INTRODUCTION TO DATA MANAGEMENT PLANS

Iowa State University Library
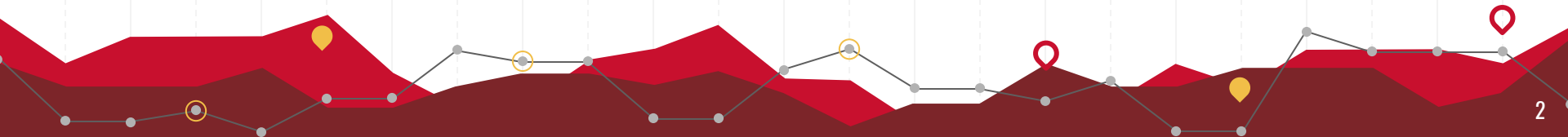
Your Instructors

Megan O'Donnell

Research Data Services Lead

mno@iastate.edu
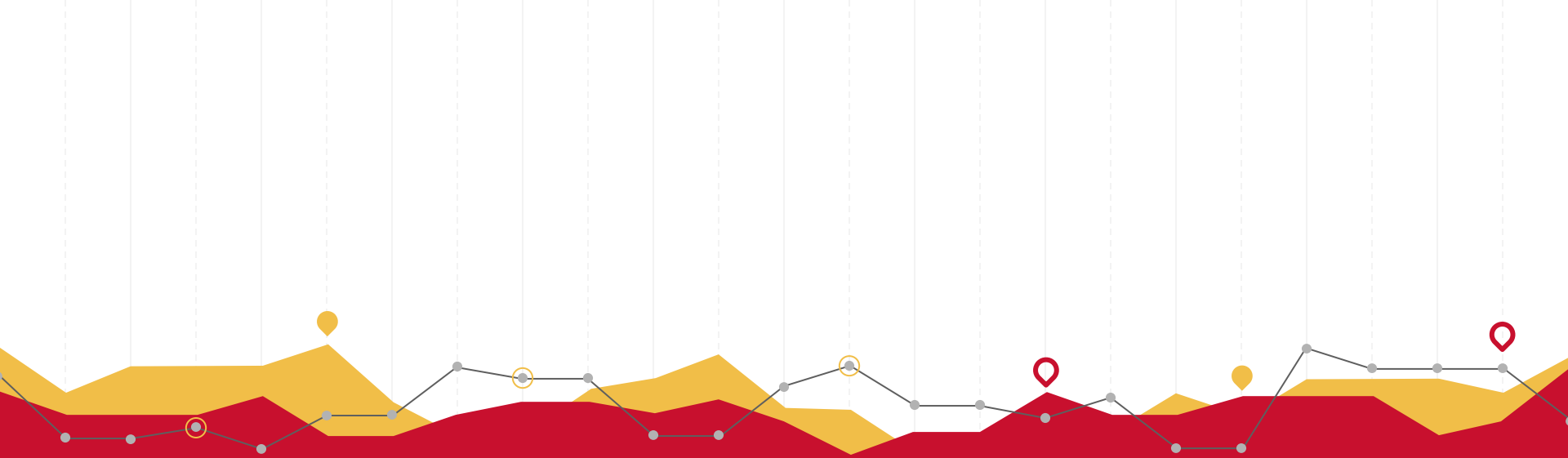
Kris Stacy-Bates

Engineering Librarian

kksb@iastate.edu

Outline

- Introduction & backstory
- Q&A $^{no.}$1
- Elements of Data Management
- Q&A $^{no.}$ 2
- Activity and discussion
- Resources
- Q&A $^{no.}$ 3

What is research data?

and what do we do with it?

1

What is data?

- Data are observed and recorded **facts**.

- Size, shape, and structure are highly variable.

What is digital research data?

- Factual **digital** material collected or generated in **pursuit** of **research or scholarship**.
  (i.e. factual digital info about the thing you are studying)

What's not research data?

- **Creative work**
  - Papers, performance recordings, illustrations, etc.

- **Business records and transactions**
  - emails, drafts, reports, lab notebooks, bills, etc.

- **Trade secrets, commercial information, confidential materials, personnel and medical information, etc.**

You can collect information on these types of documents (i.e., primary sources) but they aren't themselves "research data."

How we use research data

- **Publications** are research stories supported by **evidence** – aka **data**.

- **Tables and figures** are "evidence summaries" that support the narrative.

- The research data is **not in the paper**…

Sharing and Managing data

Changing practices and expectations

2

Why are we discussing data sharing with data management?

