NeuroData: Enabling Big Data Neuroscience

Eric Perlman
Center for Imaging Science, Johns Hopkins University

Challenge

- Neuroscientists are routinely collecting multi-terabyte datasets.
- Imaging throughput continues to increase at an amazing speed.
- Data collected are too large to visualize, analyze or store on local hardware.
- Data end up collecting dust on external hard drives and are not analyzed.

Goals

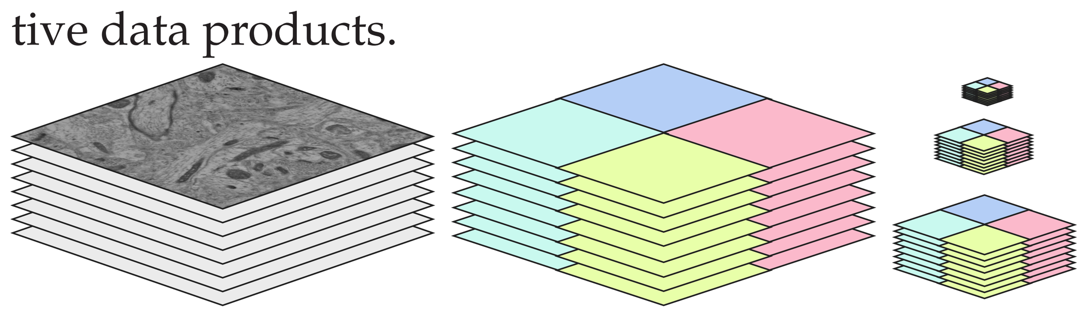
- Produce scalable tools with reproducible results.
- Integrate with existing workflows as seamlessly as possible.

Our Solution

- We build open source web-services for storing and interacting with data.
- Adopt open-source tools wherever possible.
- Use Docker containers to support identical execution in any environment (local, data center, or cloud).

Data Ingest and Storage

The Boss [5] is a multi-dimensional spatial database deployed as a managed service on Amazon Web Services (AWS), developed by JHU-APL for the IARPA MICrONS program and based on the original *ndstore* [4] architecture. A NeuroData specific deployment is used as the primary data store for both our archival and active data products.

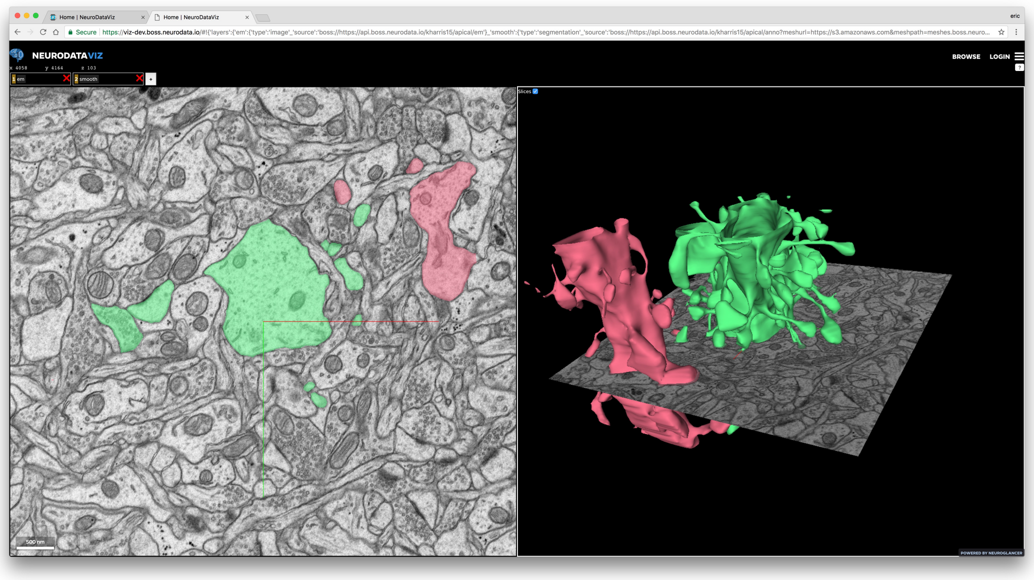


Slicing data for spatial data storage.

