

DATA SCIENCE & MANAGEMENT

Services, Training & Opportunities

Pamela Reynolds

Data Science Initiative

plreynolds@ucdavis.edu

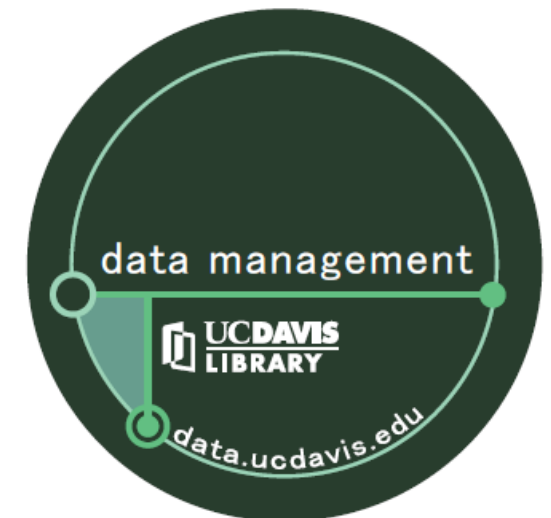
Vessela Ensberg

Data Management Program

vensberg@ucdavis.edu



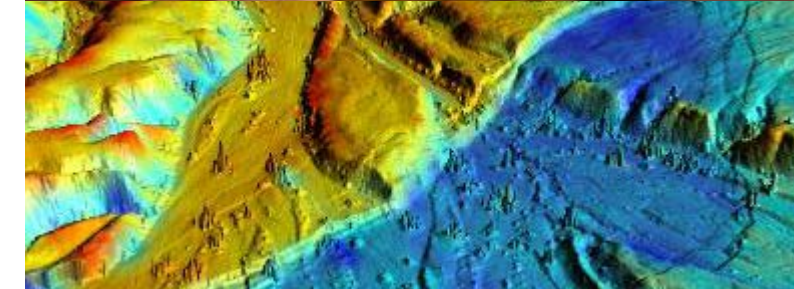
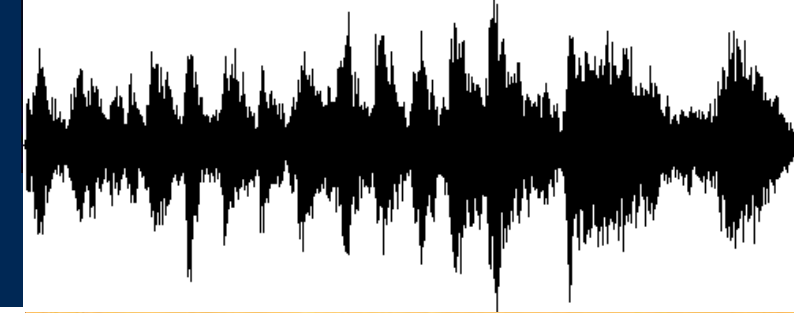
Graduate Student Orientation
September 22, 2017



“Data is plural”

facts, figures, statistics, details, specifics

- Digital information of any kind
 - Qualitative or quantitative
 - Continuous, discrete
 - Numbers, text, images, maps, audio, video, etc.



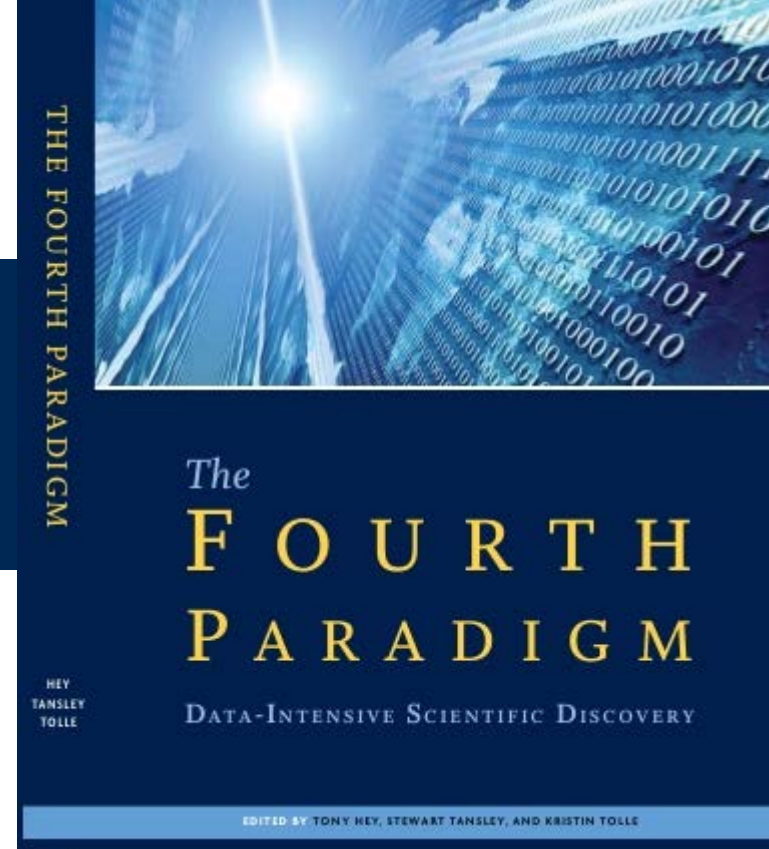
The 4th Paradigm

Data are increasing. Rapidly.

