

# **Beyond Research Data Management: Emerging Trends in Library Support for Computational Research**

CNI Spring 2017 Annual Membership Meeting #cni17s

Harrison Dekker  
University of Rhode Island

# **Data services accomplishments**

- Data archives
- Metadata standards
- Data literacy
- Research data management
- Community building

**What else can we do?**

**More attention to code and  
computation**

“An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generate the figures.”

[Buckheit and Donoho \(1995\)](#)

# Examples of support/training needs

- Data(text or numeric) - cleaning/visualization/analysis
- Database design and SQL
- File management and command line operations
- Version control
- Literate programming
- Computing environments - cloud/virtual machines/containers
- Open source code development and maintenance

# Higher level support

- Research workflows
- Computational thinking

# Strategies

- Embedded role in a course is best
- Dedicated, staffed space
- Peer support
- Workshops and other “ad hoc” instruction events

**Get involved with the research  
transparency movement**



Scientific claims should be subject to scrutiny by other researchers and the public at large. An important requirement for such scrutiny is that researchers make their claims transparent in a way that other researchers are able to use easily available resources to form a complete understanding of the methods that were used by the original

[Christensen \(2016\)](#)

# Transparency concepts

- Pre-analysis plans/Pre-specification
- P-hacking
- Publication bias
- Registration
- Replication
- Reproducibility

More info:

[Berkeley Initiative for Transparency in the Social Sciences](#), [Open Science Foundation](#), [Project Tier](#)

# **Case study**

# **ECON 196 - Practical Programming for Economics**

## **Course Objectives**

- Increase proficiency with the Stata language
- Introduce a workflow-based approach to data analysis
- Introduce emerging open source data analysis tools
- Develop teamwork and presentation skills

# Topics covered

- Introduction: Programming, reproducibility, and research integrity
- Planning, organizing, and documenting: Files and variables
- Planning, organizing, and documenting: Workflow automation basics
- Command line operations and version control with git
- Data transformations: merge, reshape, collapse
- Data cleaning and string functions
- Analysis and presentation
- Introduction to SQL
- Introduction to R

# Teaching approach

- Live coding
- In-class coding challenges
- Group and individual projects
- Open data
- Project TIER

# Conclusions

# Impetuses for change

- Undergraduates seeking to diversify skill set
- Underrepresented groups who may have not had prior exposure
- Transitioning graduate students (undergrad → grad)
- Graduate students and postdocs transitioning to new areas
- Graduate students and postdocs seeking to acquire sufficient proficiency to correctly use new methods in their research
- Technology evolution
  - Easy, cheaper access to computing power
  - Availability of expressive, extensible, open source languages like R and Python



When you consider some of the key values that drive librarianship... it becomes evident that librarians *must* take a leading role in working with big data lest this emerging specialty become the servant only of proprietary interests. Librarians stand for open access to information, for privacy rights, for serving the information needs of the community...

[Stanton \(2012\)](#)

All of this hesitancy on the part of librarians to participate in the data movement is happening at a time when we have seen an increase in the money and involvement in data initiatives from a range of other professions and academic disciplines (e.g. computer science, informatics, etc.). For me, this is an especially critical moment for librarians to talk about data and actively plan and implement our strategies collectively.

[Martin \(2015\)](#)

# Sources

Buckheit, Jonathan B. and Donoho, David L.. 1995. "WaveLab and Reproducible Research". Technical report, Stanford University.

[http://statweb.stanford.edu/~wavelab/Wavelab\\_850/wavelab.pdf](http://statweb.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf)

Christensen, Garret. 2015. "Manual of Best Practises in Social Science Research". Unpublished manuscript. <https://github.com/garretchristensen/BestPracticesManual>

Martin, Elaine R.. 2015. "The Role of Librarians in Data Science: A Call to Action." Journal of eScience Librarianship 4(2): e1092. <http://dx.doi.org/10.7191/jeslib.2015.1092>

Stanton, Jeffrey M.. 2012. "Data Science: What's in it for the New Librarian?". Blog post, Syracuse University. <https://ischool.syr.edu/infospace/2012/07/16/data-science-whats-in-it-for-the-new-librarian/>