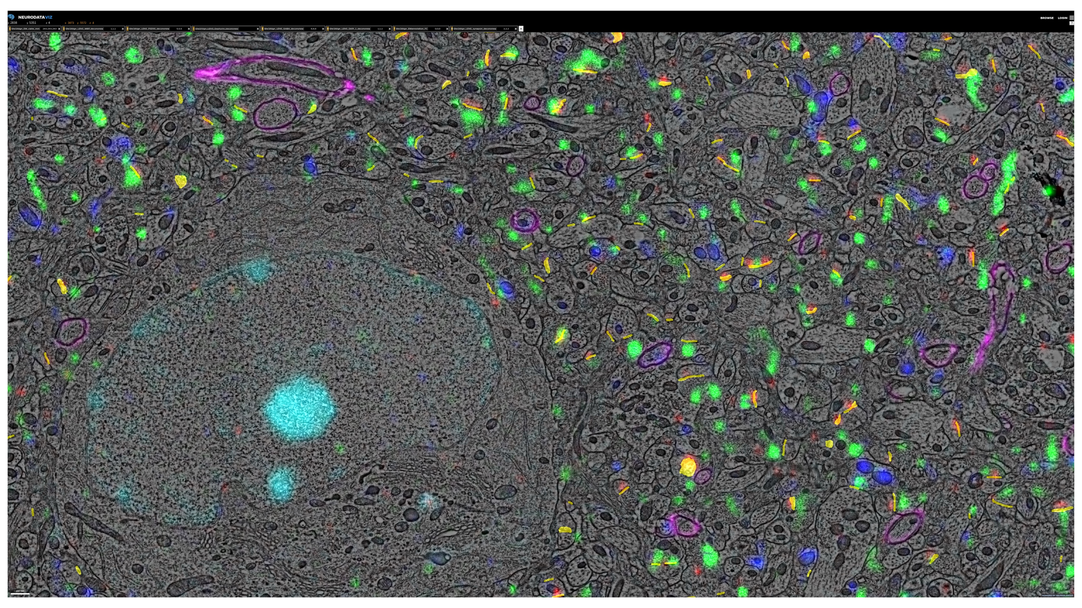
We built a tool to facilitate data ingest called *ingest_large_vol* and is currently in use by several of our data collaborators. This tool reads data in bulk and then uploads to the Boss in an optimized fashion. Further work is underway to use the cloud resources to further accelerate the ingest process.

Ben Falk (JHU), Will Gray (APL) and Dean Kleissas (former APL)

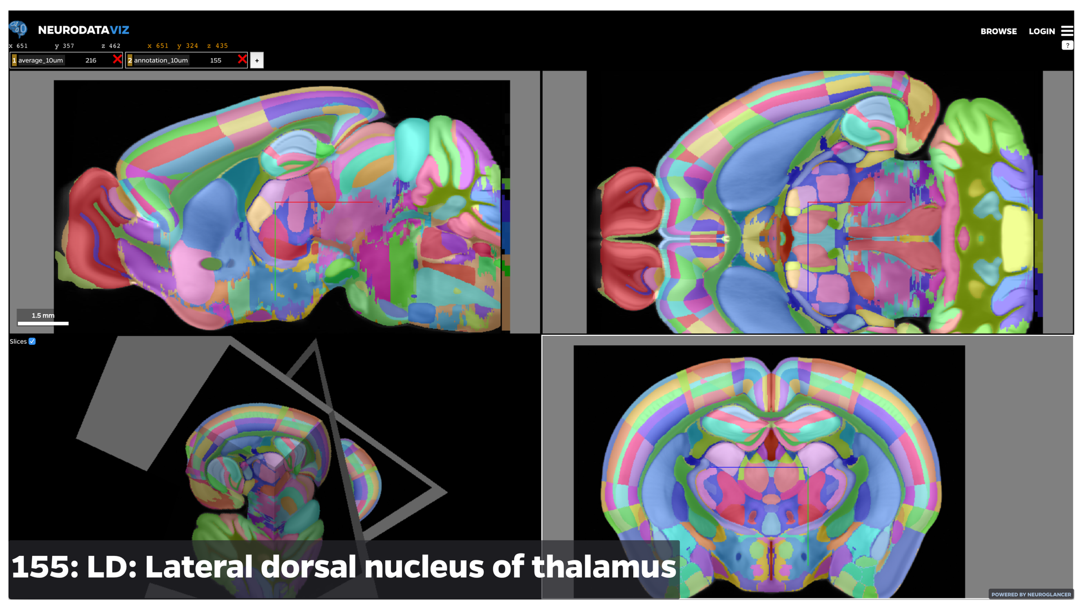
Visualization



Visualization of dendrites from within ndviz.
[Data: Harris et al. 2015]



