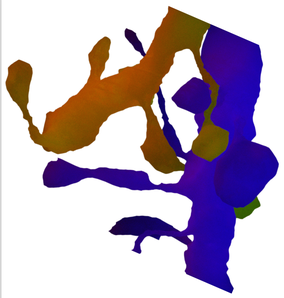
Cloud-native Infrastructure and Accessible Interfaces to Enable Petascale Neuroscience

J. Matelsky, L. Rodriguez, S. Bannowsky, S. Davis, A. Frazier, K. Porter, P. Manavalan, D. Pryor, R. Hider, Jr, T. Gion, W. R. Gray Roncal, D. Kleissas
Johns Hopkins University, Baltimore MD; Johns Hopkins University Applied Physics Laboratory, Laurel, MD
jordan.matelsky@jhuapl.com, william.gray.roncal@jhuapl.edu

Background



In order to better understand the neural circuit underpinnings of the brain, it is fundamental to determine its anatomical and spatial structure. Aiming to establish the connections between the neural pathways within the brain and its functionalities, efforts are focusing on generating "wiring-diagrams" of the brain. We can then compare the maps to functional data and draw inference about how the brain implements its computations.

To store massive, continuously-produced neural data, several teams have implemented powerful data archives, such as Johns Hopkins Applied Physics Laboratory's bossDB and HHMI Janelia's DVID.

The Problem

Current problems faced by the neuroscience community include:

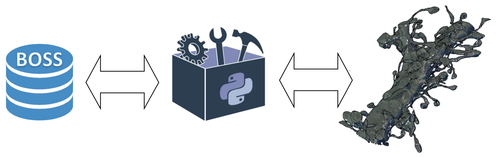
- The infrastructure necessary to easily share data does not exist
- Each data access tool uses a different format
- Tools require scientists to have coding experience

Our Solution

`intern` is a software development kit that can easily be downloaded through pip to access the bossDB data archive.

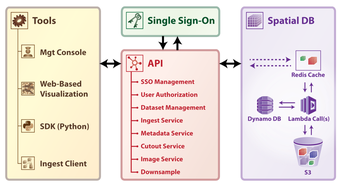
Using `intern`, we developed the following solutions:

- Data access to multiple industry-standard data-archives by extending `intern`
- A user-interface to make neuroscience easily accessible using React.js and Flask



bossDB

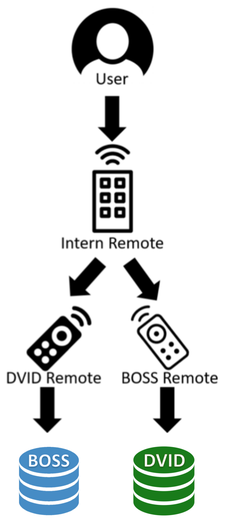
bossDB is a serverless, cloud-native spatial database for petascale imagery toward reproducible neuroscience. More details are available at **bossdb.org**.



intern Expansion

We expanded `intern` to communicate with Janelia's DVID database, focusing on upload and download consistency between DVID and bossDB.

```python
from intern.remote.boss import BossRemote
from intern.resource.boss.resource import ChannelResource
from intern.remote.dvid import DVIDRemote
# BOSS Data fetch:
boss = BossRemote({
    "protocol": "https",
    "host": "Boss api",
    "token": "User specific",})
volumeB = boss.get_cutout(
    boss.get_channel("channel", "collection", "experiment"),
    resolution, [10000, 10500], [10000, 10500], [500, 550],)
#DVID Data fetch:
dvid = DVIDRemote({
    "protocol": "http",
    "host": "DVID api",})
volumeD = dvid.get_cutout(
    dvid.get_UUID("UUID","grayscale"),
    [2300,4600],[2300,4600],[1380,1390])
```



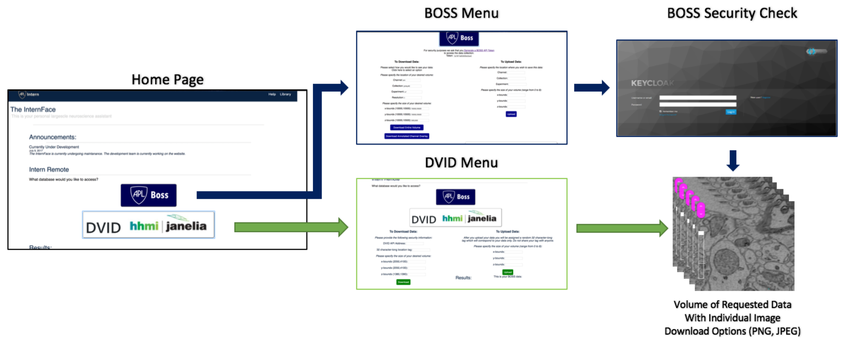
Future development will incorporate:

- recording logs
- deleting repositories
- changing server settings
- merging instances

Data Access Web Interface

We present a graphical user interface (UI) to facilitate user-friendly data-access using `intern` and related data-archive software.

This UI enables data download without code or installation steps.

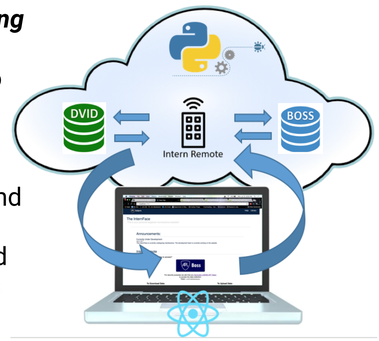


This illustrates the process of a user interacting with the new interface to obtain data without coding skills from either bossDB or DVID database.

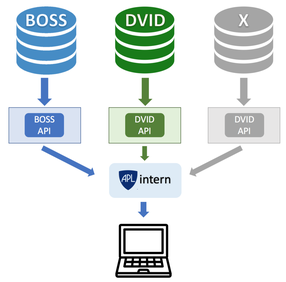
Results

- ***Enables easy data transfer from DVID to bossDB***
- Allows the user to take advantage of the database specific services such as **bossDB's authentication service** and **DVID's versioning services**

- ***Simplifies the process for retrieving and uploading data:***
  - usage of large data archives no longer requires coding experience
  - common interface for communicating with bossDB and DVID remotes
  - enables users to build tools and algorithms that work; abstracts the computer science details from the scientist



Future Projects



Our team plans to:

- Add other possible databases to `intern` remote library
  - Google Drive
  - Direct S3 Access
  - NeuroElectro
- Add support for local databases
- Improve database-specific capabilities in the web-based interface
- Facilitate transfer between archives

References and Links

References

- Kasthuri, N et. al, "Saturated Reconstruction of a Volume of Neocortex." *Cell*, 2015.
- Burns, R et. al "The Open Connectome Project Data Cluster: Scalable Analysis and Vision for High-Throughput Neuroscience." *SSDBM* 2013.

Links

IARPA Project Page
iarpa.gov/index.php/research-programs/microns
IARPA MICrONS Poster: goo.gl/BHBgQK
Boss Homepage: bossdb.org
DVID: github.com/janelia-flyem/dvid

Acknowledgements

- Caitlyn Bishop
- Joseph Downs
- Nathan Drenkow
- Marysol Encarnacion
- Pat Rivlin
- Brock Wester

We would also like to thank the JHU/APL CIRCUIT program.

This material is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2017-17032700004-005 under the MICrONS program. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation