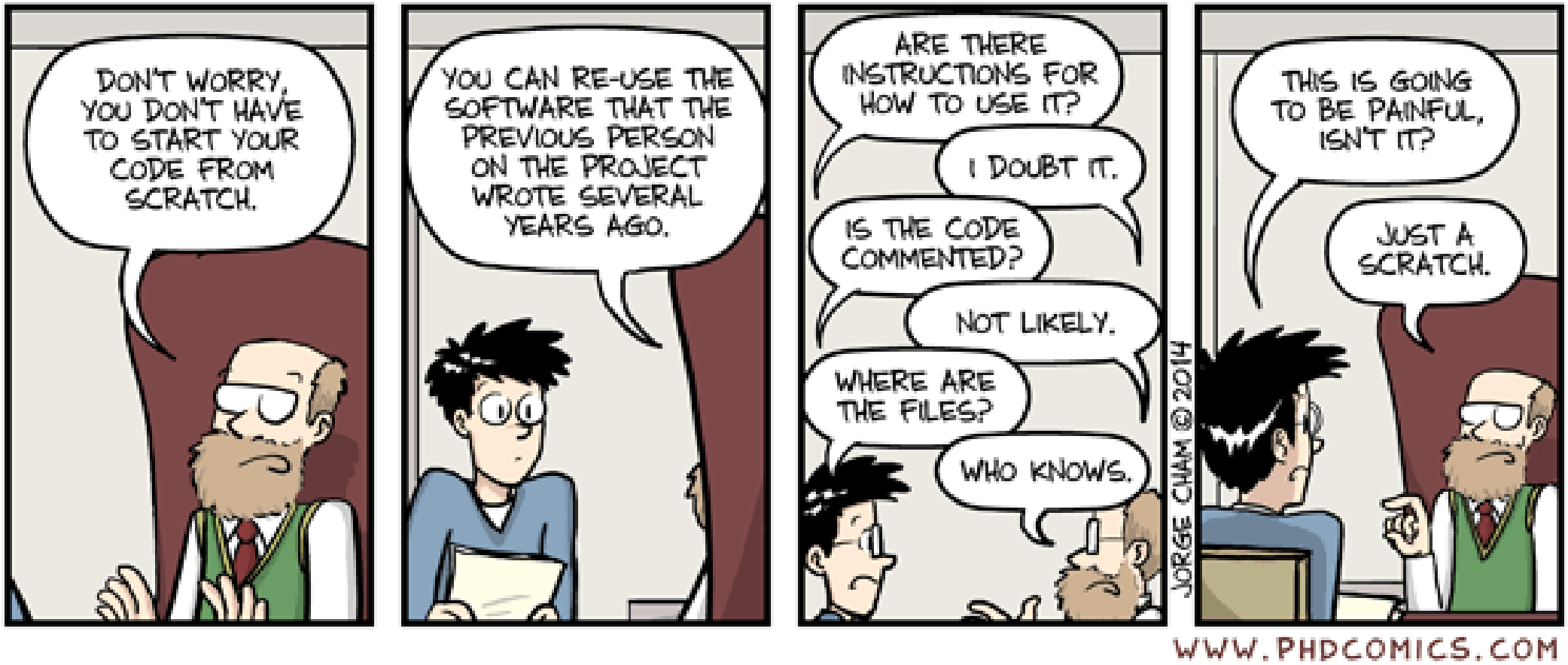
UC DAVIS LIBRARY

Keeping your data tidy and organized

Vessela Ensberg, Associate Director, Data Management
Alesia McManus, Environmental Sciences Librarian

