Getting Started in the Flaherty Lab

Patrick Flaherty

Department of Biomedical Engineering
Worcester Polytechnic Institute
Worcester, MA 01609

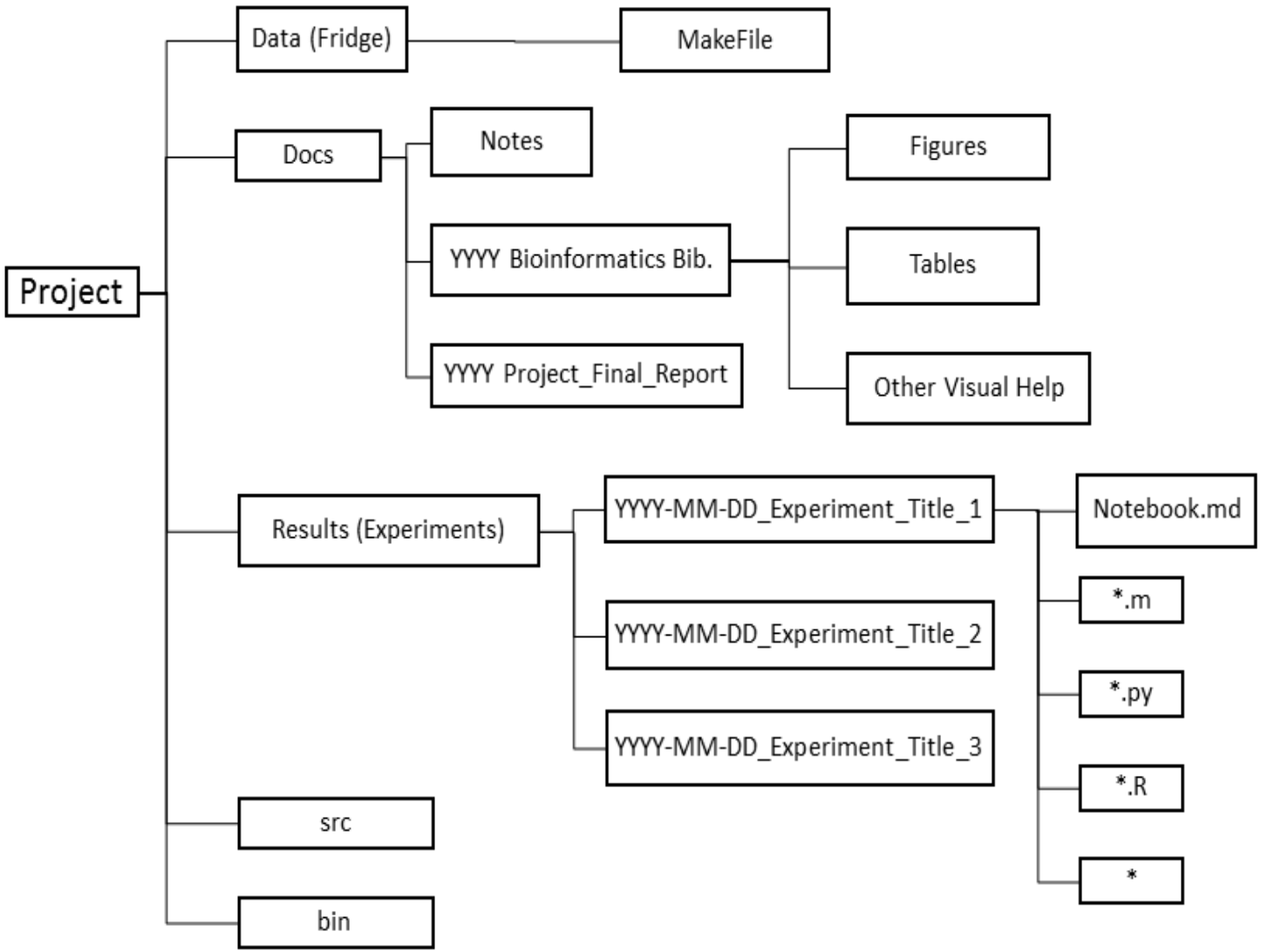
pjflaherty@wpi.edu

June 2, 2014

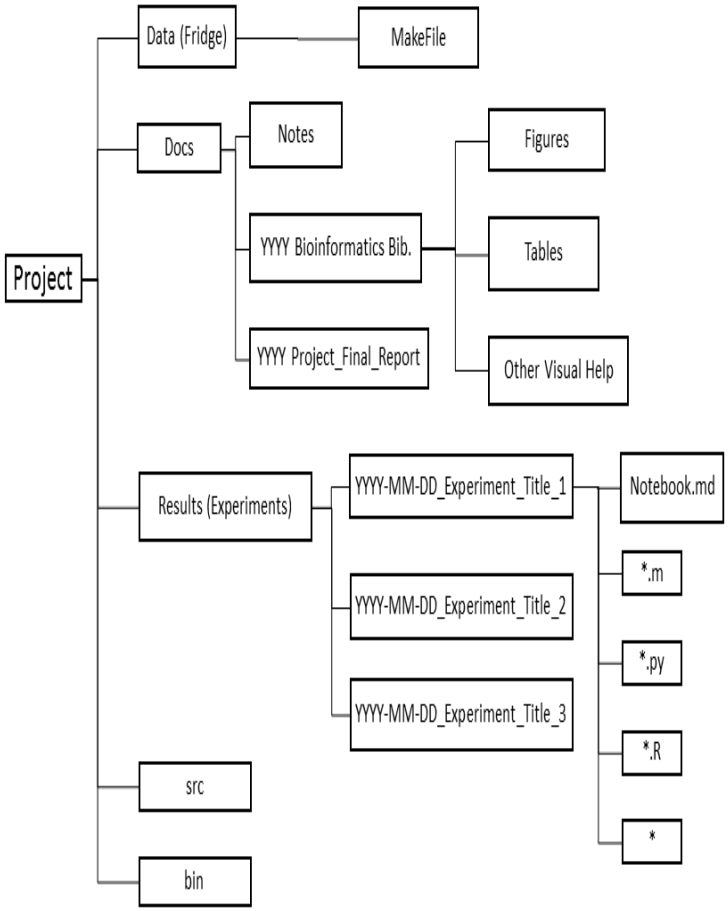
Data Management

- Bench Data
- Fridge Data
- Freezer Data

