10-30-2018

Data Curation Network Hosts First Specialized Workshop

Emily Berkowitz
University of Pennsylvania

Recommended Citation

Berkowitz, Emily, "Data Curation Network Hosts First Specialized Workshop" (2018). *ADRF Network Newsletter Interviews*. 6.
https://repository.upenn.edu/admindata_interviews/6

Data Curation Network Hosts First Specialized Workshop

Creative Commons License

October 30, 2018
By Emily Berkowitz

The Data Curation Network held its first Specialized Data Curation Workshop on October 17-18, 2018 in Henderson, Nevada. The ADRF Network spoke with [Lisa Johnston](#) from the University of Minnesota, [Cynthia Hudson Vitale](#) from Penn State University, and [Wendy Kozlowski](#) from Cornell University about how the Data Curation Network got started and about their recent workshop. This was the first workshop in a series of three training sessions that aim to increase stakeholder understanding of and capacity for data curation practices. Monica King from the ADRF Network also attended the workshop.

What is the Data Curation Network?

"The Data Curation Network [brings] together the curators who work at data repositories into a shared staffing model where we can rely on the expertise of each other's data curators," Johnston explained. For example, a researcher may excel in the use of GIS data files but lack expertise in using statistical survey formats. As part of the DCN, curators can share their knowledge and grow from the expertise of others within the Network.

"We have created a community of people for the first time can talk to each other about this new-ish area of digital data curation that's happening in a lot of libraries," Johnston said.

What is data curation?

Data curation is the work and action of data curators taken to by curators of a data repository in order to provide meaningful and enduring access to data, according to the Data Curation Network.
"An easy way to think about data curation," Johnston explained, "is trying to be the first user of the data." A "first user" might seek to answer questions such as, "Why were these data created," "What methodologies were used," or, "What can this data be used for and what permissions do I have or need in order to use it?" Data curators evaluate research data to ensure they are accessible to and digestible for future users. The data are diverse and include tabular data from the social sciences, data imagery from physics and astronomy, and geospatial data.

What are the benefits of data curation?

Data curation enables stakeholders across a variety of disciplines to open, access, and understand unfamiliar data files. Without prior training or expertise in a certain discipline, it can be difficult to know how to work with certain datasets and file types. Data curators together can identify and communicate "the best way to work with these

files and ensure that the data themselves are more reusable," Johnston said. "The ultimate outcome of all of our efforts is that the data can be used by someone else." Careful data curation can also mitigate some risks of archiving data, Kozlowski told the ADRF Network. For example, a researcher may accidentally delete data or remove an important piece of the information, which can lead to the proliferation of data in poor quality. Data curation can also prevent researchers from exposing sensitive data that compromise privacy and safety. Johnston shared an example of a data set that provided the geo-location of endangered species which could put animals in acute danger. The development of best practices through shared expertise in the DCN can help promote benefits of data curation and mitigate the risks associated with making data more accessible.

What are some key takeaways from the Specialized Data Curation Workshop?

The DCN developed the Specialized Data Curation Workshops in order to bring the practice of data curation to people outside of the Data Curation Network. The workshop took place over a day and a half and included hands-on learning and networking. A group of over 20 participants from non-profits, academic institutions, and other organizations from the United States and Canada attended with the shared goal of developing best practices for data curation.

For most of the workshop, the participants broke into small groups to work with five different datasets—including tabular, geospatial, code, survey, and image data—to walk through the steps of data curation. The workshop also allowed the participants to start developing primers to offer best practices for addressing a data curation need at their local institutions. Hudson Vitale explained, "[primers] are meant to jumpstart the curation process or have someone look at this document and then understand how they might interact with the file or format or data that they're looking at." Examples of primer topics that emerged from the workshop include Jupyter, administrative data, and GIS files. The primers will be published in spring 2019 and will be available in a public repository.

Reflecting on the workshop, Johnston told the ADRF Network, "People brought their own experiences, and what we were able to do was give them a framework to put those experiences in the context of data curation best practices. That for me has been what's missing from this field."

If you're interested in learning more about the Data Curation Network or the workshops, contact Cynthia Hudson Vitale at cuv185@psu.edu.