

Jun 21st, 1:30 PM - 3:00 PM

## Disciplinary data inventory: introductory plenary session comments

Jacob Carlson  
*Purdue University*, [jakecar@umich.edu](mailto:jakecar@umich.edu)

Follow this and additional works at: <http://docs.lib.purdue.edu/iatul2010>

---

Jacob Carlson, "Disciplinary data inventory: introductory plenary session comments" (June 21, 2010). *International Association of Scientific and Technological University Libraries, 31st Annual Conference*. Paper 9.  
<http://docs.lib.purdue.edu/iatul2010/conf/day1/9>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

IATUL presentation

Thursday, June 10, 2010

4:31 PM

My name is Jake Carlson and I am a Data Research Scientist here at Purdue.

I am a part of the Purdue University Libraries research department. The central mission of the Research department is to explore how the knowledge, skills, and expertise of librarians could be applied further upstream in the research process. Traditionally librarians have focused on the outputs of research, the books and journal articles that are produced and published, but as technology and other factors are changing the nature of science and scholarship, we are seeing that other information resources being generated over the course of the research process, research data sets in particular, also need to be considered for inclusion in the scholarly communications process.

Given this new environment, we seek to understand researcher needs with regards to organizing, managing, describing, disseminating, curating or preserving their data and to form interdisciplinary partnerships with faculty to address these needs.

Data Research Scientist sounds like an unusual title, but I want to make it clear that I am a professional librarian with an MLS degree, and that it is from the librarian perspective that I approach my work. Although, I'm comfortable with technology, I do not possess a deep level of technical expertise or training. Nor do I have formal training in a STEM discipline.

As a librarian, however, I do possess a set of skills and perspectives that are important to curating research data. I understand the principles and practices of collection development, what is needed to create and administer a diverse collection of materials to support the research and teaching needs of faculty and students. I understand taxonomies and the organization and classification of information. I have experience in conducting reference interviews and to ask questions in a way that will uncover the true needs of patrons and then to address these needs through my knowledge of services and resources that are available to them.

I have been at Purdue since 2007, a little more than three years have passed. The biggest challenge that I have encountered in my time in this position is how to apply my skills as a librarian to a new, unfamiliar and largely uncharted environment of data curation. Although we strongly believe that librarians can contribute and have a role to play in data curation, there were few if any models to follow when I started at Purdue. I knew what I was supposed to do (talk with faculty to better understand their data and their needs, and then figure out how we could work with them to address these needs), but I did not know how to go about doing it.

Early on I spent a great deal of time talking with my colleague Michael Witt to try and learn from his experience in this area. Those conversations led to our developing a list of 10 questions that could serve as the broad foundation of a data interview. We turned these questions into a poster submission and presented our work at the 2007 Digital Curation Conference in Washington D.C. You have a copy of this poster as a handout for this session.

This poster was well received at the DCC, but what has surprised me is the reception it has received since then. This poster has been download more than 1,500 times from Purdue's institutional repository and the number of downloads has been fairly consistent over time. This tells me that Michael and I really identified a need in the library science field. That beyond just myself, we as a profession are seeking models and tools to help us move collectively forward in identifying and advancing roles for ourselves in working with data.

At about the same time that Michael and I presented our poster, a grant proposal that we had submitted with the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign to the Institute of Museum and Library Services was awarded. This grant centered on investigating "who is willing to share what data, with whom, when and under what conditions" and on developing the "data curation profile", a framework to capture and compare information about a researcher's data set and his/her needs for the data. We interviewed 19 faculty for this project and asked them a variety of questions that fell into three broad categories:

1. What data do you have? (characteristics, format, lifecycle, quantity, etc.)
2. What are you doing with this data currently? (immediate value, management and handling practices, etc.)
3. What would you like to be doing with this data, but are not currently for whatever reason? (better organization, better management, description, sharing, preservation, etc.)

However, even more important than the questions that we asked was the approach that we took. Our goal was to have researchers drive this process and to have them tell us what their needs and priorities were rather than trying to dictate what we thought they should be doing. Therefore, initially the questions were kept fairly general and lots of follow up questions were asked based upon researcher's responses.

I believe that this principle, approaching data curation through enabling researchers describe their needs and priorities to me, is the central tenet of my work. Once the needs have been established and a relationship has been built, I can then respond

with my ideas and perspectives on addressing their needs, but faculty needs have to drive data curation efforts.

And with that I will turn it over to the first of our speakers...