Data Science

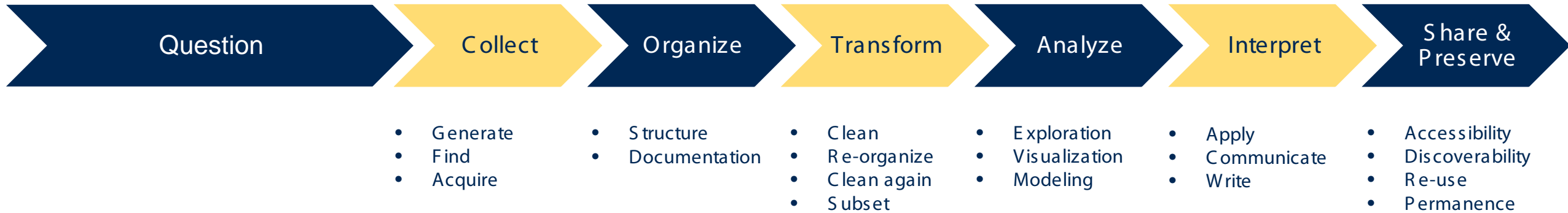
An integrative, multi-disciplinary field translating data into knowledge.

- *Empirical*
- *Theoretical*
- *Computational*
- *Exploratory*



Data-enabled research

... a non-linear pipeline





SERVICES

- Research Support
- Collaborations • Consultations
- Development • Resources



COMMUNITY

- Centers • DSI Affiliates Program
- Hackathons • Internships

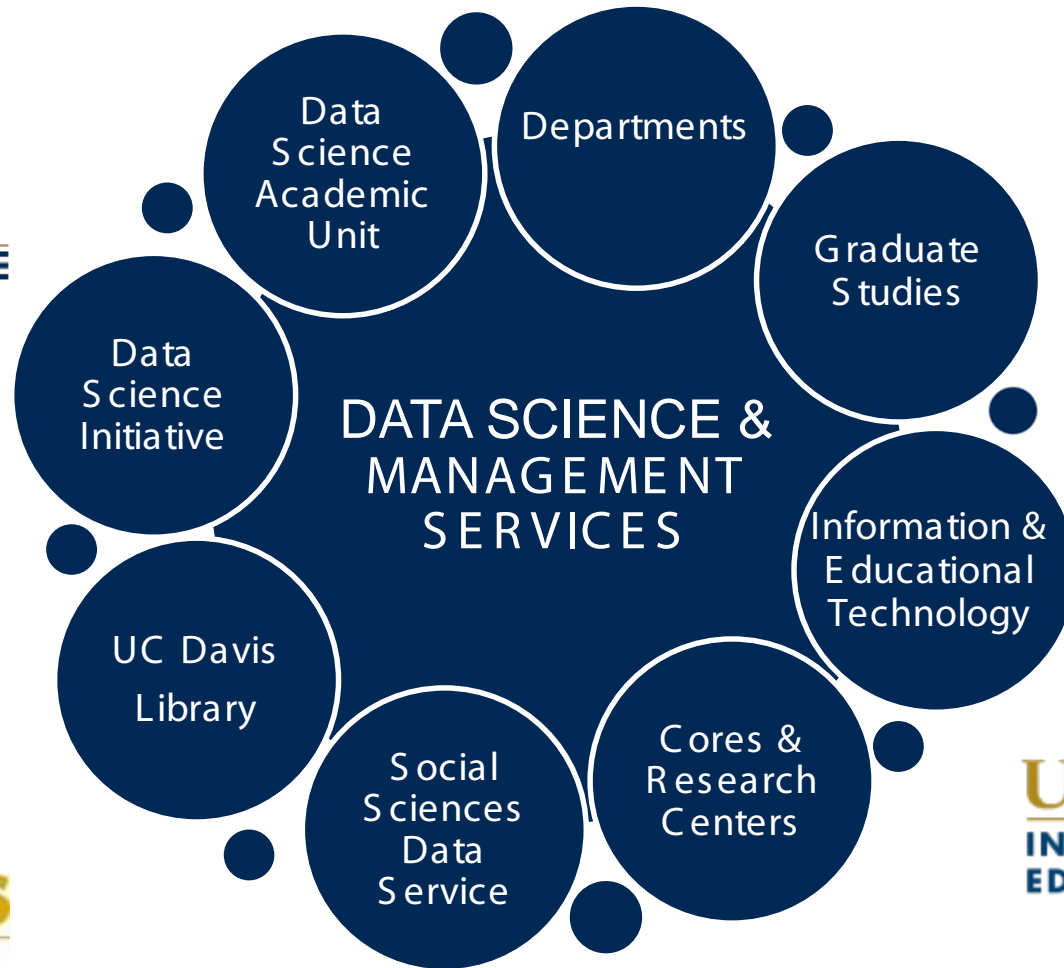
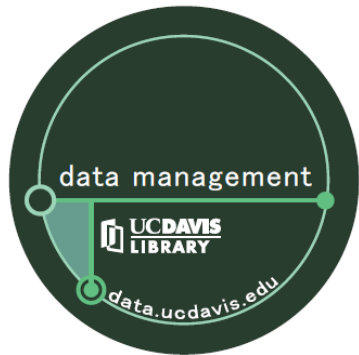


TRAINING

- Courses • Programs • Mini-classes
- Workshops • Boot Camps
- Seminars • Symposia



SERVICES



Question

Design

Collect

Organize

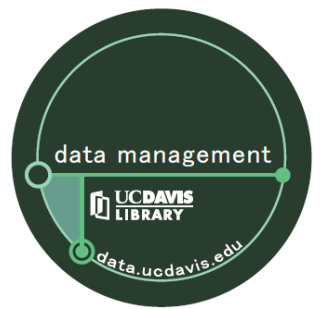
Transform

Analyze

Interpret

Share &
Preserve

- UC Davis Library
 - Research Support Services (liaison librarians)
 - Office hours & consultations
 - Finding literature, information & general statistics
 - Data Management Program
 - Office hours, consultations, projects
 - Data discovery
 - Data management
 - GIS