- Funders and journals **require it**.

- A **key part** of a data management plans and may affect:
  - Data collection and formatting
  - Budgets, timelines, and staffing

You cannot share data unless it is well managed

What should be shared?

- **At minimum:** data (and scripts) used to create published visualizations.

- Data that's **not sensitive, confidential, private, or prohibited by special contract**.
  ex: related to patents, from industry partners, location data for sensitive sites or species,

- Be **careful** and **thoughtful** with data **about people** (PII, human-subject data) and their things and culture.
  ex: re-identification risk, sensitivity of info, misuse, etc.

What is a Data Management (and Sharing) Plan?

- **In general:** an outline of what the data is and how it will be cared for -- during and after research

- **For proposals:** convince peers that the data will be managed **safely** and **responsibly** and **shared and preserved** in a similar manner
  …and that you are following rules and regulations that apply.
  … and typically **limited to 2 pages.**

Data Description

1. Suisun Marsh fish study i
standard length), water quali
1980 using a variety of samp
meter, Secchi disk, and GPS
hotspots of occurrence in the

2. Suisun Marsh ponds aqua
conditions, and primary and
and zooplankton density. Da
but in transit and at fixed loca
a, and zooplankton trawls. D
from 2017 to 2020. Compara
managed wetlands will be co

3. Waterfowl food availabilit
processed and evaluated for
experimental trials that used
Data were collected during p
across Suisun Marsh from
explicit Waterbird Agen
with food availabili
astern po

Data Management (and Sharing) Plan tips

- **Justify** restrictions and limitations, if any (safety).

- Use existing **standards** and systems whenever possible (responsible).

- Implement **quality control** and **safety** measures (sharing and preservation).

  (don't worry we'll provide examples)

Q&A No. 1

Are there any questions?

Elements of a Data Management Plan

What you need to do and plan for

3

Data Management in a Nutshell

- Own it.
- Know it.
- Organize it.
- Document it.

- Copy it.
- Secure it.
- Deposit it.

Own it

Take responsibility.

- Allocate time to care for your data.

- Know your responsibilities and your limits.

- Plans mean fewer mistakes.

- Seek help when needed.

- Start early.

Know It

You need to know **what you have** (or will have) to manage it effectively.

- **flavors**:

  quantitative, qualitative, observational, experimental, spatial, temporal, instrument-generated, models, simulations, raw, normalized, primary, secondary, etc.

- **types** and **file formats**

  - **Type**: image, **File format**: jpeg, gif, tif, png

Spicy data
- **Big** data
  - storage and processing
- **Other people's** data
  – hard to understand
- **Sensitive** data
  – may need extra security

Organize It

- Use rules and stay consistent.
    - Have a logical system.
    - Use **standards** whenever they apply (ex: international date and time standard)
- Clearly label files and folders with descriptive names.
    - Final paper vs. BIO554_F23_final_ELT
- Test it: can other people understand the names?

Document It

- Keep logs, make keys and keep them with the data.
    - readme, metadata, data dictionary, codebook, etc.
        - ex. "ph 01" = "soil ph taken at field site A with Kenzier tester."
- Assume that others will need and use your data documentation.
    - Avoid shorthand and code (unless defined)
    - Don't assume that people will "read the paper."
   - Cite and link to sources, methods, and tools.
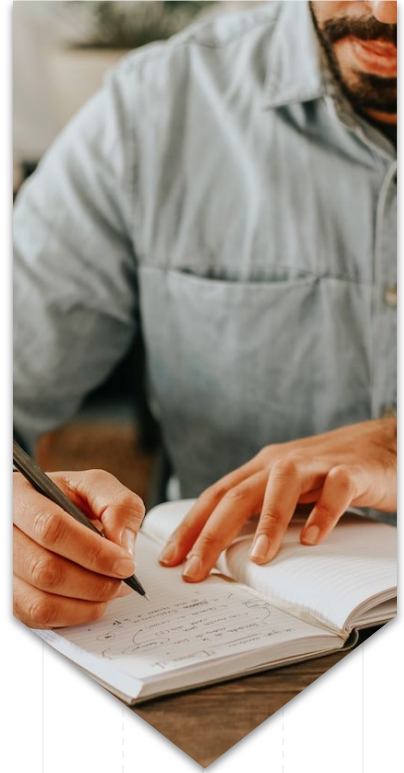
Copy It

- You must have copies and backups!

- Use version control/file versioning
  - **Save as**… create a new copy when changes are made without overwriting the previous version(s).
  - Most cloud storage supports this as does Git and GitHub.

