Promoting Open Science Throughout the Research Lifecycle: The Integrative Role of Libraries

Norm Medeiros

Associate Librarian for Collection Management & Metadata Services
Haverford College

Co-Director, Project TIER

16 June 2017

Project TIER is generously supported by the Alfred P. Sloan Foundation
What to expect

• Opportunities in academic libraries and technical services, especially research data management and digital scholarship

• An introduction to open science and prospective library services integration therein
About Me

- Why librarianship?
About Me

- University of Rhode Island ‘95

- Exciting times: online catalog; World Wide Web; improving economy
About Me

• Very traditional opening act

• Followed by “Let’s catalog the Web!”

• Last 16+ years at Haverford trying to provide valuable, cost-effective, and sustainable collection & technical services with a lean staff and modest budget.

• ALCTS: Association for Library Collections & Technical Services
About Me
Expanding Role of Academic Libraries & Technical Services

• What is technical services today? What is cataloging? What is acquisitions?

• Fewer positions, fewer MLIS students, greater inter-divisional cooperation

• More forward-facing work

• More immediate service to users

• From collections- to user-orientation
Research Data Management


• Strong utilization of metadata & organizational skills

• Consultative role on NSF, NIH, grant funding requirements
Research Data Management

Learn more about data management in this research guide. If you have observational, experimental, simulation, derived or compiled data, we can help:

- Introduce your research team or class to the Open Science Framework (OSF), to facilitate management of workflows, collaboration, version control, and data archiving.
- Assist you in developing and writing your data management plan and make sure you have included types of data, standards, policies for access and sharing, archiving and preservation plans.
- Teach a workshop so your research team or students know how to organize and manage data;
- Connect you with other VCU groups that can help with the storage, security and analysis of your data;
- Identify data repositories to preserve and/or share your data, such as the new VCU institutional repository, VCU Scholars Compass;
- Suggest metadata and documentation to make your data searchable and assure you are credited.

More help

Contact our Research Data Management Team.
Research Data Management

Data Management Services

Data Management Services assists Stanford researchers with the organization, management, and curation of research data to enhance its preservation and access now and into the future.

If you have any questions you can email us any time at ask-data-services@lists.stanford.edu.

A permanent home for your data
Easy access for everyone else

Preserve your research data in the Stanford Digital Repository and give yourself peace of mind, knowing your data are safe. And the persistent URL you’ll receive will ensure your data always have a home.

> Find out more about sharing and preserving your data

Data management plans

We let you know what’s required, what to include, where to find the information you need, and how to get it done fast with the DMPTool.

> Find out more about the Data Management Planning Tool
Research Data Management Services

Contact the Libraries' Research Data Management (RDM) Team for help with managing your research data. Email rdm-mgmt@lib.psu.edu.

We are available for the following:

- Consultations
- Workshops / Information sessions
- DMP assistance
- Guidance on tools and repositories
- And more!

Guides:

- RDM Services guide
- Guidance on funding agency public access policies
- Data management planning toolkit.

The RDM Team

- Robert Olendorf, Science Data Librarian in the Physical and Mathematical Sciences Library
- Nathan Plekulek, Geospatial Services Librarian in the Social Sciences Library.

More Penn State Resources

- Penn State DMP Guidance
- Penn State Data Management Plan Tutorial
- ScholarSphere
- DataCommons

Other Tools, Resources, and Services

- DMP Tool Online
- re3data.org - Registry of Research Data Repositories
- All-purpose repository services
  - figshare
  - Zenodo
- GitHub - Git repository on the Web, used typically for software development
Research Data Management

RESEARCH DATA MANAGEMENT SERVICE GROUP
Comprehensive Data Management Planning & Services

The Research Data Management Service Group (RDMFG) is a collaborative, campus-wide organization that assists with creating and implementing data management plans, applying best practices for managing data, and finding data management services at any stage of the research process.

Email Us  Office Hours

RECENT NEWS
Amazon Web Services Training Session
Thu, Apr 06, 2017
Amazon Web Services will be hosting an on campus training session Research Immersion Day on April 27, 2017.

