





Weekly Deliverables: 27 November 2017

AVATR

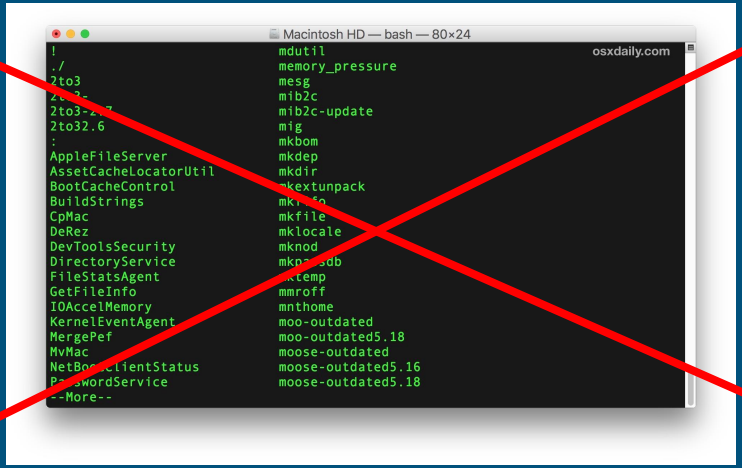
AVATR will focus on two main goals

- Refine annotation workflow
  - DoD: Demo at end of sprint
- Develop MVP LIMS system
  - DoD: Demo at end of sprint
- We will be presenting a demo for each of our goals demonstrating progress

Sprint 2	12/20	• Annotation Pipeline
		  ▪ DoD: Demo with <=2 terminal commands
		• Develop MVP for LIMS with the following features:
		  ○ Dataset registration
		  ○ Dataset queries
		  ○ Logging
		  ○ Documentation
		  ▪ DoD: Demo demonstrating features

Annotation Pipeline Iteration DoD:

- Full Demo which uses at most 2 terminal commands
  - Potentially locally hosted web-based pipeline
  - Make sure other people use it to check ease of use



MVP LIMS System

Composed of multiple DoDs:

1. Data Registration
2. Dataset Querying
3. Logging
4. Documentation (DoD: online with Sphinx)

\+ Demo

MVP LIMS System DoDs: Data Registration

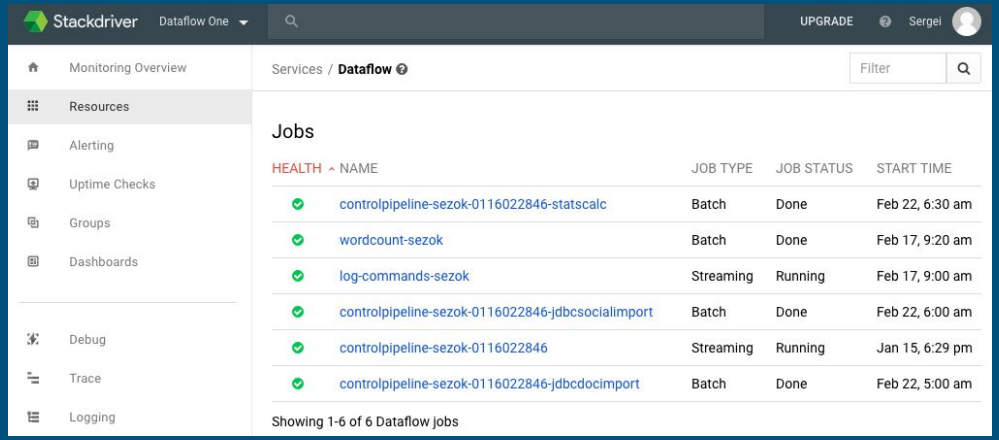
- Scrape and parse m2g.io data sets into database
- Schema of database showing data organized in a reasonable way
- Scripts to handle database migration and missing data

MVP LIMS System DoDs: Dataset Querying

- Front end that displays database in an organized structure
- During demo, metadata and links to data derivatives easily available
- Organize database into different categories
  - Subject based
  - Experiment based

MVP LIMS System DoDs: Logging

- During demo, logging is demonstrated to function correctly on Eric’s data
- Tech evals for existing logging tools
  - Stackdriver?



Next Steps

- Focus on data registration
  - Investigation of BIDS spec
- Plan out database organization
  - Preliminary schema
  - What matters in the data
  - Talk to Eric + people

Schema Proposal

The current datasets are already formatted in the current format:

- Dataset
  - metadata
    - subject
  - aligned images
    - subject
  - tensors
    - subject
  - fibers
    - subject
  - graphs
    - subject
  - quality assurance

Since many of the subtrees rely on subjects and our focus on the LIMS currently is on subjects, it makes sense for the first schema iteration to focus mostly on the subjects. This means that we should rearrange the dataset such that subjects take prevalence over the calculated plots and derivatives.

Below is our proposed solution:

- Dataset
  - quality assurance
  - subject
    - aligned images
    - tensors
    - fibers
    - graphs
    - metadata

With this schema, it should be fairly simple to perform search filters and queries of individual subject information while still keep track of the overall dataset quality.

Note that this tree structure fits the specs of a MongoDB extremely well.