- Try to follow the **3-2-1 Rule**:

  - 3 copies, 2 formats, 1 off-site
  - Using cloud storage is a shortcut: more than 3 copies are made and they're off-site.

Secure It

Plan for the "worst case" to recover quickly.

- Know who's responsible.
    - Machines and equipment – security of, backups, etc.
    - Remote access – access to, oversight, etc.
    - Physical locations – locks, fires, floods, vandals, etc.

- Ask for help and work with IT experts
    - They know the policies and standards and want to prevent problems for everyone.
    - How much and what kinds of precautions depend on laws, policies, and the nature of the research.
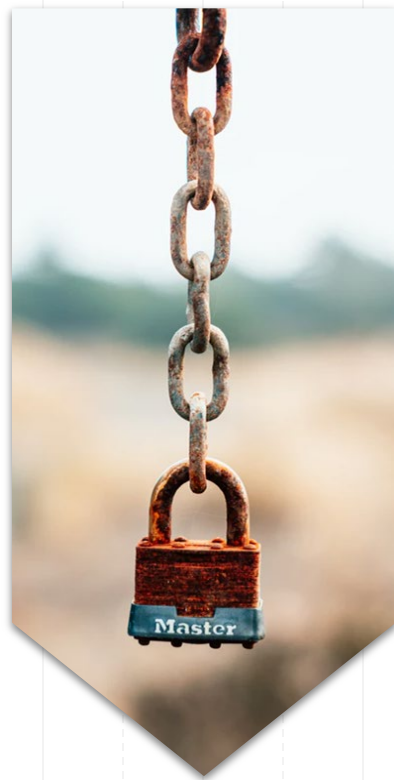
Deposit It

- Determine what to keep, what to share, and what to toss.

- Get it out of proprietary file formats!

- Find the data a *long-term home* that meets both legal requirements and contractual agreements.

- Don't use DIY solutions like lab and project websites, they don't last.

Sidebar: Data Repositories & IS

Data Repository: an online system specifically used for sharing, publishing, and archiving all types of research data.

DataShare: ISU's open access data repository.

- Archiving in DataShare is **free** and librarians will review your data and help you prepare it for sharing.

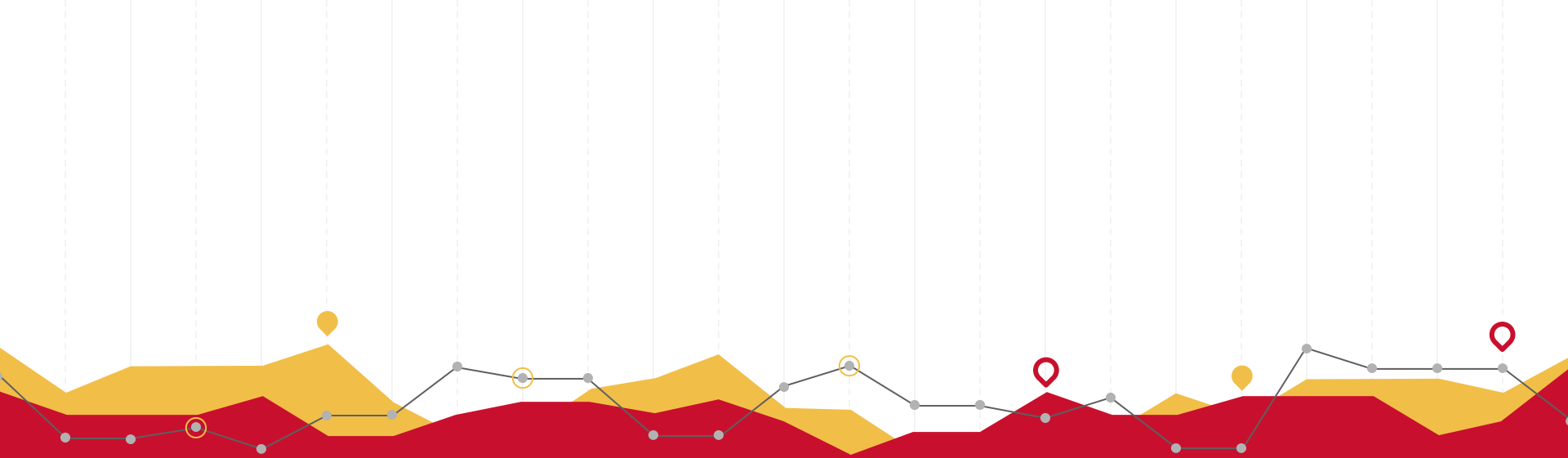- Published data sets are assigned a DOI and will show up in search results.