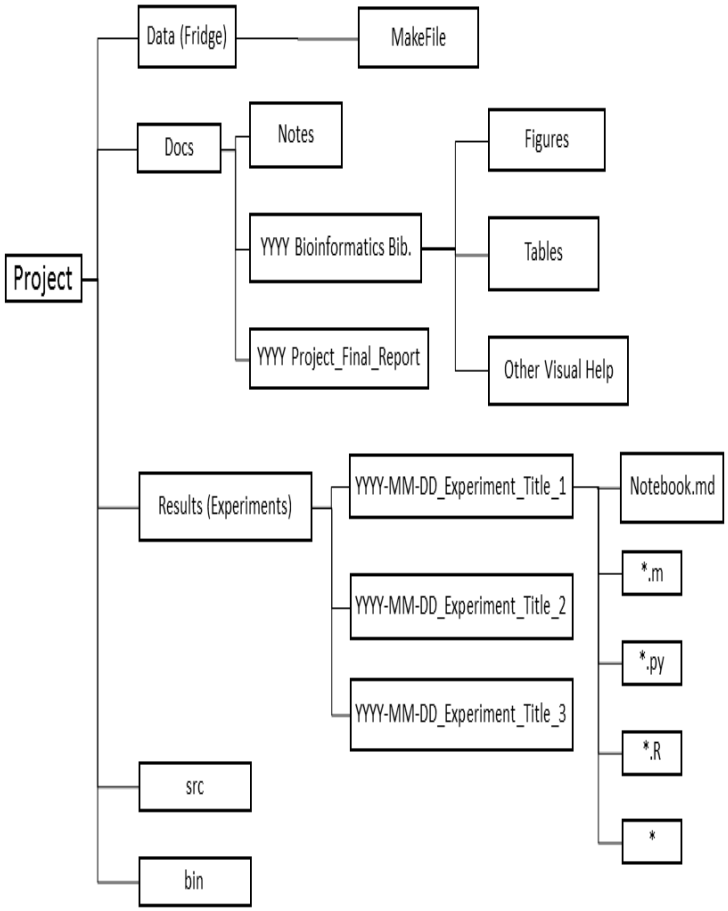
Data Management



Data Fridge


Project
Data (Fridge)
MakeFile
Docs
Notes
YYYY Bioinformatics Bib.
YYYY Project_Final_Report
Figures
Tables
Other Visual Help
Results (Experiments)
YYYY-MM-DD_Experiment_Title_1
YYYY-MM-DD_Experiment_Title_2
YYYY-MM-DD_Experiment_Title_3
Notebook.md
*.m
*.py
*.R
*
src
bin

