Collaborative data management

Vessela Ensberg, Associate Director, Data Management

Alesia McManus, Environmental Sciences Librarian

Why manage data?

Learning outcomes
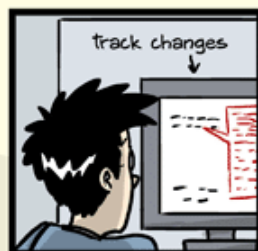
- You will practice managing a collaborative project with proper documentation
- You will know how to start and manage projects in Open Science Framework and GitHub
- You will learn about collaboration tools for your research.

Challenges in collaboration

Collaboration tools @UCDavis

- Storage
  - Box
  - Google Drive
  - AWS
  - Office 365

- Backup
  - CrashPlan

Messy files

How do you know where things are supposed to go?

Structural readMe file

- Organization of file/folder structure
- Relationship between files
- Naming conventions
- Definitions of acronyms, abbreviations
- Contact

UC DAVIS LIBRARY

Analysis readMe file

- Original dataset
- Description of the parameters/variables, units, codes
- Uncertainty, precision, and accuracy of measurements, if known
- Script running instructions
- Method(s), standards or calibrations that were used
- Specialized software
- Date dataset was last modified
- Example records for each data file (or file type)
- Contact information

http://d7.library.gatech.edu/research-data/readme

Activity: writing a structural readMe file

Develop a folder structure for the files related to either Zea mais sensor data or yield data and describe it in a readme file

https://github.com/Vensberg/collaborative_workshop_materials

UC DAVIS LIBRARY

Structural readMe file

- Organization of file/folder structure
- Relationship between files
- Naming conventions
- Definitions of acronyms, abbreviations
- Contact

Structural readMe file

Project level folder: ALIEnS

Subfolders level 1:

Zea_mais

Oidiodendron_maius

Subfolders level 2:

Zea_mais contains folders

Sensor_Raw_Data, Sensor_Exported, Sensor_Analysis, Yield_Raw and Yield_Analysis

Sensor_Analysis

Sensor_Exported

Sensor_Raw_Data

Yield_Analysis

Yield_Raw

Structural readMe file

Sensor_Raw_Data contains raw sensor data for light production (mdb file) and CO_2 measurements (bin file) of the Zea mais environment. each file is named YYYYMMDD_zea_mais_[variable]_raw, in which YYYYMMDD is the date the data were collected, and the variable is either light or CO_2.

Sensor_Exported contains files from the light and CO_2 sensors that have been migrated to csv format. Some data may be lost. The files are named YYYYMMDD_zea_mais_[variable]_migrated_raw, where [variable] is either light or CO_2, and YYYYMMDD is the date the data were collected.

YYYYMMDD_zea_mais_[variable]_raw and YYYYMMDD_zea_mais_[variable]_migrated_raw should be identical excepting data loss due to format transfer.

Sensor_Analysis contains statistical analyses of light or CO_2 fluctuations. Analyses were conducted as described in zea_mais_[variable]_analysis_readme.txt. Analyzed data are in YYYYMMDD_zea_mais_light_processed_YYYYMMDD.sas7bdat files that can be opened in SAS software. The first YYYYMMDD is the date the data were collected. The YYYYMMDD at the end of the file name is the date the data were analyzed.

UC **DAVIS** LIBRARY

Using tools to answer questions

- Do we have all the data?
  - What data were collected?
- Where are the data?
  - Collaborators at University X use Box, but we use Google Drive. How do we keep track of everything?
- Have you done the preliminary analysis?
- Who did the preliminary analysis?
- Who is supposed to collect data Z?
- When is collection of Z supposed to happen?

UC DAVIS LIBRARY

On your own

OSF
1. Sign up
2. Write two sentences in the Wiki
3. Upload a file as a part of a component and describe the file
4. Link to another storage space (i.e. GDrive)
5. Add an unregistered contributor (you can use your own e-mail)

GitHub
1. Sign up
2. Initiate a repository with a readme
3. Upload a file
4. Download GitHub for Desktop https://desktop.github.com/
5. Clone your repository locally