Question

Design

Collect

Organize

Transform

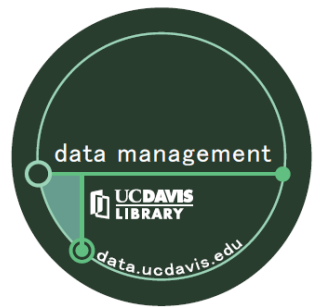
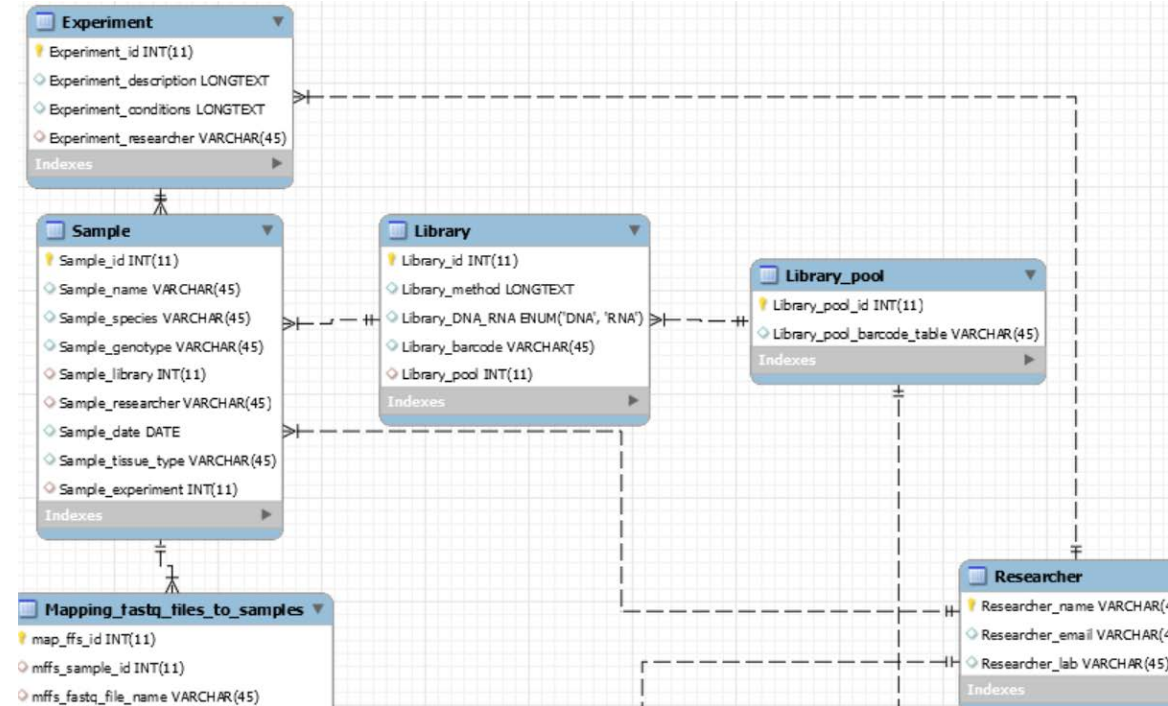
Analyze

Interpret

Share & Preserve

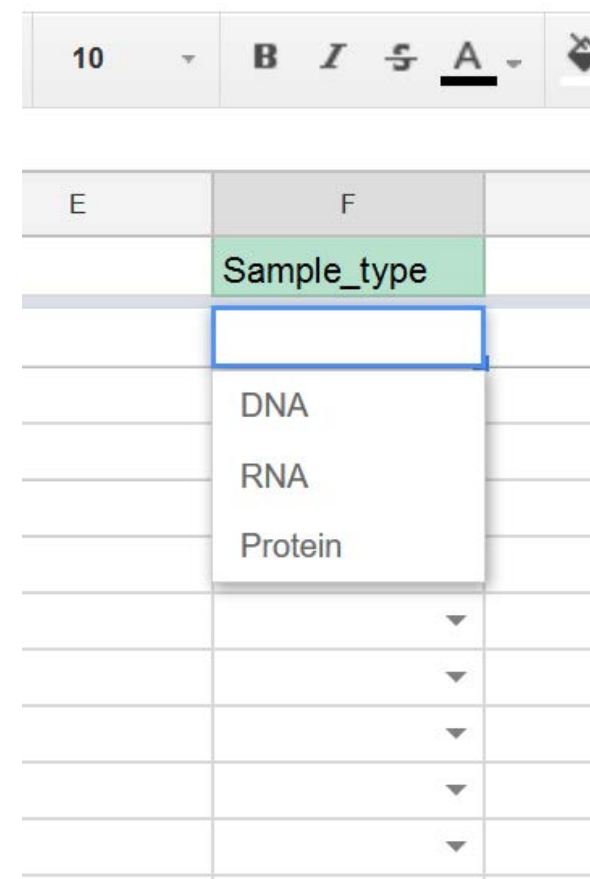
Database design

Workflows
Schema



Data collection forms design

- Design matches destination
- Data validation for consistency
- Protections and restrictions



Question

Design

Collect

Organize

Transform

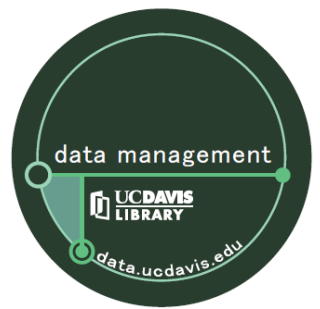
Analyze

Interpret

Share &
Preserve

Best practices tips

- 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



Question

Design

Collect

Organize

Transform

Analyze

Interpret

Share &
Preserve

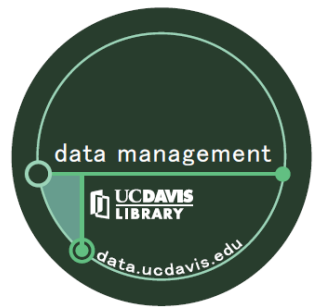
Finding data

- Need general information & statistics?

