
NeuroHDF Documentation

Release 1.0

NeuroHDF Developers

Sep 27, 2017

Contents

1	Multi-compartment neural circuitry	3
2	N-dimensional, homogeneous arrays	5
3	Multiscale image datasets	7
4	Physiology	9
5	Surfaces	11
6	Behavioral datasets	13
7	Simulation	15
8	Serial section 2D images	17
9	Evaluation of HDF5	19
10	Further reading	21

Neuroscientists need to manage and integrate data from anatomy, physiology, behavior and simulation data on multiple spatial and temporal scales and across modalities, individuals and species. Large amounts of data with complex data types are to be produced in the coming decades - and viable solutions for databasing, data sharing and interoperability of software tools [are needed](#).

“[Hierarchical Data Format \(HDF5\)](#) is a data model, library, and file format for storing and managing data. It supports an unlimited variety of datatypes, and is designed for flexible and efficient I/O and for high volume and complex data.”

NeuroHDF is an effort to combine the flexibility and efficiency of HDF5 for neuroscience datasets through the specification of a simple layout for different data types with minimal metadata. The *NeuroHDF Interest Group* consists of the members of [this group](#).

Multi-compartment neural circuitry

The SWC format became the quasi-standard for the description of single neuronal cell morphology reconstructions. For the description of larger neural circuits with many neurons and their synaptic connectivity, a new and efficient data format is needed. NeuroHDF describes a multi-compartmental neural circuit, similar to SWC, with points in 3D space (vertices) and their connectivity. Attributes like vertex (skeleton node, connector, root) or edge (presynaptic_to, postsynaptic_to) type or radius are expressed as arrays corresponding to the vertices or edges.

You can use Hdfview to inspect an [example NeuroHDF file](#). The software tool CATMAID for neural circuit reconstruction exports microcircuits in this format. Another emerging standard is [libNeuroML](#).

Useful for neuroscience data types

- Single cell morphology
- Neural circuit reconstructions

Tool supporting this specification

- [CATMAID](#) exports neural circuit reconstructions in NeuroHDF

N-dimensional, homogeneous arrays

By using a simple convention to describe metadata about the array axes, basic information is available to make sensible interpretation of the array's content.

Useful for neuroscience data types

- Electron microscopy: 3D array, 3 spatial dimension after alignment
- Optical microscopy: 4D array, 3 spatial dimension and 1 channel dimension
- Labeling
- Functional MRI/PET: 4D array, with 3 spatial and 1 temporal dimension
- Structural MRI: 3D array, with 3 spatial dimensions
- Diffusion MRI: 4D array, with 3 spatial dimension and 1 dimension for gradient directions Contain metadata tables for b-values and b-vectors

Tool supporting this specification

None so far. A [zebra fish dataset](#) available as HDF5 files uses a similar specification for axes units.

Example generation

Multiscale image datasets

As an extension to generic *N-dimensional, homogeneous arrays* representation, addition of subgroups for different scales can represent multiple (spatial) scales.

Useful for neuroscience data types

- Multi-scale electron microscopy

Tool supporting this specification

None so far.

Example generation

<Text and example dataset>

Useful for neuroscience data types

- Extracellular recordings
- Intracellular recordings
- Calcium imaging
- EEG
- MEG
- NIRS

Tool supporting this specification

None so far.

Example generation

A proliferation of file formats exist for 3D surfaces. The most widely used scheme to store surfaces is as triangular meshes, using vertices (points in 3D) and faces (usually triangles which describe the connectivity of the vertices). Additionally, values can be stored either on the vertices or faces. We propose the same convention to store surfaces in NeuroHDF. Additionally, level-of-detail meshes can be expressed with an additional group indirection.

Useful for neuroscience data types

- Cortical and subcortical surface-based atlases
- Morphology of neurons or subcellular components

Tool supporting this specification

None so far.

Example generation

CHAPTER 6

Behavioral datasets

- Behavioral experiments of tracked animals moving on a 2D plate
 - Irregular spatio-temporal data in a spatial reference system
- Questionnaire results

<Text and example dataset>

Useful for neuroscience data types

- Multicompartmental model simulation - Gaute Einevoll - Key challenges in multiscale modeling of neural tissue
- James Kozloski - The Neural Tissue Simulator

Tool supporting this specification

None so far.

Example generation

Serial section 2D images

<Text and example dataset>

Useful for neuroscience data types

- Serial section electron microscopy

Tool supporting this specification

None so far.

Example generation

Main HDF Group page <http://www.hdfgroup.org/>

Supported Libraries

Python libraries

<http://code.google.com/p/h5py/> <http://www.pytables.org/moin>

Java libraries

<http://www.ral.ucar.edu/~steves/nujan.html> <https://wiki-bsse.ethz.ch/display/JHDF5>

Matlab

<http://www.mathworks.com/help/techdoc/ref/hdf5.html>

R bindings

<https://r-forge.r-project.org/projects/h5r/>

.NET

<http://www.hdfgroup.org/projects/hdf.net/>

Advantages of using HDF5

- Compact binary data storage, extensible metadata
- Fast random and parallel access, efficient, scalable
- Widely used in High Performance Computing
- Open source and cross-platform
- [HDF5-Fast Query](#) and [paper](#)

Possible limitations of HDF5

- Difficulty to store variable-length string properties.
- Deleting a dataset does not free the space on disk. Requires rewriting the file.
- Many read/write operations on the same HDF5 file might be limited.
- [Delete or update a dataset in HDF5?](#)
- [Evaluating HDF5: What limitations/features does HDF5 provide for modelling data?](#)

Biological image formats

- [Unifying biological image formats with HDF5](#)
- [BioHDF for next generation sequencing](#), [The Case for HDF](#), [Introduction to BioHDF](#)
- “Our current estimates are that there are approximately 80 proprietary file formats for optical microscopy alone (and not including other common imaging techniques) that must be supported by any bioimage informatics tool that aims to provide a generalizable solution. In short, the lack of standardized access to data makes the generation of informatics tools quite difficult.” [Reference](#)
- [Bio-Formats Java Library](#)
- [A HDF5 I/O plugin for ImageJ](#)

Neuroimaging formats

- [Nibabel: BIAP3 - a JSON nifti header extension](#)
- [Extending Nifti discussion](#)
- [MINC2](#)

Visualization formats

- [Fiber Bundle HDF5 library](#)
- [Introduction to HDF5 and F5](#)

Microscopy formats and metadata

- Open Microscopy Environment (OME): [Metadata matters: access to image data in the real world](#)

Climate and Forecast formats

- NetCDF Climate and Forecast (CF) Metadata Convention

General

- [HDF5 forum](#)

Global object model

- Object Model for Neuroinformatics
- Common Data Model for Neuroscience Data and Data Model Exchange

Software tools using HDF5

Python

- [neo for electrophysiolog](#)
- [pandas](#)
- [PyMVPA](#)
- [larry](#)
- [ilastik](#)
- [stimfit](#)

Java

- [neuroConstruct](#)

Matlab

- [Matlab natively supports HDF5](#)