SERVICE SPOTLIGHT
eCommons for sharing and preserving data
You can use eCommons, Cornell's digital repository, to make your data available

CORNELL SERVICES
- Collaboration tools
- Data collection and analysis
- Data sharing
- High performance computing
- Intellectual property and copyright
- Metadata
- Security, privacy, and confidentiality

Ex Libris Bluegrass Users Group Annual Conference, Bowling Green, KY, 16 June 2017
Research Data Management

The Research Data Management team is ready to consult on how to collect, document, organize, store, and preserve your research data throughout your research project. We can help you to understand and meet funder requirements, including assistance in drafting and reviewing data management plans, selecting appropriate active data storage, choosing a repository, and preparing research data for end-of-project archiving. Come by or email us to talk about best practices to insure your research data will continue to be available when you and others need it in the future.

To learn more, check our Research Data Management Guide with the most current information available, including requirements for NSF, NIH, NEH, DOT, IES, DOE, and NOAA.

Research Data Life Cycle

Data Discovery → Re-Use → Data Archive → Deposit

Proposal Planning Writing → Project Start Up → Data Collection → Data Analysis → Data Sharing → End of Project
Effective data management can increase the pace of the research process, contribute to the soundness of research results, and meet funding agency requirements by making research data easy to share.

Supporting data management at every point in your research
Digital Scholarship

• A thousand points of light

• High profile

• Active role in generation of scholarship

• Working without a net

• ALCTS program, “Creating the Future of Digital Scholarship Together: Collaboration from Within Your Library” (Saturday, 6/24, 8:30-10:00am, MCP, W185a)
Digital Scholarship Center

3D Printing and Laser Cutting Now Available!

Get the Digital Scholarship Center Newsletter

About the DSC, DSC Events No Comments

New Innovative Teaching with Makerspace Technologies Grant

Deadline: May 1, 2017

About the DSC, Services No Comments

Makeover Monday with Tableau Software

HOUSE of DATA

Live Stream of QUEER ENCODING: DIVERSE IDENTITIES

Measuring Similarity Between Texts in Python

By Luling Huang

DS Grad Students, DS Methods, DS Projects No Comments
About Digital Scholarship

Digital Scholarship can be defined in many ways. At the University of Washington, Digital Scholarship has been defined as:

“Digital Scholarship” is defined as any scholarly activity that makes extensive use of one or more of the new possibilities for teaching and research opened up by the unique affordances of digital media. These include, but are not limited to, new forms of collaboration, new forms of publication, and new methods for visualizing and analyzing data. - Demystifying the Digital Humanities, University of Washington

Digital scholarship is often composed of works that are born digital, multimedia, database technology-based, analysis of other born digital material, digital text and images, digital music or art, and data sets. Much of this scholarship is never intended to be formally published. This form of scholarly data, presentations and dissemination represents a shift away from publishing and the kind of scholarship that we have traditionally collected and preserved in libraries, and is a natural evolution and adaptation of digital technology to scholarship. - UW Libraries

The goals of this website are to:

• raise awareness of digital scholarship happening at the University of Washington
• connect University of Washington faculty, staff, and students with the assistance needed at any point during a digital scholarship project

Have questions about digital scholarship? Ask us!
Digital Scholarship

Welcome

The Center for Digital Scholarship, a cross-departmental group in the Brown University Library, supports digital scholarship for the Brown community and beyond by supporting scholarly and academic activities that are conducted or enhanced through the use of digital technology, or that engage with its effects.

Featured Projects

Risorgimento
U.S. Epigraphy Project
Mapping Colonial Americas Publishing
Brown Digital Repository

How Can I Work With CDS?

Get help with a data management plan
Interact with a large-scale, high resolution display
Visualize data

Data Curation Visualization

Ex Libris Bluegrass Users Group Annual Conference, Bowling Green, KY, 16 June 2017
Digital Scholarship

Haverford Libraries supports digital projects that complement traditional scholarly research, while opening new opportunities for the creation of knowledge.

We support the teaching of digital scholarship in the classroom, including support for assignments, course design and digital pedagogy.