Liaison librarian



- Need raw data?
- Need data from an agency?



Question

Design

Collect

Organize

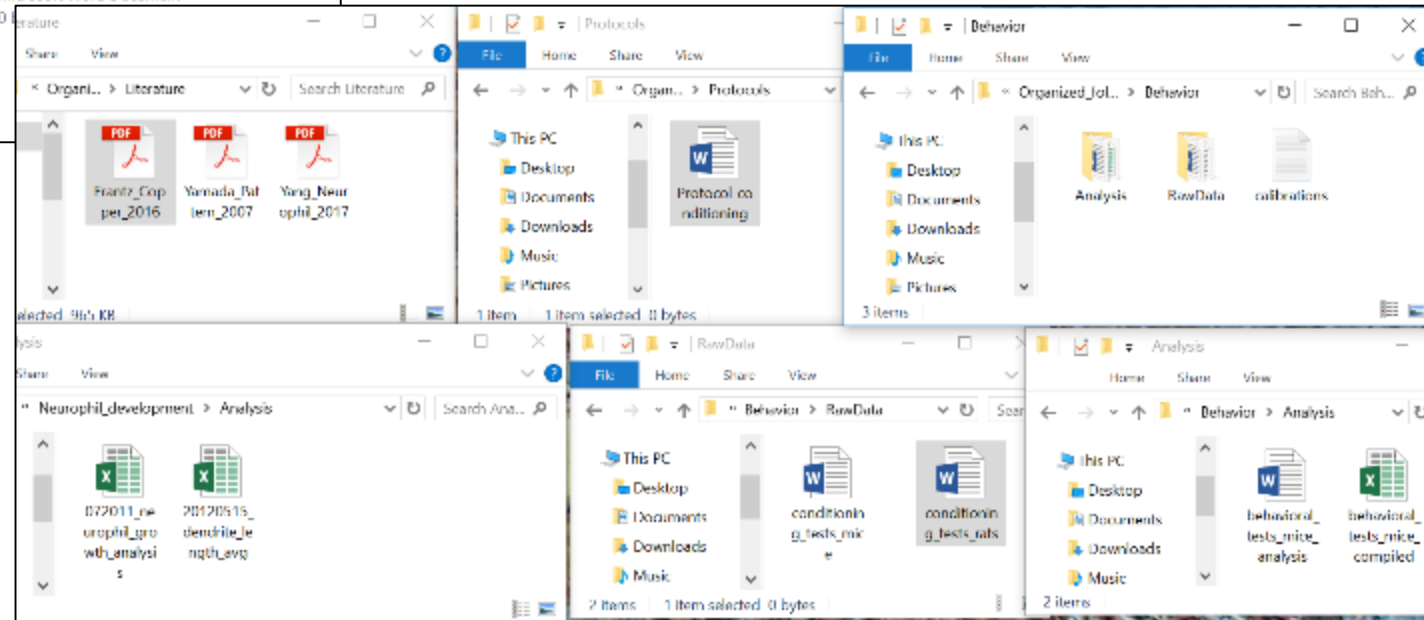
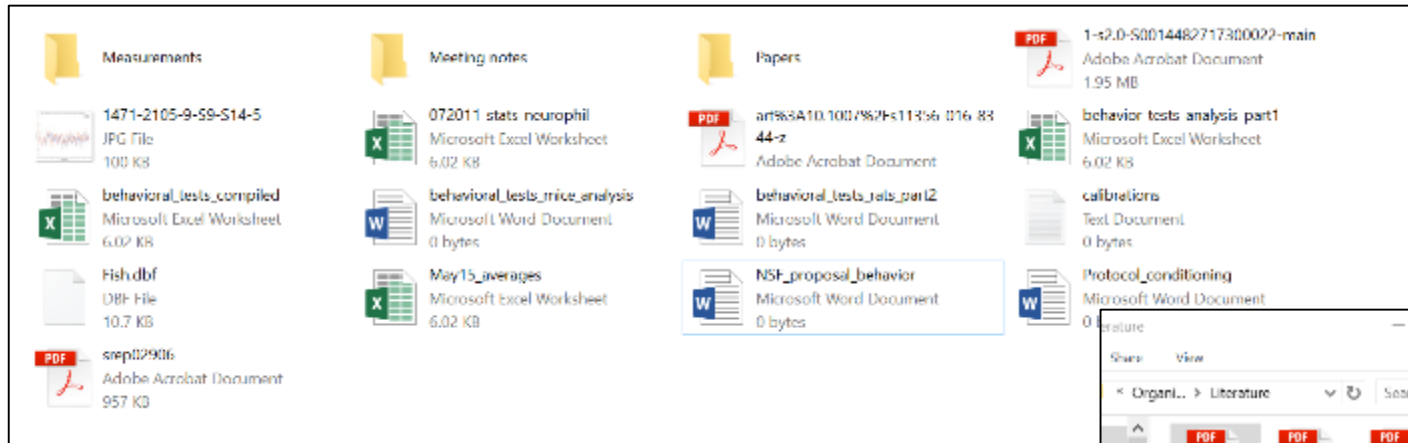
Transform