Contact datashare@iastate.edu to include DataShare in your DMP!
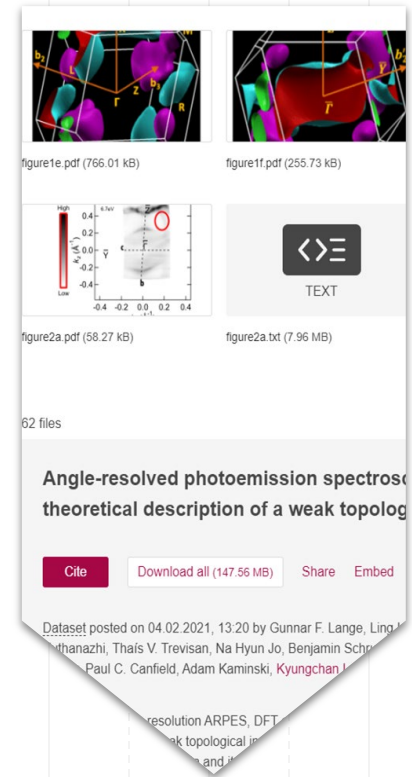
Q&A No. 2

Are there any questions?

DMP Bingo

4

Apply what you've learned!

Rules and Regulations

1. Bingo is a game of chance.

2. You will mark off squares that match what you find – or don't find in a DMP.

3. 5 squares in a row, column, or diagonal is a **bingo**!

B	I	N	G	O
Data is stored on workstations	Provides file size estimates	Uses cloud storage	Provides a target date for sharing data	Using a data repository
States data reuse conditions	Data seems secure and safe	Undefined acronym(s)	Mentions metadata	File formats are common or open
Data on sensitive subjects is well explained	Data does not seem secure and safe		Data is not documented	Can't tell which data will be shared
Data is not sensitive	Identifies responsible parties	Data will be shared promptly	Data is documented!	"Sharing" through publications
USB or hard drives	Has copies or backup system	Can't tell which data will be preserved	Not using a data repository	File formats are weird or proprietary

How the Game is Played

1. Read and review your DMP
   1. Poli Science Polling, or
   2. Plant Breeding Partnership

2. Work together and try to score a BINGO!
   (You can use the internet if needed)

B	I	N	G	O
Data is stored on workstations	Provides file size estimates	Uses cloud storage	Provides a target date for sharing data	Using a data repository
States data reuse conditions	Data seems secure and safe	Undefined acronym(s)	Mentions metadata	File formats are common or open
Data on sensitive subjects is well explained	Data does not seem secure and safe		Data is not documented	Can't tell which data will be shared
Data is not sensitive	Identifies responsible parties	Data will be shared promptly	Data is documented!	"Sharing" through publications
USB or hard drives	Has copies or backup system	Can't tell which data will be preserved	Not using a data repository	File formats are weird or proprietary

DMP Bingo Discussion Topics

- What was done well?

- Was anything missing?

- Could anything be improved?

- Would you fund research with this data management plan?

B	I	N	G	O
Data is stored on workstations	Provides file size estimates	Uses cloud storage	Provides a target date for sharing data	Using a data repository
States data reuse conditions	Data seems secure and safe	Undefined acronym(s)	Mentions metadata	File formats are common or open
Data on sensitive subjects is well explained	Data does not seem secure and safe		Data is not documented	Can't tell which data will be shared
Data is not sensitive	Identifies responsible parties	Data will be shared promptly	Data is documented!	"Sharing" through publications
USB or hard drives	Has copies or backup system	Can't tell which data will be preserved	Not using a data repository	File formats are weird or proprietary

Resources for Writing a DMP

and for data management in general

5

DMP Tool
(dmptool.org)

- **Templates** and **guidance**
- Walks you through funder **requirements**.
- Highly recommended and free!

dmptool.org

Library Resources

- **Your Librarians**!
  - We offer consultations and draft reviews (email datashare@iastate.edu)

- **Data Management Plan Guide** tinyurl.com/iastateDMP
  - Covers everything we went over today and more.
  - Has a worksheet, FAQs, definitions, and examples.
  - Links to policies, requirements, and learning tools.

Need help?

Data Management Questions
datashare@iastate.edu

IT Questions
researchit@iastate.edu

Quick Access

IOWA STATE UNIVERSITY
DataShare

DMPTool

Q&A No. 3

Are there any remaining questions?