

# HDF5 in NWB

Andrew Tritt

Computational Research Division  
Lawrence Berkeley National Laboratory

HDF5 and Its role in Exascale, Cloud, and Object Stores – SC19

November 19, 2019

[nwb.org](http://nwb.org)

---

# Overview

---

**Motivation:** community science needs a stable target for data

## NWB – An Ecosystem for Neuroscience Data Standardization

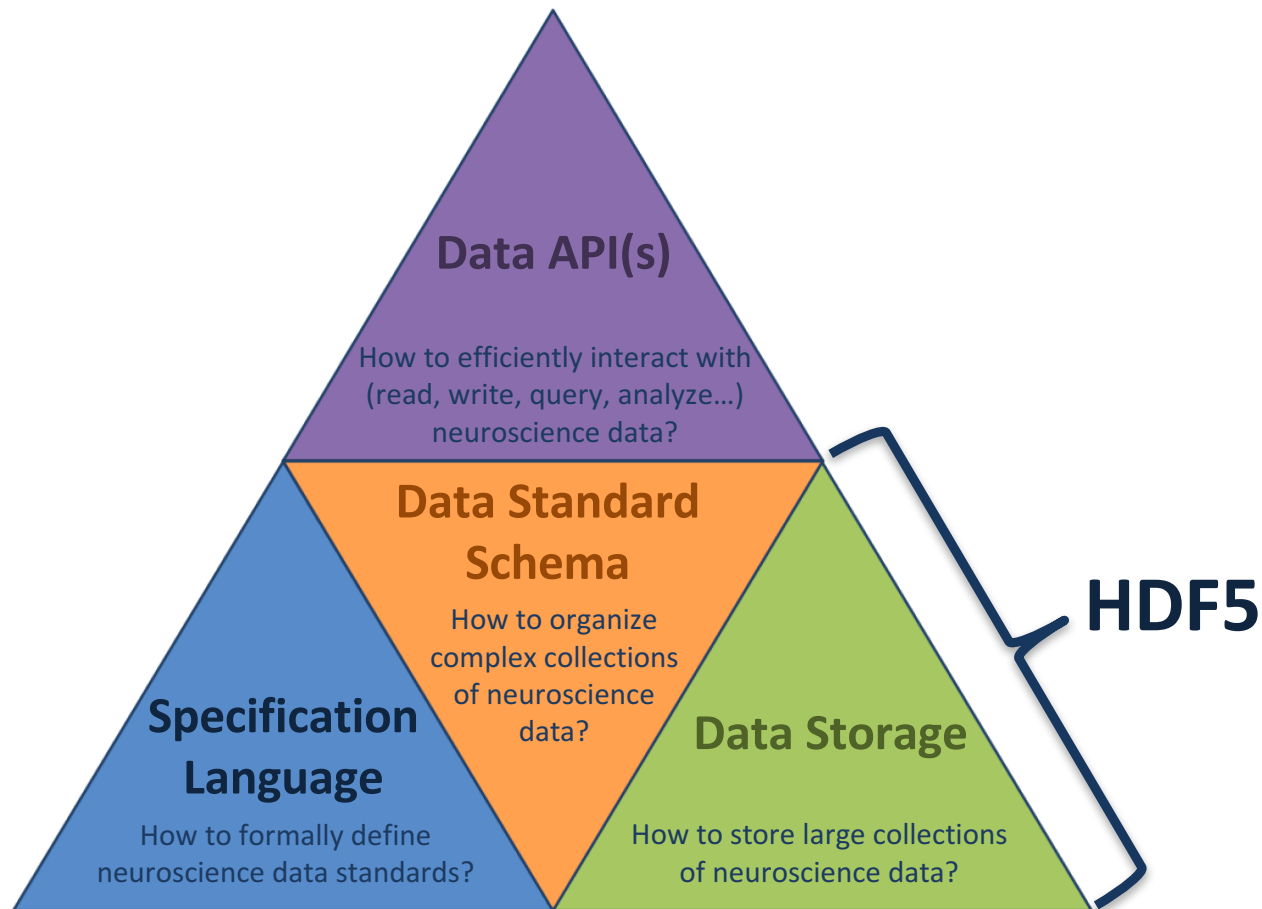
- a unified data format for neurophysiology data
  - focus on dynamics of neurons and groups of neurons
- an ecosystem of tools, methods, and standards for:
  - storage
  - sharing
  - analysis tools

**Goal:** develop a next generation data format and software ecosystem that will enable standardization, sharing, and reuse of neurophysiology data



