

AAHSL Data Management Training Scholarship Report

I chose to attend a two-day Data Carpentry workshop offered as part of the Research Data Access and Preservation (RDAP) summit in Atlanta, GA in May, 2016. University data services need to be ready to help researchers comply with the public access requirements for data that supports published papers, which includes cleaning up data and providing scripts for any analysis done on the data. Over half (53.14%) of our sponsored program awards come from the federal government, so the 2013 Office of Science and Technology Policy (OSTP) memo, which requires public access to data and data management plans for NIH, FDA, and other federal agencies, will have a big impact on our researchers. At the same time, the NIH is pushing for transparency in research and researchers must include information on reproducibility in their grant narrative. NIH funding accounts for 31% of our total sponsored programs awards, making this the top agency for funding, so it is very important for me to help be researchers comply with award requirements surrounding reproducibility and public access to and sharing of data. I requested this award to update and expand my data skills, including hands-on practice with advanced data cleaning, preparation, sharing, and analysis skills.

Data Carpentry (<http://www.datacarpentry.org/>) trains researchers in the core data skills for efficient, shareable, and reproducible research practices. Workshops are based on the Software Carpentry strategy of collaborative development of hands-on, interactive lessons, that are domain-specific so participants learn the skills most relevant to their domain. The workshop I attended used an ecology dataset and went through lessons in:

- Spreadsheet Best Practices
- Open Refine
- Introduction to R
- Introduction to SQL
- Analysis and Visualization in R

Our group was lucky because as well as the two official Data Carpentry instructors, a couple of knowledgeable colleagues agreed to assist, so there was lots of help throughout the workshop. Having time free from work interruptions to devote to practice with Open Refine, R, and SQL was very valuable. I now have a better understanding of how researchers can clean up the spreadsheet data, run statistics on that data, and visualize the results. I also have some good information on best practices for spreadsheets and setting up SQL databases, that will be

helpful when I'm asked to consult about data collection. And importantly, I learning about scripting in R, which allows for the saving of a series of commands used for analysis. This is a crucial part of showing reproducibility in research.

As a data librarian who needs to be able to go back and train researchers and students, the free lesson materials in multiple domains are an excellent resource and make planning workshops much easier. Not only do I have ready-made data sets, but I have some idea of how long each lesson will take so I can plan properly.

I am very grateful to AAHSL for this scholarship and feel it has improved my skills, making it easier to consult and teach in the necessary areas of reproducibility and data analysis.

Respectfully submitted,

Margaret Henderson, MLIS, AHIP
Associate Professor
Director, Research Data Management
VCU Libraries
Virginia Commonwealth University