes how to organize large collections of neuroscience data using basic primitives, e.g., Files, Groups, Datasets, Attributes, and Links to describe and hierarchically group data. The role of the data storage then is to store large collections of neuroscience data. In other words, the role of the storage is to map NWB-N primitives (and types, i.e., `neurodata_types`) to persistent storage.

How are NWB-N files stored?

The NWB-N format currently uses HDF5 as primary storage mechanism. The mapping of the NWB-N format to HDF5 files is described in more detail in [Section 2](#).

Are backends other than HDF5 supported?

NWB-N currently only officially supports HDF5 as main storage backend. However, the PyNWB API has been designed to enable the design of custom read/write backends for the API, enabling other storage backends to be mapped to NWB-N.

The NWB-N format currently uses the [Hierarchical Data Format \(HDF5\)](#) as primary mechanism for data storage. HDF5 was selected for the NWB format because it met several of the project’s requirements. First, it is a mature data format standard with libraries available in multiple programming languages. Second, the format’s hierarchical structure allows data to be grouped into logical self-documenting sections. Its structure is analogous to a file system in which its “groups” and “datasets” correspond to directories and files. Groups and datasets can have attributes that provide additional details, such as authorities’ identifiers. Third, its linking feature enables data stored in one location to be transparently accessed from multiple locations in the hierarchy. The linked data can be external to the file. Fourth, HDF5 is widely supported across programming languages (e.g., C, C++, Python, MATLAB, R among others) and tools, such as, [HDFView](#), a free, cross-platform application, can be used to open a file and browse data. Finally, ensuring the ongoing accessibility of HDF-stored data is the mission of The HDF Group, the nonprofit that is the steward of the technology.

Format Mapping

Here we describe the mapping of NWB primitives (e.g., Groups, Datasets, Attributes, Links etc.) used by the NWB format and specification to HDF5 storage primitives. As the NWB-N format was designed with HDF5 in mind, the high-level mapping between the format specification and HDF5 is quite simple:

Table 2.1: Mapping of groups

NWB Primitive	HDF5 Primitive
Group	Group
Dataset	Dataset
Attribute	Attribute
Link	Soft Link

Key Mapping

Here we describe the mapping of keys from the specification language to HDF5 storage objects:

Groups

Table 2.2: Mapping of groups

NWB Key	HDF5
name	Name of the Group in HDF5
doc	HDF5 attribute doc on the HDF5 group
groups	HDF5 groups within the HDF5 group
datasets	HDF5 datasets within the HDF5 group
attributes	HDF5 attributes on the HDF5 group
links	HDF5 SoftLinks within the HDF5 group
linkable	Not mapped; Stored in schema only
quantity	Not mapped; Number of appearances of the dataset.
neurodata_type	Attribute neurodata_type
namespace ID	Attribute neurodata_namespace

Datasets

Table 2.3: Mapping of datasets

NWB Key	HDF5
name	Name of the dataset in HDF5
doc	HDF5 attribute doc on the HDF5 dataset
dtype	Data type of the HDF5 dataset
shape	Shape of the HDF5 dataset if the shape is fixed, otherwise shape defines the maxshape
dims	Not mapped
attributes	HDF5 attributes on the HDF5 group
linkable	Not mapped; Stored in schema only
quantity	Not mapped; Number of appearances of the dataset.
neurodata_type	Attribute neurodata_type
namespace ID	Attribute neurodata_namespace

Note:

- TODO Update mapping of namespace ID
 - TODO Update mapping of dims
-

Attributes

Table 2.4: Mapping of attributes

NWB Key	HDF5
name	Name of the attribute in HDF5
doc	Not mapped; Stored in schema only
dtype	Data type of the HDF5 attribute
shape	Shape of the HDF5 dataset if the shape is fixed, otherwise shape defines the maxshape
dims	Not mapped; Reflected by the shape of the attribute data
required	Not mapped; Stored in schema only
parent	Not mapped; In HDF5 all attributes are explicitly tied to the parent.
value	Data value of the attribute

Links

Table 2.5: Mapping of links

NWB Key	HDF5
name	Name of the HDF5 Soft Link
doc	Not mapped; Stored in schema only
target_type	Not mapped. The target type is determined by the type of the target of the HDF5 link

CHAPTER 3

Release Notes

Version: 1.0.0 and later

Created separate reStructuredText documentation to discuss and govern storage-related concerns.