Analyze

Interpret

Share & Preserve

File organization



Question

Design

Collect

Organize

Transform

Analyze

Interpret

Share &
Preserve




Supporting documentation


- Notebooks
- Logs
- ReadMes
- Wikis



Wikispaces



 avas.geojson	add Leona Valley	19 hours ago
 digitizing_ava_boundaries.md	Update digitizing_ava_boundaries.md	2 months ago
 us_by_ava.csv	Added US Overview data.	5 months ago

 README.md

American Viticultural Areas (AVA) Project

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

<http://instructionaldev.umassd.edu/wikis/>



Question

Design

Collect

Organize

Transform

Analyze

Interpret

Share &
Preserve

S tructural readMe

- S tructure of files/folders

e.g., The top-level folders in this project are Behavior, Literature, Meeting Notes, Neurophil development, Papers and Protocols. Behavior folder contains: 1) calibrations.txt file that contains food dispenser calibrations that need to be followed to set up each conditioning experiment; 2) RawData folder; 3) Analysis folder.

- Naming conventions

The RawData folder contains observations named test_type_subject.doc.

- Relationship between files

eg. TestType is the same in the name of the protocol and the RawData and Analysis files.

- Definitions of acronyms, abbreviations, jargon

- Contact



Question

Design

Collect

Organize

Transform

Analyze

Interpret

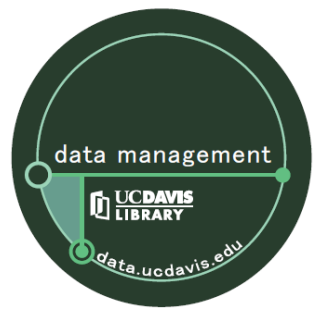
Share &
Preserve

Sharing data: FAIR

- Where does your data fit best?
- How are people going to find it?
- Can it work with other data?
- Is it re-usable on a 5-10 year timeline?



ICPSR



Question

Design

Collect

Organize

Transform

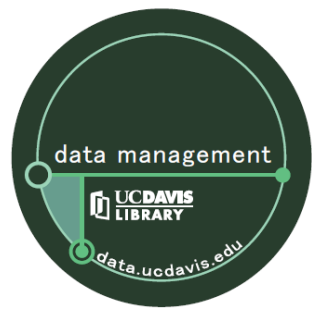
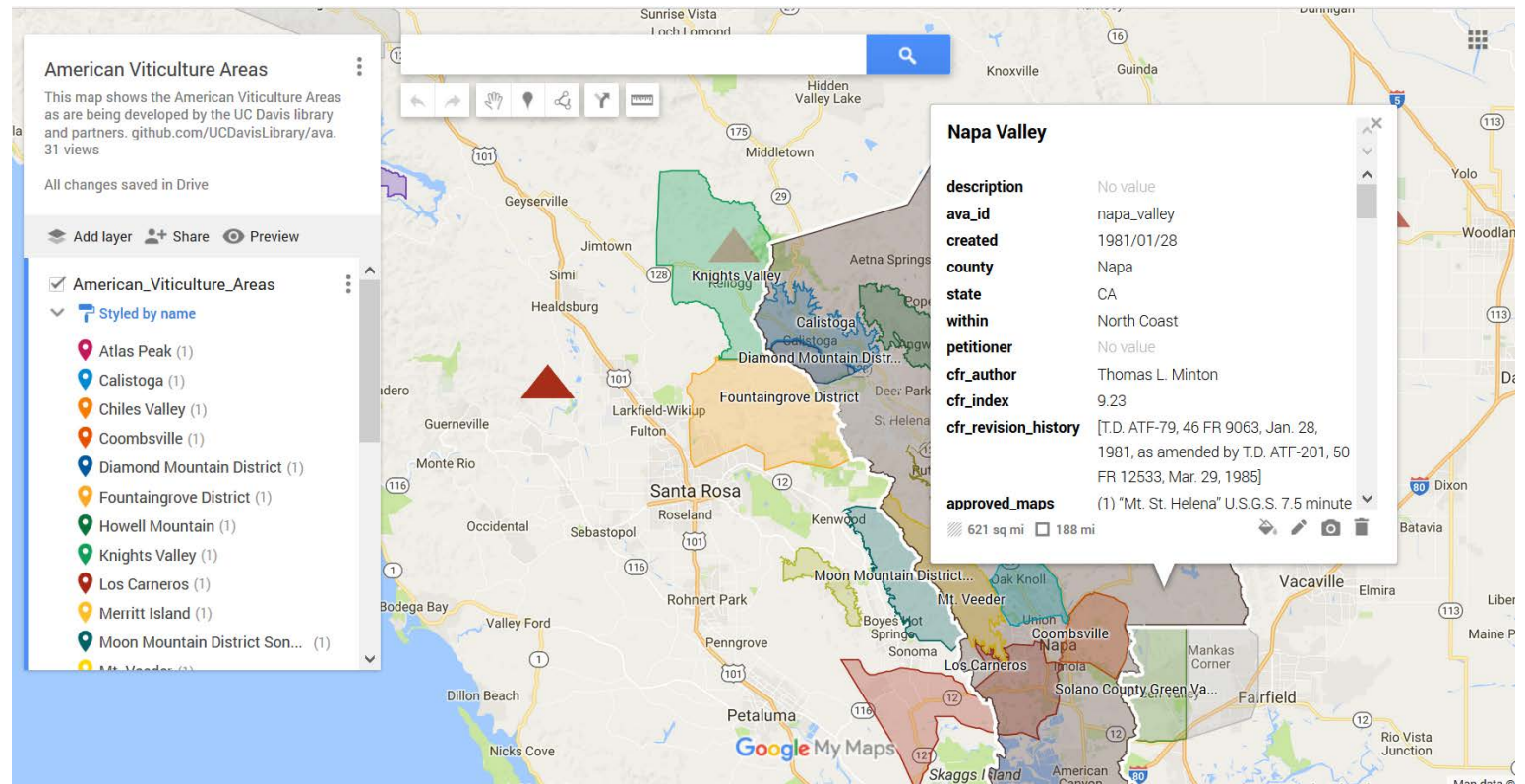
Analyze