Use the filters below to explore current projects, student work and opportunities:

- All
- Faculty Projects
- Quaker & Special Collections
- Student Work and Opportunities
- Course Support
- People

TICHA
A DIGITAL TEXT EXPLORER FOR COLONIAL ZAPOTEC
#OpenScience

- Transparency of method, data, and analysis to facilitate verification, replication, and extension of scientific findings

- Relationship of Open Science to Open Access & Open Source

“Making reproducibility of your work by peers a realistic possibility sends a strong signal of quality, trustworthiness, and transparency.”

What problem is open science trying to solve?

→ Credibility crisis in science
Could bogus scientific results be considered false advertising?

Could a scientific paper ever be considered an advertisement?

That was the question posed to a Tokyo court, in a criminal case where prosecutors argued — at the behest of Japan's ministry of health — that a peer-reviewed paper containing faked data should be considered "fraudulent or exaggerated advertising" under that country's laws.

In that case, however, the argument didn't work. In March, the court decided that a fraudulent paper was not false advertising, allowing a pharmaceutical researcher at a Novartis subsidiary to escape jail time. The decision also cleared Novartis of charges and helped the company avoid a ¥4 million ($35,373 USD) fine.

According to an article in The Japan Times, the presiding judge Yasuo Tsujikawa said:

Read the rest of this entry »

Gender-based violence researcher now up to 10 retractions for plagiarism

A publisher has retracted all of the papers it published by a researcher in Nigeria, citing plagiarism.

The papers, all about terrorism and gender-based violence, were written by Oluwasen Bamidele. The journal editors and the publisher, Taylor & Francis, decided to retract nine papers by Bamidele because of the overlap to other works — which he also failed to reference.
Reproducibility Project: Psychology


- 39% found to replicate

“A large portion of replications produced weaker evidence for the original findings despite using materials provided by the original authors, review in advance for methodological fidelity, and high statistical power to detect the original effect sizes.”

Reproducibility Project: Cancer Biology

• 50 articles published in *Nature*, *Science*, and *Cell* published between 2010 and 2012

• Early returns: 2 replicated, 1 failed, 2 inconclusive
Social Sciences Replication Project

- 21 experimental studies published in *Nature* and *Science* in 2010-2015; have clear hypotheses, statistically significant results, and convenient samples (frequently students)

- Replication criteria: Effect in the same direction and p-value < 0.05

- To be completed in September 2017

- Replication reports will be posted on the SSRP website: [www.socialsciencesreplicationproject.com](http://www.socialsciencesreplicationproject.com)
Is Economics Research Replicable?

• 59 articles from 13 leading journals

• same data and code (!)

• 49% found to replicate, even after author assistance; worse rate (33%) without author assistance

Reproducibility vs. Replication

• Reproducibility
  Verification that identical inputs (data & method of analysis) yield identical results; commonly referred to as “computational reproducibility”

• Replication
  Testing effect using same methodology on a different sample
A Frequent Scenario

“everything you need to know is in the article”
Scientific Value vs. Scientific Reward
(truth seeking)  (funding, tenure)

➔ Incentive to publish is powerful

• Drive funding, promotion, acclaim

• The need for significant results ($p < .05$)

• Most prestigious journals will not publish insignificant results
Threats to reproducible science

P-hacking

→ **Manipulating data to achieve a positive effect**

- Confirmatory (Hypothesis testing) vs. Exploratory (HARKing; data dredging)
- Results in false positives (type I error)
- Contributes to “file drawer” problem
- *Researcher degrees of freedom*: Dropping observations/variables; cutting short data collection; testing different analytical techniques
Let’s learn more about P-hacking from John Oliver
(Pre-)Registration/Pre-Analysis

- Making available hypotheses, study design, materials, and methodological plan before data collection begins
- Decreases HARKing and dredging for significance
- Also decreases file drawer problem

“The strongest form of pre-registration involves both registering the study … and closely pre-specifying the study design, primary outcome and analysis plan in advance of conducting the study or knowing the outcomes of the research.”

— A Manifesto for Reproducible Science
A New Publishing Paradigm?

