

Practical data management: Enabling graduate students and staff to function as ethical actors

Heather Coates, IUPUI University Library

Digital Scholarship & Data Management Librarian

hcoates@iupui.edu

http://ulib.iupui.edu/digitalscholarship/dataservices

INTRODUCTION

High quality data is desired by all researchers, but few receive formal training in data management. Graduate programs provide in-depth training related to issues of research design and analysis, but fail to adequately equip students with specific strategies for managing research data that also ensure research integrity. Though some training programs sponsored by the National Institutes of Health (NIH) and National Science Foundation (NSF) include a responsible conduct of research (RCR) education requirement, it is typically addressed by discussing theoretical situations. Most research institutions rely on the CITI Training Modules and seminar courses. Data management is one of nine components of the RCR. The RCR framework is a useful one for teaching practical data management strategies because it speaks to outcomes that researchers care about - research integrity and data validity.

OBJECTIVE

In partnership with other campus units, develop training opportunities that enable graduate students and research staff to implement specific and effective data management strategies promoting research integrity and data validity.

TRAINING DESIGN & FORMAT

- Practical Data Management Workshop Series
 - ◊ Targeted to the field of study
 - ◊ Examples, worked & practice exercises
 - ◊ Real-world data sets
- Tutorials & resources embedded into graduate research courses
- Stand-alone tutorials
- Data management subject guide

REFERENCES

1. Steneck, N. H., & United States. (2004). ORI introduction to the responsible conduct of research. Rockville, Md: U.S. Dept. of Health and Human Services, Office of Research Integrity.
2. UK Data Archive. (2009). Managing and sharing data: A best practice guide for researchers. Colchester: UK Data Archive.
3. Inter-university Consortium for Political and Social Research (ICPSR). (2012). Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (5th ed.). Ann Arbor, MI.
4. Society for Clinical Data Management. (2011). Good Clinical Data Management Practices. Washington, DC.

MISSION ALIGNMENT

- IUPUI: "to advance...the intellectual growth of its citizens...through research and creative activity, teaching and learning, and civic engagement"
- University Library: "Inform. Connect. Transform."
- Campus & Program Accreditation
- Fill in gaps between existing requirements for RCR education and mentorship

"In general terms, responsible conduct in research is simply good citizenship applied to professional life."

Office of Research Integrity, 2007

PLAN

- Recognize benefits of data management planning in ensuring data validity and enabling data preservation, sharing, and re-use
- Develop clear description of the data to be generated and produced during the study
- Define data validation and analytical processes
- Define expected outcomes for data
- Identify and develop key project documents
- Identify stakeholders in data management and scholarly products
- Define roles and responsibilities for data management activities
- Meet funding agency and publisher requirements for data management planning and practice

DISCOVER & INTEGRATE

- Identify level of data sensitivity and legal and ethical obligations
- As a data producer, understand the considerations for sharing data
- As a data producer, choose an appropriate level of access given ethical and legal obligations
- As a data consumer, understand restrictions and permissions for re-use, re-distribution, and derivatives
- Enable proper attribution through citation mechanisms
 - ◊ Ex: Assign URIs or DOIs to the data, at the appropriate level of specificity for consumption
- As a data consumer, document data source for attribution As a data producer, provide sufficient context for discovery, integration, and re-use
- Document the process for integrating data sets
 - ◊ Ex: Map data sets and create a final data model

PRESERVE

- Recognize the need to preserve data as a vital part of the scholarly record
- Recognize the dependence of preservation on good data management throughout the data life cycle
- Ensure the availability of data for preservation through robust storage and backup processes
 - ◊ Ex: Maintain three copies of all files
 - ◊ Ex: Use checksums to verify file integrity
- Implement strategies to facilitate long-term access and preservation of files
 - ◊ Ex: Choose open standard, common file formats
 - ◊ Ex: Use metadata and documentation to provide a digital provenance trail
 - ◊ Ex: Implement appropriate versioning tools or strategies

COLLECT & ASSURE

- Implement study procedures outlined in the protocol and data management plan to facilitate data integrity through accurate, consistent, and logical data collection
- Consider provenance issues in data assurance
- Develop tools and processes supporting quality assurance (QA), quality control (QC), and continuous improvement (CI)
- Implement practical QA, QC, CI strategies that minimize opportunity for human error
 - ◊ Ex: Develop standardized collection tools
 - ◊ Ex: Collect raw data (i.e., calculate values later)
 - ◊ Ex: Identify missing values with specific codes
 - ◊ Ex: Maintain consistent data typing
 - ◊ Ex: Use robust data entry procedures (i.e., blind double entry, automatic validation, field limits, etc.)

DESCRIBE

- Understand the need for a detailed, accurate, and organized description of the study and resulting data
- Understand the function of documentation and metadata and the need for standards in both
- Implement the use of standards in both study and data description products
 - ◊ Ex: Choose discipline-specific documentation formats (i.e., readme file, codebook, procedures manual, lab notebook)
- Develop appropriate documentation to ensure data integrity and facilitate analysis and write-up
 - ◊ Ex: Study protocol
 - ◊ Ex: Documentation detailing validation and cleaning procedures
- Create appropriate metadata to facilitate sharing, discovery, and re-use
 - ◊ Ex: Study metadata to enable discovery
 - ◊ Ex: Metadata on specific data points to enable data integration and re-use