**NEURODATA**  
WITHOUT BORDERS

# Main components of the NWB ecosystem



---

# Our opinions

---

## What works well

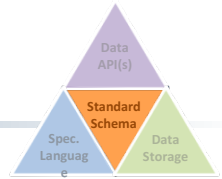
- Storing n-dimensional arrays
  - Including organizing
- Referencing objects
- Including metadata
- Chunking and compression

## What needs improvement

- Higher-order data structures
  - Ragged arrays
  - Tabular data
  - Sparse matrices
- API complexity
  - Modular storage
    - Cross-file referencing
  - Performance optimization

# High-level data structures (Part 1/2):

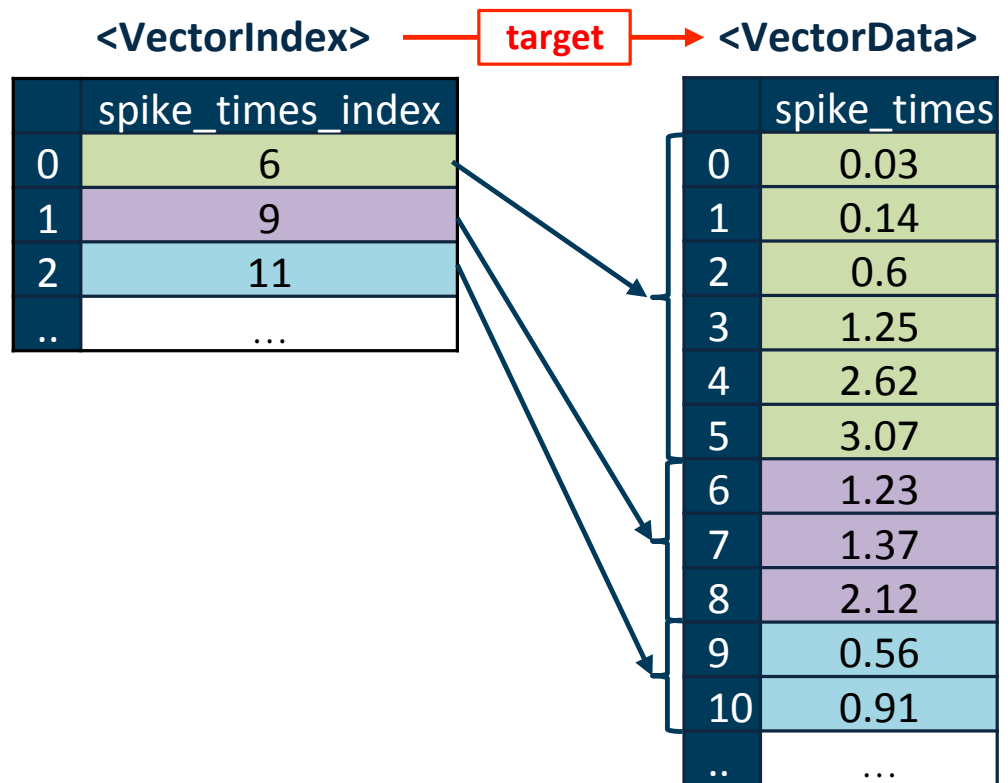
## Ragged arrays



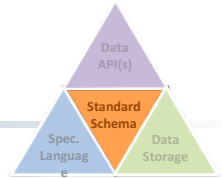
**Ragged array:** An array where each element of the array is itself an array of variable length.

0	0.03	0.14	0.6	1.25	2.62	3.07
1	1.23	1.37	2.12			
2	0.56	0.91				
..	..	..	..	..		

## Ragged arrays in NWB:N 2.0:



# High-level data structures (Part 2/2): Tables



Row-based  
table

electrodes (Dataset)						
	id	x	y	z	imp	location
0	0	0.1	0.3	0.4	4.5	'CA1', ...
1	1	0.5	0.4	0.2	10.4	'CA1', ...
2	2	1.5	2.6	1.8	1.1	'CA2', ...
..	...					

Column-based  
table

electrodes (Group)							
	id	x	y	z	imp	location	...
0	0	0.1	0.3	0.4	4.5	CA1	
1	1	0.5	0.4	0.2	10.4	CA1	..
2	2	1.5	2.6	1.8	1.1	CA2	..
..	...	...	...	...	...	...	...

Hybrid table

electrodes (Group)					
	id	(x,y,z)	imp	location	...
0	0	(0.1, 0.3, 0.4)	4.5	CA1	
1	1	(0.5, 0.4, 0.2)	10.4	CA1	..
2	2	(1.5, 2.6, 1.8)	1.1	CA2	..
..	...	...	...	...	...