Visualization of conjugate EM and array tomography.
[Collman, Unpublished]

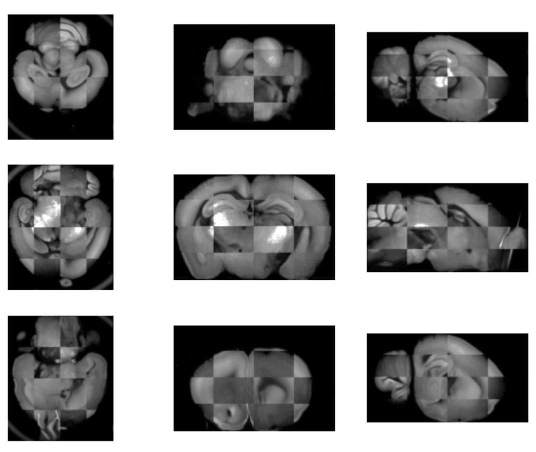


Visualization of the Allen Mouse Reference Atlas

The open source Web viewer Neuroglancer [3] handles the data retrieval and data rendering for 3D microscopy data and volumetric annotations. By building *ndviz* on top of Neuroglancer, we can immediately begin to develop interesting data visualizations specific to the problems we want to solve without having to re-engineer the fundamentals. Our extensions include both support for custom data sources and to provide new capabilities qualitatively exploring the data we work with. Examples include display of adjacent layers to evaluate alignment quality or querying an atlas for details on an annotation.

Alex Baden (JHU), Jeremy Maitin-Shepard (Google)

Registration



Using a *checkerboard* diagnostic to qualitatively evaluate a registration.

The ndreg library combines the LDDMM algorithm with the NeuroData infrastructure for registering pairs of images. Our primary use has been for mapping imaged brains to and from the appropriate reference atlases.

Vikram Chandrashekhar (JHU), Kwame Kutten (JHU), Daniel Tward (JHU)

