Hello!

Erin Thomas
Presenter

Katie Wampole
Presenter

Kris Stacy-Bates
Presenter

Megan O’Donnell
Host & Zoom Wrangler
What we’re going to do today

- Background
- Parts of the Data Management Plan (DMP)
- Start a draft DMP in DMPTool
- Review Resources
- Q&A
Why is a DMP important?
High Level Goals

**Funders**
- More value for their money.
- Important data is *shared* and *preserved* responsibly and systematically.
- More dependable results.

**Researchers**
- Recognition for work that isn’t a formal publication.
- Collaboration opportunities.
- More dependable results.
U.S. DOT Workflow

**Submit Research Proposal**
- Draft DMP!
  - Obtain ORCID for each researcher
  - Include and obtain approval of DMP

**Perform Research**
- Follow data management practices in approved DMP
- Submit research description and updates to TRB's Research in Progress (RiP) database
- Project records automatically appear in Research Hub (RH)

**Package Research Results**
- Archive final dataset(s)
- Include ORCID and funding agreement number with data and on tech report documentation page

**Report Research Results to DOT**
- Submit research results (files or URLs) with ORCID, RH Display ID, and funding agreement number
Iowa DOT Data Management Plan

- Stacks and aligns with US DOT
- Consists of two parts:
  - Cover Sheet
  - **Narrative** - today’s focus

### Data Management Plan (DMP) for Iowa DOT Research Projects

<table>
<thead>
<tr>
<th>Research Data Management Documentation Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of Principal Investigator(s) or Contractor(s) and ORCID number</td>
</tr>
<tr>
<td>Current Project Title, and all previous project titles:</td>
</tr>
<tr>
<td>Iowa DOT Project Manager</td>
</tr>
<tr>
<td>Iowa DOT Project number</td>
</tr>
<tr>
<td>Other contract or grant numbers</td>
</tr>
<tr>
<td>Iowa DOT Research assigned project Digital Object Identifier (DOI), or researcher acquired DOI</td>
</tr>
<tr>
<td>TRB Research in Progress (RIP) Title, Agreement Number, and URL</td>
</tr>
<tr>
<td>Project Duration (projected): Start Date:</td>
</tr>
<tr>
<td>Do the data management requirements of the US DOT “Plan to Increase Public Access to the Results of Federally-Funded Scientific Research” apply to this project: Yes or No; and if No, why not:</td>
</tr>
<tr>
<td>Name(s) of Federal Funders, Funding Program Name(s), Agency Code(s) and or Contract/Grant numbers</td>
</tr>
<tr>
<td>DMP Version</td>
</tr>
<tr>
<td>Data DMP amended, if any:</td>
</tr>
<tr>
<td>Name and ORCID number of each author</td>
</tr>
<tr>
<td>Facilitator links or identifiers assigned to this project, datasets, reports, or peer reviewed publications generated by this project</td>
</tr>
<tr>
<td>Name and URL of all peer reviewed publications which have been generated from this project</td>
</tr>
</tbody>
</table>

This table is to be filled out as completely as possible before the beginning of the project, and updated as needed, including at the end of the project, and after, as derivative publications are created.

[Note: Guidance for using this template can be found at: [http://publications.iowa.gov/id/reprint/21912](http://publications.iowa.gov/id/reprint/21912)]
Today’s workshop includes time to start a draft in DMPTool.

- Visit [dmptool.org](http://dmptool.org) and login or create an account.
- Click Create Plan.
- Choose Iowa Department of Transportation in the last drop-down box.
Required components of the DMP
Description
Technical Information

- Types of data: tabular, sensor, text, imaging, audio-visual, etc.
- Types of file formats: .csv, .txt, .tiff, .pdf, etc.
- Note special tools or software used to work with the data.
Relational Information

- Additional outputs associated with the data (Papers, reports, etc.)
- Was the data part of a unique event? Or can it be reproduced?
- Give identifiers associated with the data when available (DOI commonly used).
Documentation & Organization
Organization Goals

- Stay consistent. Make sure the organization means something.
- Usable by all project members (including yourself).
- Readable by both humans and machines.
Documentation Goals

- Data is only useful if it has context and meaning.
- Benefit you, your team, and other potential partners/viewers.
  - Where/how data was collected, generated, complied, etc.
  - What was done to data (raw vs cleaned data).
  - File naming conventions and organization.
### Codebook/Data Dictionary

Tabular file that has information about...