Interpret

Share & Preserve

GIS data management

- Geocoding
- Documentation
- Analysis
- Cartography



Office Hours

- Wednesdays, 2-4 P M
DSI conference room (S hields Library 362)
- GIS: Wednesdays, 2-4 P M,
Map R oom, S hields Library (basement)



data.ucdavis.edu

Data Management

Contact us

If you have questions, want to schedule a consultation or want to involve us in a project, you can reach us at dataserv@ucdavis.edu, fill out the form below or contact the program staff directly.

Ask Me

HOW CAN WE HELP YOU?

You can use the text box to ask a question or request a consultation or a workshop.

EMAIL *

We will respond to your request to the e-mail address you provide.

SUBMIT

Vessela Ensberg

Associate Director, Data
Management Program
vensberg@ucdavis.edu
530-752-4914



Michele Tobias

GIS Data Curator
mmtobias@ucdavis.edu
(530) 752-7532



Question

Design

Collect

Organize

Transform

Analyze

Interpret

Share &
Preserve

Working across the pipeline

- Software development
- Data acquisition, cleaning, record matching
- Web scraping
- Database queries
- Data extraction
- Exploratory data analysis & visualization
- Statistical/machine learning & modeling
- Text mining
- Natural language processing
- Scaling computations
- High-performance computing
- Data technologies
- Presentation visualization
- Software engineering
- and more!



Question

Design

Collect

Organize

Transform

Analyze

Interpret

Share &
Preserve

Working across the pipeline

- Software development
- Data acquisition, cleaning, record matching
- Web scraping
- Database queries
- Data extraction
- Exploratory data analysis & visualization
- Statistical/machine learning & modeling
- Text mining
- Natural language processing
- Scaling computations
- High-performance computing
- Data technologies
- Presentation visualization
- Software engineering
- and more!



Question

Design

Collect

Organize

Transform

Analyze

Interpret

Share &
Preserve

– Office hours (Fall 2017)

- Duncan Temple Lang
 - Thursdays 12-1pm
- Carl Stahmer
 - Thursdays 1-2pm
- Postdocs & Affiliates
 - Tuesdays 1-3pm



<http://DSI.ucdavis.edu>
datascience@ucdavis.edu

– Consultations

- By appointment

– Collaborative projects

- Annual call, but come talk to us!





SERVICES

- Research Support
- Collaborations • Consultations
- Development • Resources



COMMUNITY

- Centers • DSI Affiliates Program
- Hackathons • Internships



TRAINING

- Courses • Programs • Mini-classes
- Workshops • Boot Camps
- Seminars • Symposia



- Courses
 - Statistics, Computer Science, Mathematics, Engineering
 - Other disciplines
- New Data Science Academic Unit
- Graduate Designated Emphasis
 - A “minor” for PhDs
 - Thesis must have a Data Science component
 - DS member on qualifying committee
 - Required courses & electives
- Graduate Academic Certificate
 - Required courses & electives

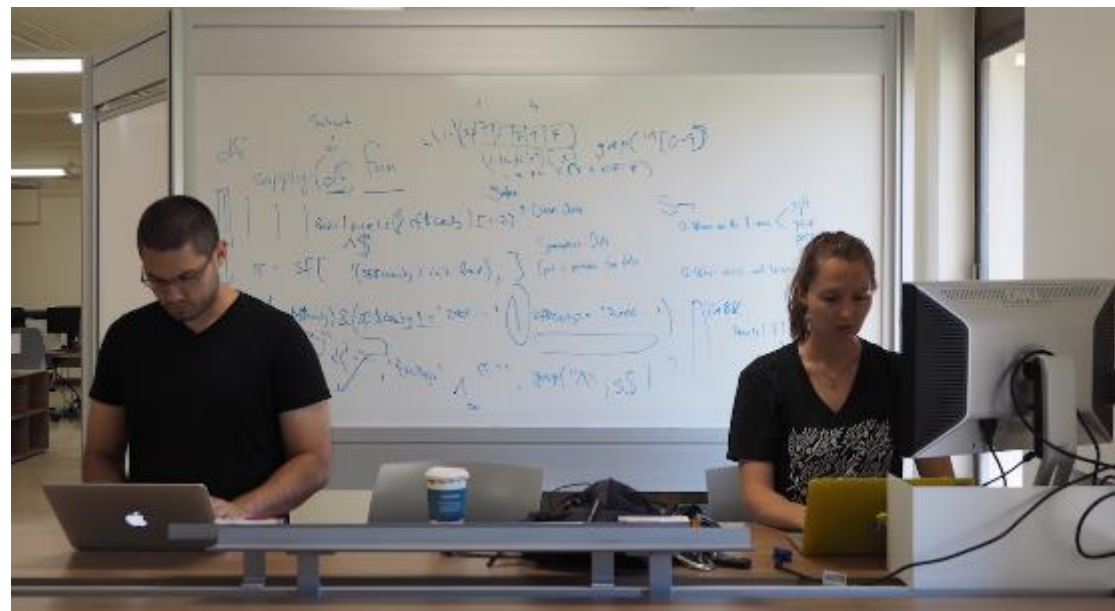
*See DSI website & email
datascience@ucdavis.edu
to learn more!*