Alignment and Interrogation of Raw Data

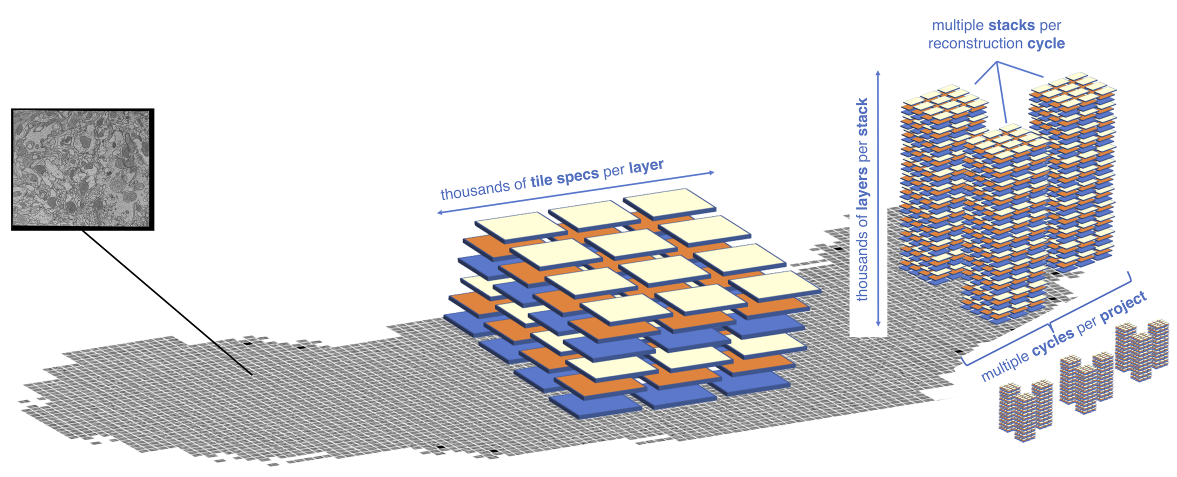
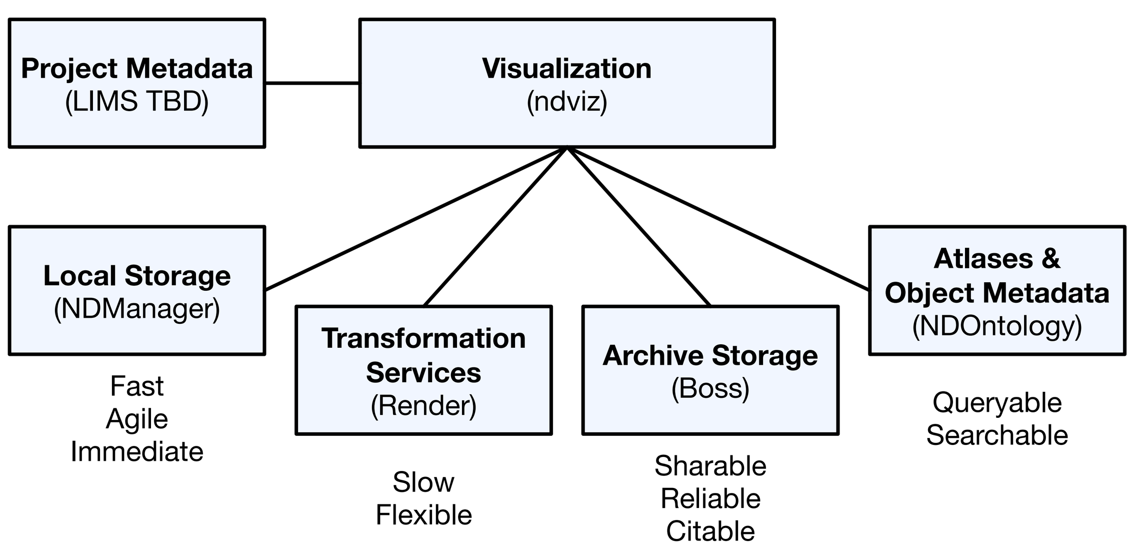


Image reconstruction stored in render.
Figure courtesy of Eric Trautman

Render [2] provides a framework for storing tile metadata and transformations and materializing transformed tiles using a rich suite of web services. These services enable us to dynamically interact with the data, enabling rapid visualization to assess both imaging and alignment quality throughout the entire data processing pipeline.
We have extended Render to support cloud deployment. With this approach, we can dynamically scale the capacity as needed to supported computationally expensive tasks like rendering out complete datasets.

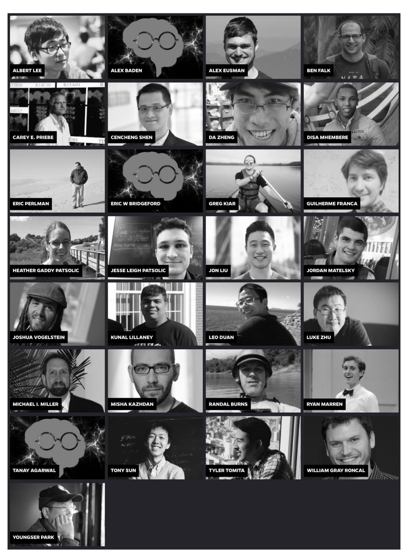
Forrest Collman (AIBS), Eric Perlman (JHU), Stephan Saalfeld (HHMI), Eric Trautman (HHMI)

Microservices



We are building our architecture as a collection of microservices.

NeuroData Team



The NeuroData team is large and evolving.
See `https://neurodata.io/people/` for a current list.

References

[1] NeuroData website. `https://neurodata.io/`.

[2] Render Tools and Services. `https://github.com/saalfeldlab/render/`.

[3] Neuroglancer. `https://github.com/google/neuroglancer/`.

[4] Burns et al. *The Open Connectome Project data cluster: Scalable analysis and vision for high-throughput neuroscience.* SSDBM 2013

[5] The Boss. `https://github.com/jhuapl-boss/`.

Acknowledgements

NeuroData is supported by awards from NIH, NSF, IARPA, DARPA, Johns Hopkins University, and the Kavli Foundation. Specific award information can be found at https://neurodata.io/about/.

The NeuroData team is large and evolving.
See `https://neurodata.io/people/` for a current list.