IUPUI University Library Center for Digital Scholarship

Heather Coates, Digital Scholarship & Data Management Librarian

Improving data management in academic research: Assessment results for a pilot lab

Background

Common practices for data collection, storage, organization, documentation, sharing, re-use, and preservation are often suboptimal. Issues often arising from common data practices include data loss, corruption, poor data integrity, and an inability to demonstrate the provenance (i.e., the origin) of the data. Ineffective data management can result in data that are unusable for re-use and re-analysis. However, effective data management practices exist to support data integrity, interoperability, and re-use. These practices maximize the value and potential impact of any particular dataset. In light of the gap between common practice and known effective strategies, we developed an intensive lab curriculum to train students and research support staff in implementing these strategies. This lab addresses the lack of formal data management training available on our campus and targets key processes in the data life cycle, promoting strategies that facilitate generation of quality data appropriate for re-use.

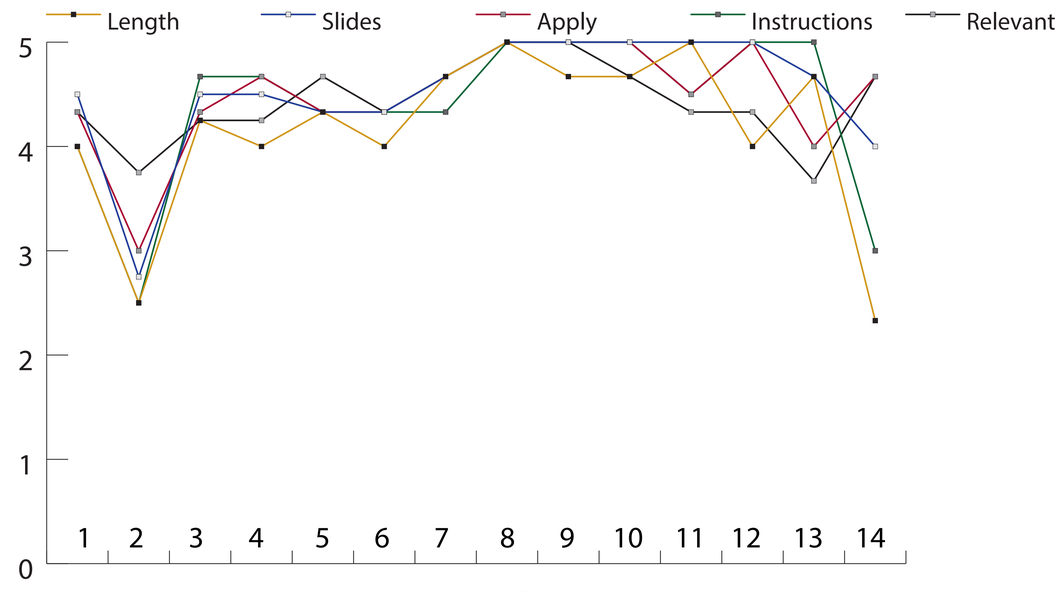
**See handout for learning outcomes & activities*

Instructional Products

- Data Management Plan
- Map of data outcomes
- Documentation Checklist
- Data Quality Standards
- Screening & Cleaning Checklist

Heather Coates
hcoates@iupui.edu
@landPangurBan
coateshl.wordpress.com

Evaluations



1: Intro to Research Data Management
2: Data Management Plans & Planning
3: Ethical & Legal Obligations
4: Organizing Data & Files
5: Project & Data Documentation
6: Quality Assurance & Quality Control
7: Data Collection
8: Data Entry
9: Data Screening & Cleaning
10: Automating Tasks
11: Data Rights, Protection, Access
12: Data Sharing & Reuse
13: Data Attribution & Citation
14: Overall

Rating Questions

- Length: The length of the workshop was appropriate for the content.
- Slides: The slides provided the information necessary to complete the exercises.
- Apply: The exercises will help me to apply the strategy(ies) in my own research.
- Instructions: The instructions for the exercises were clear.
- Relevant: The strategy(ies) are relevant to my own research.

Rating Scale

1. Strongly Disagree
2. Disagree
3. Neutral
4. Agree
5. Strongly Agree

Oops! No Product Assessment

- Time cut short due to weather.
- Case study not detailed enough.

Muddiest Points

- Case study was too vague
- Analysis procedures
- Terminology for data mapping
- Metadata and why it's important

Valuable Points

- Data cycle/project life cycle
- Creating data management plans
- Proper technique for file naming
- Backup systems for files
- Master files
- Version control
- QA vs. QC
- Discussing consistency in data
- Setting up variables
- Cleaning checklist
- Using scripts (i.e., automation)

Feedback Informing Revisions

√ Further develop and focus the case study to support meaningful completion of activities in the allotted time.

√ Change the scheduling format: 4 sessions of 2 hours each.

√ Include more information about campus resources.

√ More emphasis on the importance of selecting appropriate tools (e.g., "Excel is horrible for analysis - stress this more.")

√ More emphasis on documenting study changes.

More Information

1. Coates, H. L. (2014). Research & Practice Notebook: Data Management Lab 2.0 [Web log posts]. Retrieved from http://coateshl.wordpress.com/research-notebook/.
2. Coates, H. L. (2013). Developing a data management lab: Teaching effective methods for health and social sciences research. Poster presented at the Data Information Literacy Symposium: West Lafayette, IN.
3. Coates, H. (2012). Practical data management: Enabling graduate students and staff to function as ethical actors. Poster presented at the 2013 conference of the Association of College and Research Libraries, Indianapolis, IN.

IUPUI