Why manage data



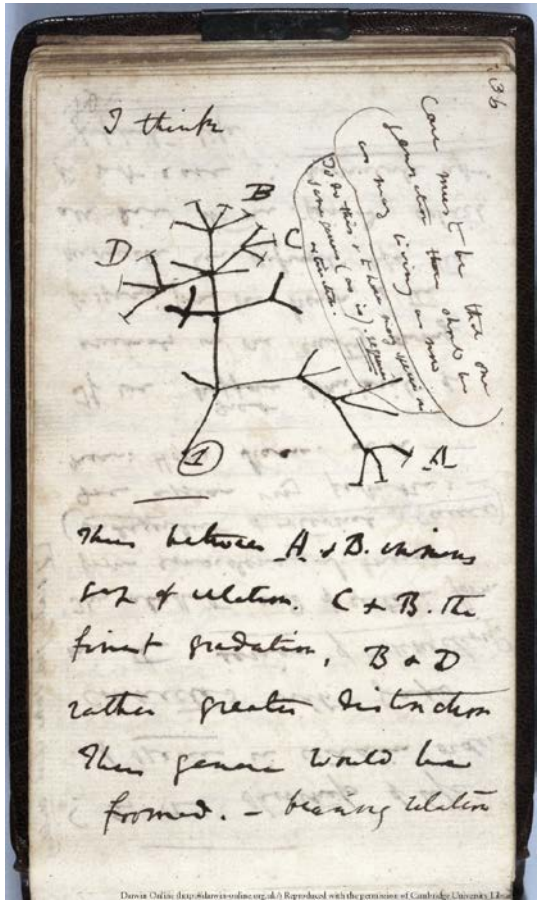
UC DAVIS LIBRARY

Learning outcomes

- Keeping track of multiple files generated during a project
- Setting up data validation in spreadsheets to avoid inconsistent entries

UC DAVIS LIBRARY

Historical look at notebooks



[page 14a]

As the Sandstone in Island contains Salt: some veins nearly 2 inches thick —

Lignite. black [glossy] coal —

parts vegetable structure. —

Difficulty of understanding Salt deposit in such open districts —

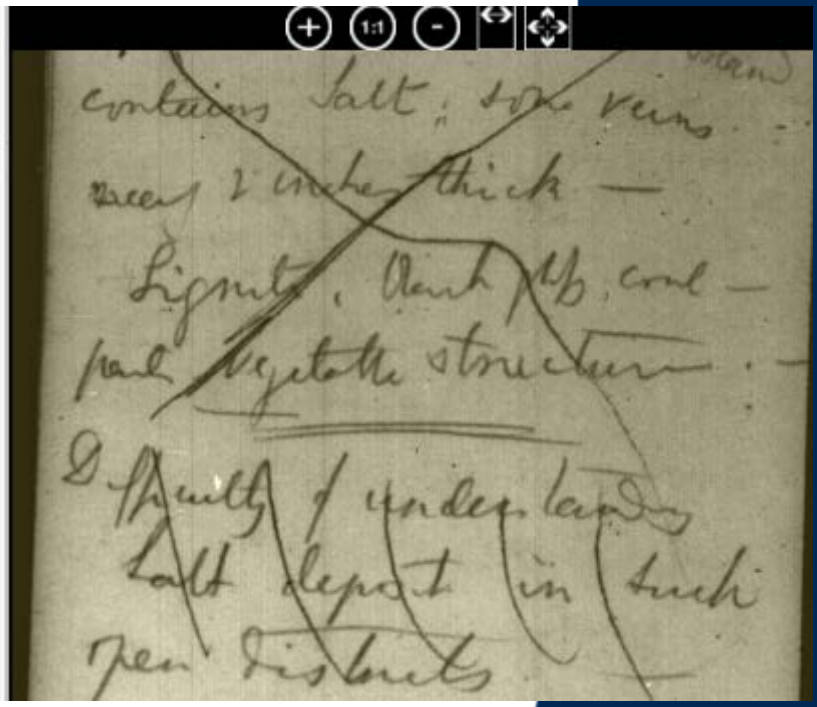
[page 15a]

Was Bellavista destroyed in 1846

Cliffs of Banos del Pujio about 200 ft above sea — 200 above Riman

Amancares 200 ft above sea 2 rivers unite Islands

Cannot understand, A. Cruikshank[1] some pebble at about that elevation on some of the hills



http://darwin-online.org.uk/content/frameset?viewtype=side&itemID=CUL-DAR121.-&pageseq=38
http://darwin-online.org.uk/content/frameset?pageseq=1&itemID=EH1.17&viewtype=side

Describe your environment

- Date
- Time
- GPS coordinates
- Temperature
- Humidity
- Equipment
- Sketches
- Vegetation cover
- Flora: identity and number
- Fauna: identity and number

UC DAVIS LIBRARY

Structure of a lab notebook entry

- Date
- Title
    - Rationale
- Protocol/Procedures
- Results
- Conclusion

} Table of Contents

UC DAVIS LIBRARY

Table of Contents

Date	Title	Page
09/27/2017	Determining salt-tolerance threshold of clover-alpha-alpha cross. Exp. 112	17
10/03/2017	Exp. 112 Results and conclusion	35
10/08/2017	Ammonia concentrations at Glory Hole	40
10/10/2017	Ammonia concentrations at Glory Hole. Results.	43