- Workshops
 - Weekly workshops (Fridays 9am-noon)
 - Boot camps
 - Mini-Courses
- Seminars
 - Un-Seminars (Tuesdays noon-1pm)
- Affiliates Program
 - Brown-bag seminars
 - Reading groups
 - Group projects
 - Coffee hours/socials

Apply
now!
dsi.ucdavis.edu





- **DSI Workshops**

- Weekly workshops (Fridays 9am-noon)
- Boot camps
- Mini-Courses

- **Seminars**

- Un-Seminars (Tuesdays noon-1pm)

- **Affiliates Program**

- Brown-bag seminars
- Reading groups
- Group projects
- Coffee hours/socials

Apply
now!
dsi.ucdavis.edu

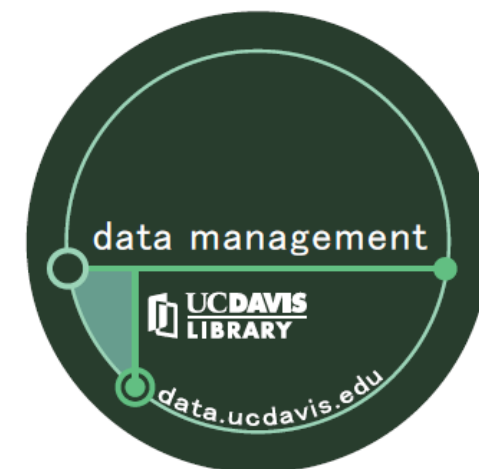
GIS Workshops

#maptimeDavis

Mondays 2:00-4:00 in the DSI Classroom

DM Workshops

- Integrated with classes
- Self-standing
 - Documentation
 - Data validation
 - Data sharing



COMMUNITY

- DSI
 - DataScience-Announce Listserv
 -  @UCDavisDataSci
 - Affiliates Program
- Davis R-Users Group (D-RUG)
- DIB Meet & Analyze (biological) Data (MAD)
- Center for Spatial Services
 - Geospatial listserv
- Bioinformatics Core
- Visualization & Interface Design Innovation



RESOURCES

- Information & Education Technology (IET)
- IET Virtual Lab
- Survey software
- User-friendly databases



- HPC
- Open Resources
- Community



Open Science Framework



GitHub

ORCID



SACRAMENTO
WOMEN
IN DATA

Links



stackoverflow

RESOURCES

For more, see the DSI & DMP websites

- UC Davis Library services <https://www.library.ucdavis.edu/services/>
 - Data Management Program <https://www.library.ucdavis.edu/service/data-management/>
 - Data Discovery <https://www.library.ucdavis.edu/service/data-management/find/>
- Data Science Initiative <http://DSI.ucdavis.edu>
- Department of Statistics
 - Statistics Lab <https://www.stat.ucdavis.edu/stat-lab/index.html>
 - Statistics RTG <http://www.stat.ucdavis.edu/research/nsf-rtg/>
- Social Sciences Data Service <http://www.ssds.ucdavis.edu/>
- Information & Education Technology <https://iet.ucdavis.edu/>
 - IET Virtual Lab <http://virtuallab.ucdavis.edu/>
- Davis R-Users Group <http://d-rug.github.io/>
- Visualization & Interface Design Innovation <http://vidi.cs.ucdavis.edu/new/>
- Keck Caves <http://keckcaves.org/>
- Bioinformatics Core <http://bioinformatics.ucdavis.edu/>
- Center for Spatial Sciences <https://spatial.ucdavis.edu/>
- Computational Infrastructure for Geodynamics <https://geodynamics.org/>
- Data Intensive Biology <http://dib-training.readthedocs.io/en/pub/>
- Software Carpentry <https://software-carpentry.org/>
- Stack Overflow <https://stackoverflow.com/>
- R Spatial <http://www.rspatial.org/>
- GradPathways <https://grad.ucdavis.edu/professional-development/gradpathways>