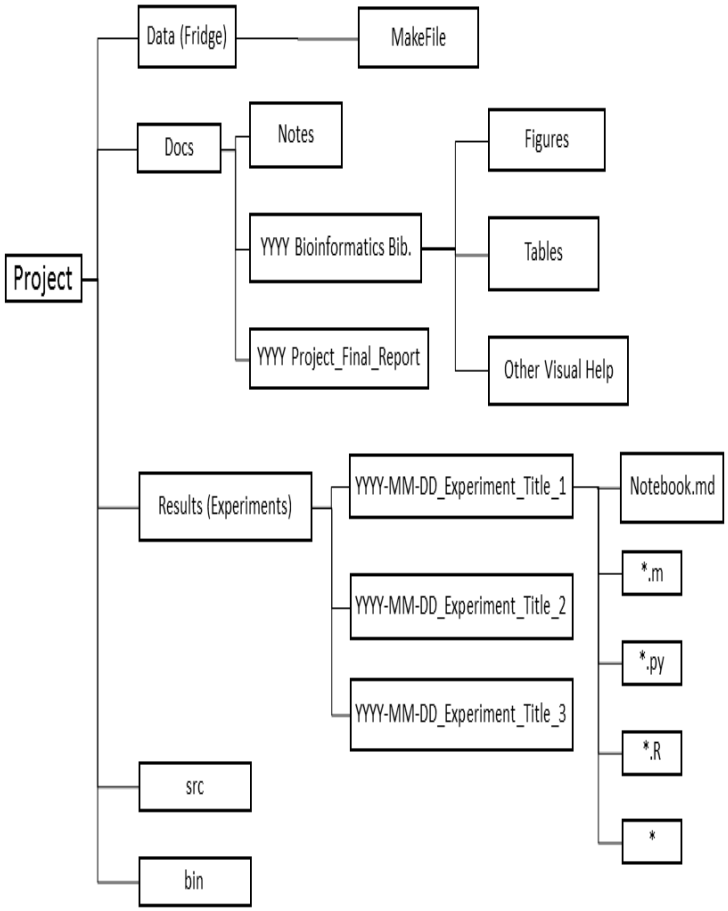
Data Management - Data Fridge



Data Fridge

- Makefile

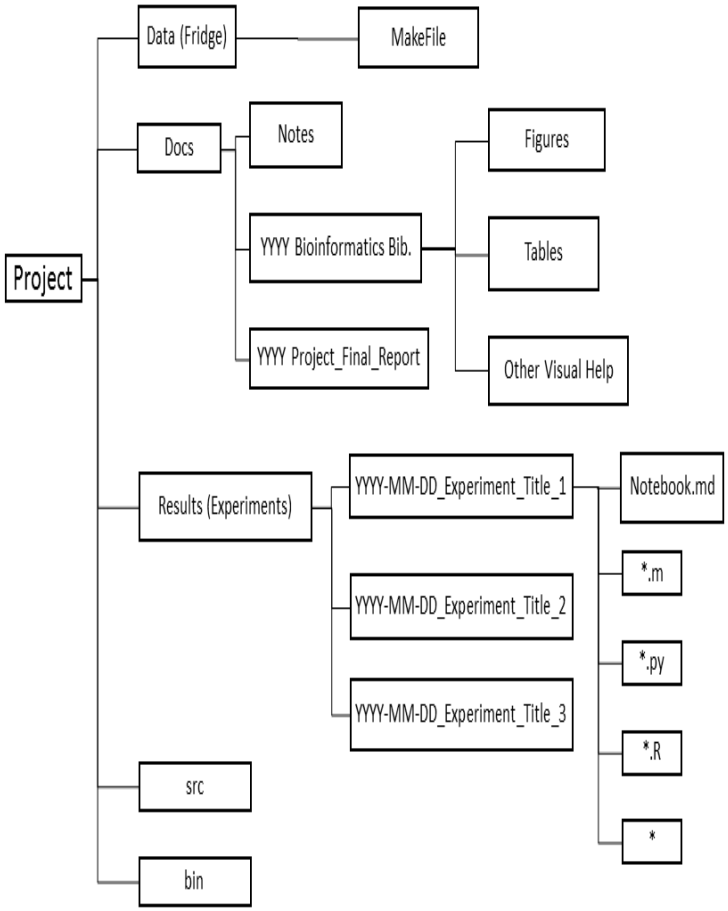
Data Management - Data Fridge



Data Fridge

- Makefile
- target:dependent
  [tab] rules

Data Management - Data Fridge



Data Fridge

- Makefile
- target:dependent
  [tab] rules
-

```
yourfile.dc : yourfile.pileup
    pileup2dc yourfile.pileup > yourfile.dc
yourfile.pileup : yourfile.bam
    samtools ipileup yourfile.bam -o yourfile.pileup
yourfile.bam : yourfile.sam
    samtools view -b -S -o yourfile.bam
```

Data Management - Docs

- Notebook
- Figures and Tables
- Project Final Report

Data Management - src (source)

Completed code for your data analysis program.

DataManagement - bin

Executable files, such as:

- Resources helpful to the experiment
- Compiled programs - execute with one command

Version Control

Version Control

- Why Version Control?

Version Control

- Why Version Control?
- What are the options for Version Control?

Version Control

- Why Version Control?
- What are the options for Version Control?
- Why GIT?

Version Control - GIT

- GIT GUI or Command Line
- clone, fetch, pull
- stage, commit, push
- Conflict Management

Flaherty Lab Environment

Flaherty Lab Environment

Flaherty Lab Environment

- Redwood Server

Flaherty Lab Environment

- Redwood Server
- Amazon Machine Image

Flaherty Lab Environment

- Redwood Server
- Amazon Machine Image
- Starcluster? MPI? What else?

Flaherty Lab Environment - Redwood Server

- Flaherty Lab Linux Server for Computation
- 64 core + 256GB RAM
- 9TB high speed drive + 1TB solid state drive via NFS

Flaherty Lab Environment - Amazon Machine Image (AMI)

- When you need more than 64 cores for calculation..

Flaherty Lab Environment - Amazon Machine Image (AMI)

- When you need more than 64 cores for calculation..
- numpy, scipy, h5py, pytables, matplotlib, pyramid, scikit-learn, pandas, statsmodels, networkx, theano, gdal, pysal, and shapely
- on-demand price: 0.145 dollar per worker per hour

  spot instance price: 0.018 dollar per worker per hour

Thank you!