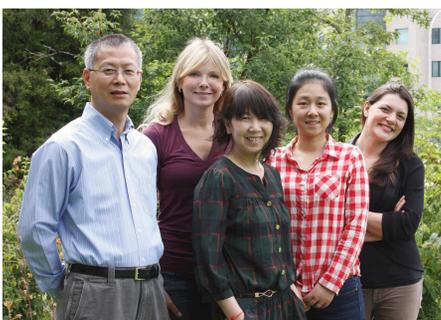


# Framing a Repository Solution for Bioimaging Datasets: Preserving *Utility*, not Bits

Kyle Banerjee  
banerjek@ohsu.edu

Robin Champieux  
champieu@ohsu.edu

## IT'S ALL ABOUT PEOPLE: DATA PRODUCERS AND DATA USERS



Mao Lab, Vollum Institute, OHSU

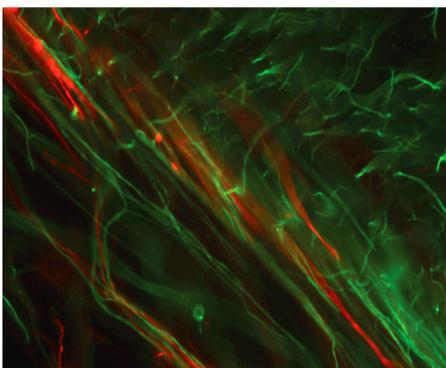
## THE DATA

≈ 14,000 files of high resolution mesoscopic mouse brain images

≈ Two-color phosphorescent labeling

≈ .5 TB data per mouse

≈ 40 TB total



## INCREASED NEED FOR DATA SERVICES

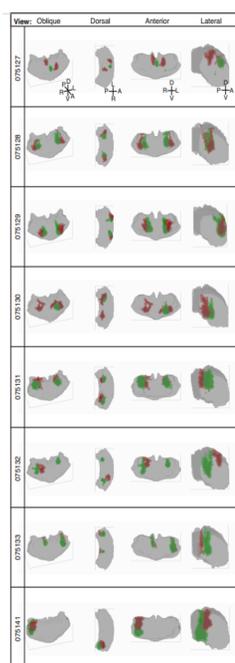
### Funder and journal policies

All OHSU funders and 39% of journals of interest mandate data sharing

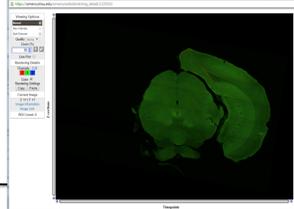
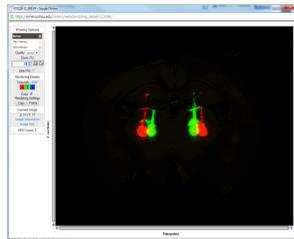
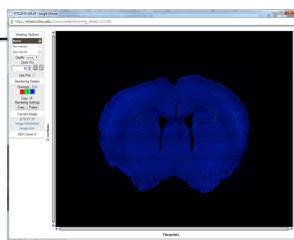
### Researchers say they need help!

- 2/3 produce image data
- 1/3 have datasets > 100 GB
- 12% have datasets > 1 TB
- Help describing and storing data
- Most solving their own data problems
- Status quo is expensive, ineffective, and not scalable

## BRAIN INDEX

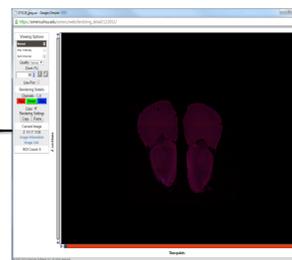


75 MOUSE BRAINS



Brain 075128 (125 IMAGES)

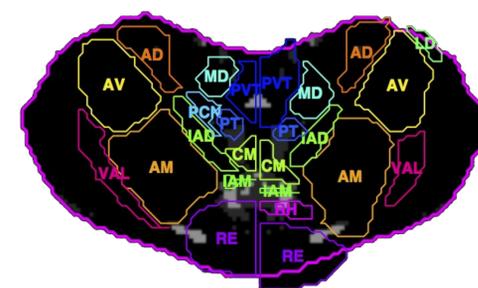
## AVI IMAGE SEQUENCE



PROCESSED DATA

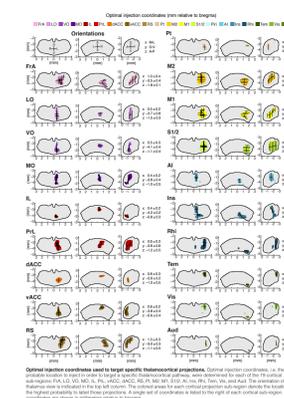
RAW DATA

## PROJECTION MAPS



19 cortical targets  
+ 50 jpg & 1 AVI per target

## OPTIMAL INJECTION SITES



ANALYZED DATA

ANALYSIS

## WHAT ARE WE TRYING TO FACILITATING?

### DISCOVERY, ACCESS & REUSE

- Increase the value of data
- Affect impact of the data, the researcher & the institution
- Accelerate discovery

## WHY IS THIS HARD?

### SIZE, QUANTITY & COMPLEXITY

- Ingest & download
- Meaningful, actionable, sustainable description
- Applicable & flexible groupings
- Useful visual interpretation
- Long term access

## BALANCING TENSIONS BETWEEN EASY DEPOSIT & USABILITY

### THE IMPORTANCE OF HOW YOU SHARE

- Organizing things so they are easily discoverable, searchable, and understandable is time consuming
- Simply depositing project data as is can render it nearly or completely unusable

## OUR SOLUTION

### TECHNOLOGY & PEOPLE

- Customized OMERO platform
- Server (48 GB RAM, direct attached storage, dual 8 core processors)
- Off-site taped back-up
- Ontology integration & annotation
- LOD with SPARQL endpoint
- Data curator