RESOURCES



datascience@ucdavis.edu



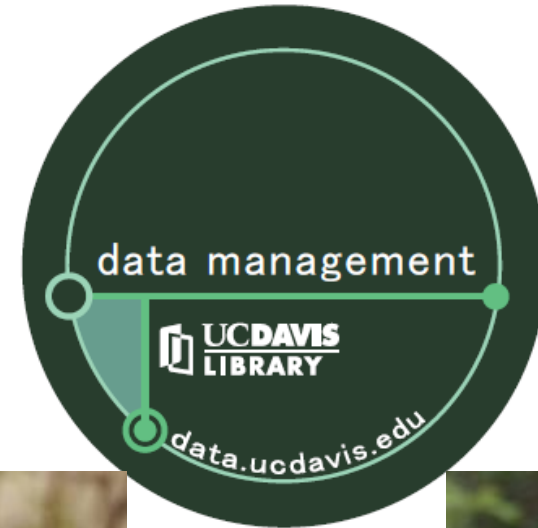
Duncan Temple Lang



Carl Stahmer



Pamela Reynolds



Vessela Ensberg

vensberg@ucdavis.edu

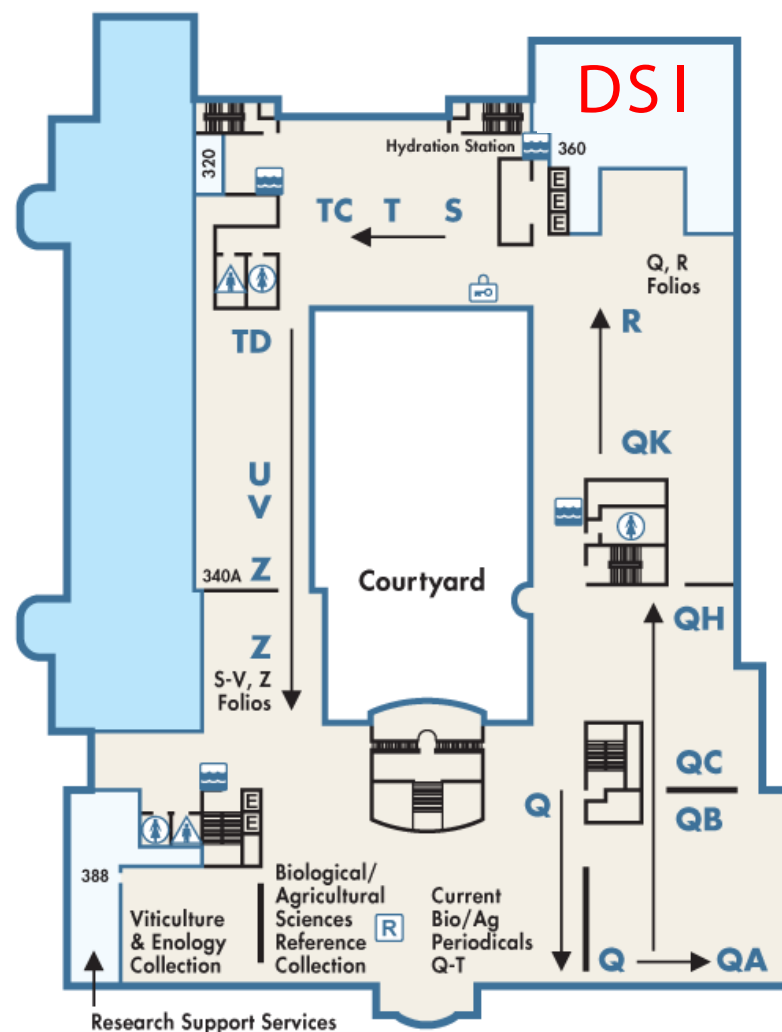


Michele Tobias

mmmtobias@ucdavis.edu

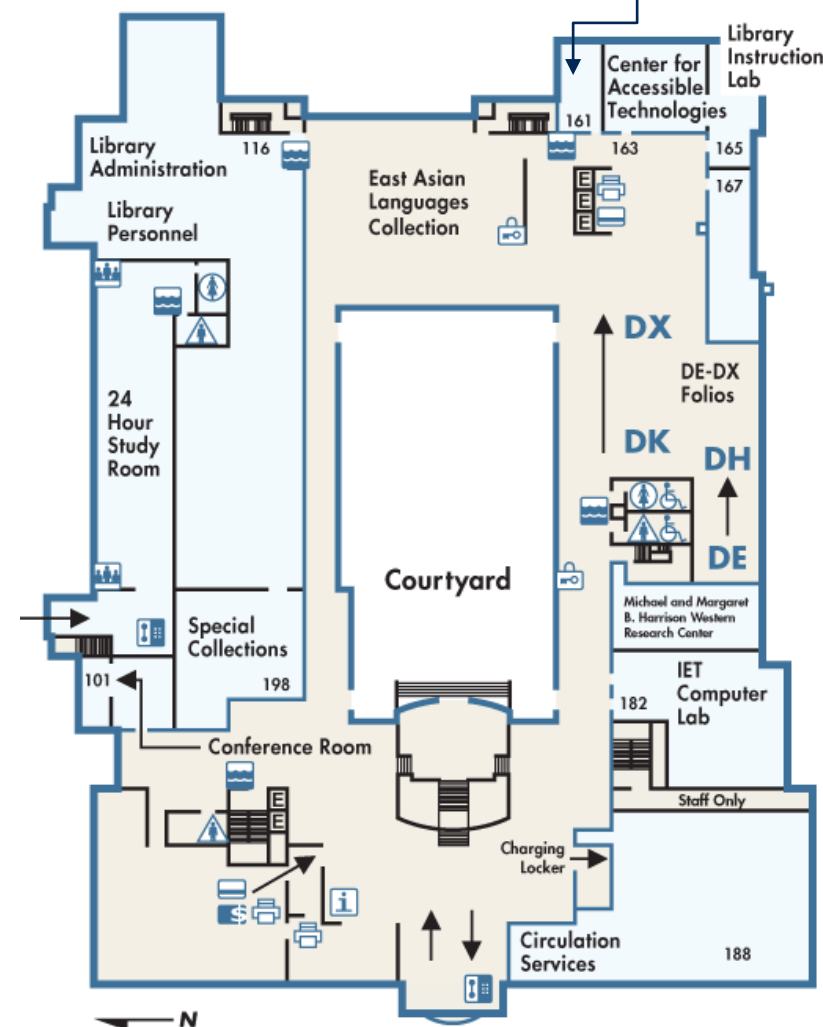
dataserv@ucdavis.edu

Data Services @ Shields Library



3rd Floor

Data
Management



1st Floor