Version: Before 1.0.0

Before version 1.0.0 there was no official separate document governing NWB-N storage concerns as HDF5 was the only supported storage backend with implicit mapping between HDF5 types and NWB-N language primitives.

Contents

- *Credits*
 - *Authors*
 - * *NWB-N: Version 1.1.x and later*
 - * *NWB-N: Version 1.0.x and earlier*
 - *Acknowledgments*
 - * *2016 - 2017*
 - * *2015 - 2016*
 - *Copyright*
 - *Licence*

Authors

NWB-N: Version 1.1.x and later

Version 1.1.x of the NWB-N format, Version 1.2.x and later of the specification language, and Version 1.x and later of the NWB storage documents have been created by Oliver Ruebel and Andrew Tritt et al. in collaboration with original members of the NWB pilot project at the University of California in Berkeley and San Francisco and at the Allen Institute for Brain Science.

NWB-N: Version 1.0.x and earlier

Version 1.0.5g (and earlier) of the NWB file format and Version 1.1c (and earlier) of the specification language were created by Jeff Teeters et al. as part of the first NWB pilot project. The documents governing the specification language and format have been adopted from the final versions released by the original NWB pilot project.

Acknowledgments

2016 - 2017

Note: #TODO Add Acknowledgement for LBL Kavli project and acknowledge other projects as well

2015 - 2016

The Neurodata Without Borders: Neurophysiology Initiative is funded by GE, the Allen Institute for Brain Science, the Howard Hughes Medical Institute (HHMI), The Kavli Foundation and the International Neuroinformatics Coordinating Facility. Our founding scientific partners are the Allen Institute, the Svoboda Lab at the Janelia Research Campus of HHMI, the Meister Lab at the California Institute of Technology, the Buzsaki Lab at New York University School of Medicine, and the University of California, Berkeley. Ovation.io is our founding development partner. Ken Harris at University College London provided invaluable input and advice.

Copyright

“nwb-schema” Copyright (c) 2017, The Regents of the University of California, through Lawrence Berkeley National Laboratory (subject to receipt of any required approvals from the U.S. Dept. of Energy). All rights reserved.

If you have questions about your rights to use or distribute this software, please contact Berkeley Lab’s Innovation & Partnerships ce at IPO@lbl.gov.

NOTICE. This Software was developed under funding from the U.S. Department of Energy and the U.S. Government consequently retains certain rights. As such, the U.S. Government has been granted for itself and others acting on its behalf a paid-up, nonexclusive, irrevocable, worldwide license in the Software to reproduce, distribute copies to the public, prepare derivative works, and perform publicly and display publicly, and to permit other to do so.

Licence

“nwb-schema” Copyright (c) 2017, The Regents of the University of California, through Lawrence Berkeley National Laboratory (subject to receipt of any required approvals from the U.S. Dept. of Energy). All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University of California, Lawrence Berkeley National Laboratory, U.S. Dept. of Energy nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

You are under no obligation whatsoever to provide any bug fixes, patches, or upgrades to the features, functionality or performance of the source code ("Enhancements") to anyone; however, if you choose to make your Enhancements available either publicly, or directly to Lawrence Berkeley National Laboratory, without imposing a separate written license agreement for such Enhancements, then you hereby grant the following license: a non-exclusive, royalty-free perpetual license to install, use, modify, prepare derivative works, incorporate into other computer software, distribute, and sublicense such enhancements or derivative works thereof, in binary and source code form.

CHAPTER 5

Indices and tables

- `genindex`
- `modindex`
- `search`