<table>
<thead>
<tr>
<th>Variable_label</th>
<th>Variable_name</th>
<th>Measurement_unit</th>
<th>Allowed_Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q24</td>
<td>Question 24: do you own a pet?</td>
<td>Numeric</td>
<td>1 = yes; 0 = no; blank = NULL</td>
</tr>
<tr>
<td>Q25</td>
<td>Question 25: Do you like cats?</td>
<td>Numeric</td>
<td>1 = yes; 0 = no; blank = NULL</td>
</tr>
<tr>
<td>Q25.1</td>
<td>Question 25 part 1: Why?</td>
<td>None</td>
<td>free text</td>
</tr>
</tbody>
</table>

An excerpt from a codebook.
Plain text file that has information about...

**COLLECTION INFORMATION:**

Time period: Data collected between 2006 and 2019, uploaded to repository in October of 2021

Location: Eagle Lake vicinity, Lassen County, California, USA

**FILE DIRECTORY**

- Eagle Lake Physiology
  - README.pdf
  - DataCodebook.xlsx
  - ELPphysData.csv

**FILE LIST**

- README.pdf: README file
- DataCodebook.xlsx: Codebook containing explanation of variables found in data file
- ELPphysData.csv: Comma-separated file containing information on experimental animals and physiological measures

**DATA COLLECTION METHODS**

These data were collected from natural populations of western terrestrial garter snakes (Thamnophis elegans). Blood samples were collected at time of capture and following a 3hr capture-restraint protocol (stress-induced). Samples were analyzed for circulating plasma corticosterone, blood glucose, and heterophil-to-lymphocyte ratios. All data were collected under Iowa State University IACUC protocol 3-2-5125-J under Dr. Anne Bronikowski.

**SOFTWARE**

Name: R
Version: Macintosh 4.0.2
URL: [https://www.r-project.org/](https://www.r-project.org/)
Developer: The R Foundation for Statistical Computing

Name: SAS
Version: 9.4
Developer: Analytics Software & Solutions
DMPTool
Questions 1 and 2
Security
Data Security: Physical Considerations

In a disaster situation…
◉ Do you have backups?
◉ Who owns your shared/cloud folders?
  ➢ What happens if they leave?

Think about:
◉ Making multiple backups (3-2-1).
◉ Using version control.
Security: Access Concerns

Do you have data that might require access restrictions?
- Privacy and confidentiality
- Access limitations
- Safeguards to prevent unauthorized access

Who can help with these concerns?
You don’t have to share everything.
If you plan to share data...

- **What** data will be suitable to share?
- **When** will it be available?
- **Who** will use the data?
- **How** will they get access?
- Does your documentation support re-use?
Keep in mind!

- If you don’t share, you should provide a justification for why.
- “Data available upon request” is not a sharing policy.
- Data repository: a support infrastructure for making data available and **findable**.
- Preparing your data for sharing may require additional work.
DMPTool

Questions 3 and 4
Preservation
What do we mean by preservation?

- Providing for data to be used after research is completed.
- Usually performed by special systems that safeguard file stability, recoverability, and reusability over time.
- Be selective in which data to preserve.
Keep your data usable for the long term

- Get it out of proprietary file formats.
- Find the data a long-term home that meets requirements and contractual agreements.
- Don’t use DIY solutions like lab and project websites - they don’t last.
- Do budget for the resources needed for preserving your data.
So where should I preserve my data?

- A data repository
  - If you’re an ISU affiliate, you can use ISU’s DataShare
  - Open systems like Dryad, Dataverse, Figshare, and Zenodo
- Other community-accepted data storage facility
  - Must meet US DOT requirements for preservation
(Your data should) be persistent

- Data must be discoverable and accessible as well as preserved.
- Datasets need persistent identifiers (such as DOIs).
- IDOT data must be publicly accessible for 10 years from the end of the contract period.
Roles & Responsibilities
With great data comes great responsibility

PIs are ultimately responsible for project data, but usually delegate some roles.
People Information

- Who is in charge of what?
- Who is the back-up for what?
- Is contact information needed?
Possible roles for team members regarding data include:

- Training
- Quality control and monitoring backups
- Updating the data management plan
- Ensuring public access, such as reporting data sets to the U.S. DOT’s Repository and Open Science Access Portal (ROSA P)
- Long-term care of the data
DMPTool

Questions 5 and 6
Where to get more help

- Iowa DOT Project Manager
- Iowa State Univ. Library
  - Data Management Plan Guide
  - datashare@iastate.edu
- US DOT
  - Public Access Plan FAQs (data section is halfway down)
  - Creating Data Management Plans
  - Data Repositories Conformant with the DOT Public Access Plan
    - Guidelines for Evaluating Repositories for Conformance
Questions?