- Publication agreements based on hypothesis and planned analysis, not results -- similar to grant awards
- Removes motivation to manipulate significance tests
- Transparency and Openness Promotion (TOP) Guidelines (over 3,000 journals & organizations)

“Transparency, open sharing, and reproducibility are core values of science, but not always part of daily practice. Journals, funders, and scholarly societies can increase reproducibility of research by adopting the Transparency and Openness Promotion (TOP) Guidelines and helping them evolve to meet the needs of researchers and publishers while pursuing the most transparent practices.”
Pre-Registration Challenge

• $1,000 to 1,000 researchers who agree to submit a registration on OSF and subsequently publish their findings in a participating journal (780 registered thus far)

• Sponsored by Center for Open Science

• A registration on OSF
• **Mission:** increase openness, integrity, and reproducibility of research

• **Product:** OSF (Open Science Framework) Platform for pre-registration/pre-analysis, data management, collaboration, and publication
Project TIER: Teaching Integrity in Empirical Research

• Promote a systemic change in professional norms related to transparency and reproducibility of empirical research in the social sciences

• Providing comprehensive replication documentation for research involving statistical data should be as ubiquitous and routine as providing a list of references

• Authors should view this documentation as an essential component of how they communicate their research to other scholars

• Readers should not consider a study to be credible unless such documentation is available
Project TIER: Teaching Integrity in Empirical Research

WHAT'S NEW

Applications are being accepted for the next faculty development workshop to be held March 31-April 1, 2017 at Haverford College. Information and registration are available here.

Project TIER: Teaching Integrity in Empirical Research

TIER Protocol Documentation

- Original Data
  - Original data files
  - Importable data files (if necessary)
  - Metadata
    - The Metadata Guide
    - Supplementary metadata documents (if necessary)
- Documents
  - The final paper
  - The Data Appendix
  - The Read Me file
- Analysis Data
  - Analysis data files
- Command Files
  - Command files
Project TIER: Teaching Integrity in Empirical Research
Project TIER: Teaching Integrity in Empirical Research

"Replication Documentation"
Project TIER: Teaching Integrity in Empirical Research

- Replication Documentation for Midlife Crisis Paper
- Analysis Data
  - country-analysis.dta
  - individual-analysis.dta
- Command Files
  - 1-import-wdi.do
  - 2-import-pew.do
  - 3-processing.do
  - 4-data-appendix.do
  - 5-analysis.do
- Documents
  - 1-ReadMe.pdf
  - 2-Midlife-Crisis-paper.pdf
  - 3-data-appendix.pdf
- Original Data
  - importable-pew.dta
  - original-pew.sav
  - original-wdi.xlsx
Project TIER: Teaching Integrity in Empirical Research

TIER Protocol 3.0: Template

Contributors: Norm Medeiros, Richard Ball

Forked from osf.io/7g6cn on 2016-09-16 12:43 PM

Date created: 2016-09-08 11:17 PM | Last Updated: 2016-10-14 10:37 AM

Create DOI / ARK

Category: Project

Description:
This project is designed to support Haverford College economics majors who produce empirical theses. The structure is based on the TIER Documentation Protocol. Additional information about Project TIER is available at http://projecttier.org

License: No license

Wiki

This template was constructed for use by individuals who wish to follow the TIER Protocol for conducting and documenting an empirical research project.

Citation

osf.io/ybzxe

Components

Add Component

Link Projects
Project TIER: Teaching Integrity in Empirical Research
These Workshops will introduce participants to the TIER protocol for replicable empirical research. They are intended for faculty members interested in teaching their own students to follow this protocol to document the statistical work they do for senior theses, other independent research projects, or papers written for classes.

The Workshops are held on the campus of Haverford College, and last one and one-half days.

**Spring 2017 Workshop**

The next Faculty Development Workshop will take place at Haverford College on March 31-April 1, 2017.
Technical Data Skills for Reproducible Research

Harrison Dekker and Paula Lackie

The fact that raw data is rarely usable for analysis without significant work is a point I try hard to make with my students. I told them “do not underestimate the difficulty of data preparation.” When they turned in their projects, many of them reported that they had underestimated the difficulty of data preparation.
Promoting Open Science Throughout the Research Lifecycle: The Integrative Role of Libraries

Norm Medeiros
Associate Librarian for Collection Management & Metadata Services
Haverford College
Co-Director, Project TIER

16 June 2017

Project TIER is generously supported by the Alfred P. Sloan Foundation

Ex Libris Bluegrass Users Group Annual Conference, Bowling Green, KY, 16 June 2017