

Project organization and management.

<https://datacarpentry.org/organization-genomics/02-project-planning/index.html>

https://github.com/shannonekj/Data_Organization_and_Management

Meet & Analyze Data

Shannon E.K. Joslin

06 November 2019

Overview

- Sending & receiving samples
- Storing data
- Setting up your project
- Working with data

Sending samples to the facility

- UC Davis DNA Technologies Core sample submission:
<https://dnatech.genomecenter.ucdavis.edu/sample-submission-scheduling/>

Retrieving sample sequencing data from the facility

- UC Davis lab information management system (slims):
<https://slims.bioinformatics.ucdavis.edu/login.php>
- ALWAYS ask for an md5sum

```
md5sum <file>
```

```
md5sum -c <file_with_list_of_md5sums>
```

Data!

Working with data

- FARM
 - Managed by Computer Science and Engineering Systems & Cluster Support
 - Documentation: <https://wiki.cse.ucdavis.edu/support/systems/farm>
 - RILAB Farm Documentation: <https://github.com/RILAB/lab-docs/wiki/Using-Farm>
- Cabernet
 - Managed by UC Davis Bioinformatics Core
 - Training: <https://ucdavis-bioinformatics-training.github.io/2018-June-RNA-Seq-Workshop/monday/cluster.html>

Back up your data!

- Come next week to learn where!

Data danger

- Always change raw data permission

```
chmod a=r <file>
```

Data Managment

- Please clone the following:

```
git clone https://github.com/shannonekj/Data_Organization_and_Management.git
```

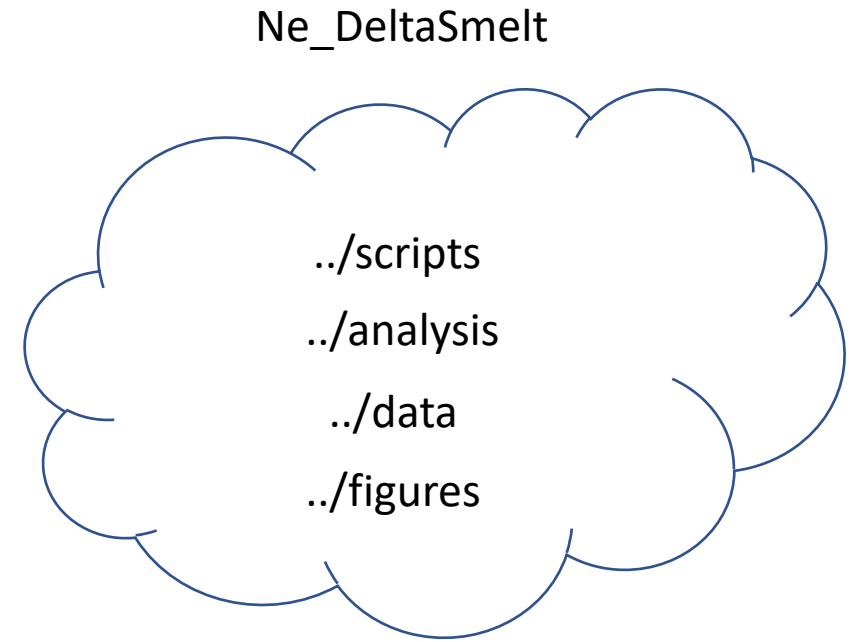
Structuring Projects

- Keep track of what you are doing
 - README files
- Consistency is key!
 - Naming files
 - Directory structure
 - Run scripts

Tricks for making your project repeatable

/Pliocene/shannonekj/Desktop/work/github/Ne_DeltaSmelt/scripts
/Pliocene/shannonekj/Desktop/work/github/Ne_DeltaSmelt/analysis
/Pliocene/shannonekj/Desktop/work/github/Ne_DeltaSmelt/data
/Pliocene/shannonekj/Desktop/work/github/Ne_DeltaSmelt/figures

- Use relative paths
 - Directories starting from the root will change
- Use conda environments
 - Software versions



Don't waste your time

- Use conda to install!
 - Can create a file to initialize conda environments or modules so you don't have to remember!
- Try scripts on a subset of data

Additional Resources

- Books:

- Bioinformatics Data Skills by Vince Buffalo <http://vincebuffalo.org/book/>

- Papers:

- Best Practices for Scientific Computing
<https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745>
 - Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3945096/>