Sprint 2:
20 December 2017

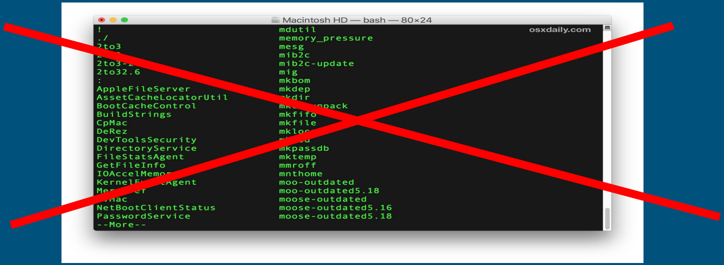
AVATR

AVATR High-Level Deliverables and Motivation

- DoD:
  - Refine annotation workflow
    - DoD: Demo at end of sprint
  - Develop MVP LIMS system
    - DoD: Demo at end of sprint
- Motivation:
  - Annotation workflow
    - Neuroscientists have a lot to gain from computational techniques, but don’t have the time for the learning curve to use them
    - We make that learning curve incredibly small, empowering more
  - LIMS system
    - When people do science without a LIMS system, things are messy
    - We help people understand what they are doing and keep their data organized

Annotation Pipeline: Improve Workflow

- DoD:
  - Demo
    - Have the annotation workflow require two terminal commands at most to run from end to end. The first terminal command should be to pull data from the BOSS. The second terminal command should be to push data to the BOSS.
- Exceeded DoD, only 1 command to start server, everything web-based



```
!                               mdutil
./                              memory_pressure
2to3                            mesg
2to3-                           mib2c
2to3-2.7                        mib2c-update
2to32.6                         mig
:                               mkbom
AppleFileServer                 mkdep
AssetCacheLocatorUtil           mkdir
BootCacheControl                mkextunpack
BuildStrings                    mkfifo
CpMac                           mkfile
DeRez                           mklocale
DevToolsSecurity                mknod
DirectoryService                mkpassdb
FileStatsAgent                  mktemp
GetFileInfo                     mmroff
IOAccelMemoryInfo               mnthome
KernelEventAgent                moo-outdated
MergePef                        moo-outdated5.18
MvMac                           moose-outdated
NetBootClientStatus             moose-outdated5.16
PasswordService                 moose-outdated5.18
--More--
```

Annotation Demo

- Not shown:
  - What an end user does only once at the very beginning
    - Cloning repo from github, `pip install -r requirements.txt`
    - Getting an account from the boss
    - Getting the api token and putting it in neurodata.cfg
- Shown:
  - what an end user would do every time they make annotations
    - Creating the upload channel in the Boss
    - Navigating to the data in ndwebtools
    - Initializing the annotation server
    - Pulling the image
    - Annotating
    - Pushing the annotations

MVP LIMS System: Dataset Registration

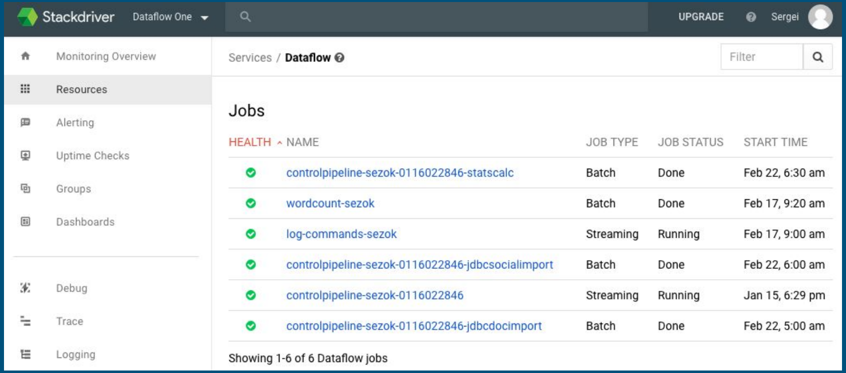
- **DoD:**
    - We will create csv and web parsers to scrape m2g.io dataset information into a database. The data should be organized in a manner such that CSV metadata information and appropriate links to graph are inserted in relation with each subject.
  - **Scrapers:**
    - CSV scraper to extract subject metadata from dataset covariates.
      Web parser for m2g.io to extract links for aligned images, tensors, fibers, graphs, QA plots, and version of code.
  - **Database Schema:**
    - Formatted database schema for structuring scraped data into database.
      Schema should be designed based off of feedback from Eric and Jovo.
  - **Database Manager:**
    - Scripts to load data correctly into database based off of specified schema.
    - Reach: Scripts to handle database migrations (schema change), missing data, and delete entries.

MVP LIMS System: Data Querying

- DoD:
  - MVP LIMS System Demo
    - The LIMS will have a route to display m2g.io datasets in an organized tree structure. Users will be able to explore individual subject information by clicking on a specific dataset and a specific subject.
    - Once a specific subject is selected, metadata information and relevant links related to the subject should be displayed in an organized interface.

MVP LIMS System DoDs: Logging

- DoD:
  - Tech Evals
    - Have at least two tech evaluations on current pipeline logging tools
      - StackDriver
      - DataJoint
    - Reach: Demo running pipeline on Eric's data and results being logged correctly.



Jobs

HEALTH ^	NAME	JOB TYPE	JOB STATUS	START TIME
✓	controlpipeline-sezok-0116022846-statscalc	Batch	Done	Feb 22, 6:30 am
✓	wordcount-sezok	Batch	Done	Feb 17, 9:20 am
✓	log-commands-sezok	Streaming	Running	Feb 17, 9:00 am
✓	controlpipeline-sezok-0116022846-jdbcsocialimport	Batch	Done	Feb 22, 6:00 am
✓	controlpipeline-sezok-0116022846	Streaming	Running	Jan 15, 6:29 pm
✓	controlpipeline-sezok-0116022846-jdbcdocimport	Batch	Done	Feb 22, 5:00 am

Showing 1-6 of 6 Dataflow jobs

MVP LIMS System DODs: Documentation

- DoD:
  - Thorough documentation outlining how each feature in the LIMS works will be up in Sphinx.
- Documentation is instead in Jupyter Notebooks
  - Reasons:
    - Easier to write and provide figures
    - Wanted more time to write better documentation since this is a MVP (will change a lot)

LIMS MVP Demo

- Not shown:
  - What an end user does only once at the beginning
    - Initializing mongodb
    - Cloning the repository (link)
- Shown:
  - What an end user does every time they want to use the LIMS system
    - Building the database
    - Initializing the LIMS server
    - Navigating the LIMS system

Next Steps

M2g.io LIMS:

- Add continuous integration (automated testing)
- Add visualizations