Naming conventions

- Develop naming conventions and folder hierarchy early:
  - YYYYMMDD
  - Keywords: location, variables, conditions
  - Between 25 and 60 characters
  - Beware of capitals, special characters and spaces
  - Use underscores between components

UC DAVIS LIBRARY

https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-naming

Examples using file naming conventions

20170908_luquillo_*amazona_vittata_count_nesting.xlsx*
(55 characters long)

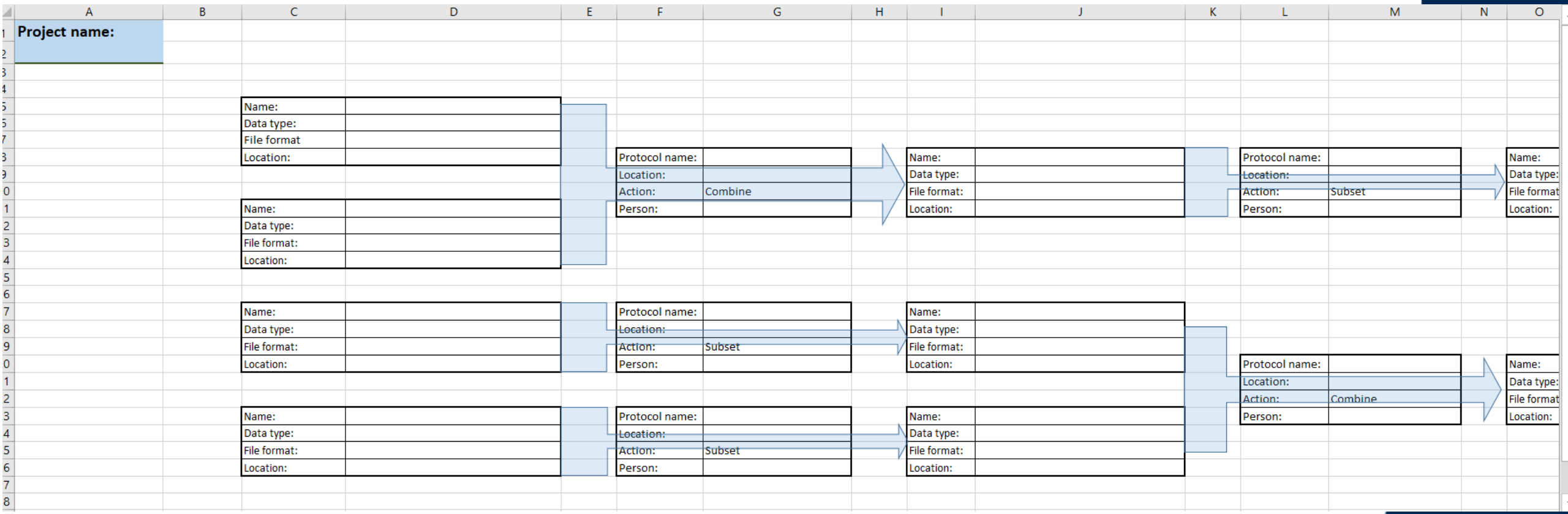
20170908_luqillo_amazona_vittata_count_rooting.xlsx

20170905_vivaldi_amazona_vittat_count_nesting.xlsx

20170905_vivaldi_amazon_vittata_count_rooting.xlsx

UC DAVIS LIBRARY

Who needs document workflow?



Activity: filenaming for "Advance Life-Support for Inhospitable Environments in Space"

--Adapted from New England Collaborative Data Management Curriculum

UC DAVIS LIBRARY

How you don’t want your data to be

Location	Soil sample weight	Species	Count	Coordinates
Dutton Island	256 g	Great egret	7	38.076976, -121.972561
Honker Bay	0.356 kg	Heron	9	38.070764, -121.916514
New York Slough	345 g	Goliath heron	three	38°01'55.2"N 121°51'40.2"W

UC DAVIS LIBRARY

Keeping data tidy

- One data type per cell
- One data point per cell
- Avoid highlighting
- Use clear variable names
- Use data validation—pre-defined consistent categories

UC DAVIS LIBRARY

Data validation in spreadsheets

Demonstration and activity

UC DAVIS LIBRARY

UC DAVIS LIBRARY

Questions?

vensberg@ucdavis.edu, office hours: Wednesdays, 2-4 PM

alemcmanus@ucdavis.edu, by appointment