

Interview: Jake Carlson – Purdue University

[Note: As of August 2014, Jake is employed by the University of Michigan.]

1. Tell me about the product or service that you provide in the roles that you play at your library?

I think the purpose of my job – I come from a very traditional background so before I came to Purdue I was at Bucknell University and I was a subject librarian for the social sciences. In that capacity I delivered reference services, I taught information literacy-type courses and I built collections and resources to support the area of social sciences. I think when I was hired at Purdue really looking at what kind of roles or services or support can we offer to faculty working with data it was really focused on what can librarians do? How can we take the knowledge and perceptions and skill sets that we have and translate that into a newish environment of working with data.

My role then, has been to try and draw bridges between traditional areas of librarianship and this new role and responsibility for librarians. If I'm known at all for the work that I have done it is probably the Digital Curation Profiles toolkit, which essentially can be likened to a reference interview on steroids. It is really based on the reference skills and abilities that librarians have. So in traditional reference, someone approaches you needing to find X or gather information on Y – you give a back and forth with the questions you ask them and then steer them towards a particular resource that will satisfy their information need. With the Data Curation Profiles Toolkit it is really the same kind of scenario only it is more proactive reference – we talk to them about what they are currently doing, talking about a particular dataset that they are working with or generating or using somehow, you're learning the data lifecycle, you are learning how they are currently managing and working with that dataset and then as you're talking you're uncovering areas of need. One example is a researcher really wishing they can share their data with their collaborators, had means to publish my dataset through means that will show impact. Essentially while they aren't coming to librarians, by talking to researchers about their situation we are uncovering needs that then once distilled and summarized in the Profiles, you can act on these needs potentially. Therefore here you can find researchers resources that will help them or identify a particular repository that could hold their particular dataset or a metadata standard that can be applied.

If there is no resource there we have tried to take on the challenge through collaborations with faculty. And that sort of encompasses the translation from reference traditional to new reference.

With the data information literacy project that I am currently leading, we're looking at how can we take what librarians have learned about information literacy type courses and translate that into the new data environment. ACRL has the 5 info literacy standards – how relevant are those when applied to the activities surrounding data such as producing data. Our first article that we published about this portal back in April 2011, we found that the ACRL has this implicit assumption that you're talking with information consumers so people are coming to you when they want information. Or you're teaching people how they are going to use and evaluate data that they are going to consume somehow. With data, we are focusing more on producers of data, so how are they building the dataset? How are they working with the data for their own particular needs? Are they extracting data from that for their own particular research areas? There is that distinct difference, but a lot of the skills and abilities of finding data to then help shape or inform their idea or their own dataset certainly comes up as a skill that they graduate students potentially need.

As a result, we came up with 12 data competencies or area that we think graduate students need to possess skills or knowledge in order to be successful. This will help them do their research in the lab but also later on in their career on their own. We're looking at what are the skills that they need? Are the twelve that we found the right ones and are there particular facets of those skills that have more of a perceived need than others? From that knowledge how can we translate that into something that librarians can act on and help support. The 12 competencies are fairly wide and vast and it may mean that librarians aren't involved in all of those but that is certainly true for information literacy too; it is not just information literacy equals libraries but there are a wide variety of facets that could or potentially should be covered by other folks with particular skill sets. So it is not that libraries are not going to be sort of 'owning' this, but we would play a role in it as well as support in terms of getting people together to offer these sorts of things.

2. How can other librarians use this product or service?

3. How has your library reached out to your institutional community and how have you earned support for this particular service?

Our perception is that it is really important for people in this area to not go in and sell the service but to seek to understand and THEN be understood. We try and go out and approach this tabula rasa, and certainly we have our own assumption and we have our own potential places where we want to explore, but really being able to listen first and gather information through the profiles or through other means and then to respond. In this case this means not really looking at what is our understanding and how can we help, but more looking at what is the community's understanding and how does it relate to our understanding in terms a larger push for needs or support that can feed into their work. For example, the data information literacy project required an environmental scan of literature being produced in the discipline of faculty partners. For computer science, for example, we went out and did article reviews for software code; how do they document their code as well as organize and share it. That gave us a disciplinary understanding before going in and trying to offer an educational program. We then went in and did a series of interviews in the local lab with graduate students to get a sense of what the lab practice was. Once we were armed with this information, then we translated that into an educational program that really aligned with disciplinary understanding and local practice.

For most of these interviews we did have a preexisting relationship with faculty members but the context of that relationship was very different than what we tapped it for in this instance. We were familiar with these faculty members in a different context than what we did which was introducing something new in that we wanted to start a conversation about their data.

4. What skills or experience do you think health sciences librarians need to acquire to meet the needs of e-science and data management and can you provide examples of the skills and services that you or your other staff have in this particular area?

The skills that I have and that I bring to this area are really the traditional skills of a librarian. As I mentioned before I came from a more traditional library environment; I don't come from a science background so I don't have indepth knowledge of biology, chemistry or computer science – I have a social science background. I see the social science background that I bring as being valuable because I complete these interviews and I try to get a sense of the lay of the land to understand their needs, it is almost like an ethnographic approach – going to observe and talk to researchers in their environment about what they do to really inform the efforts and the tools and resources that we are working to develop (at Purdue). Really it is that kind of approach that has been very helpful for me in this role.

I also don't have a depth of tech knowledge – I'm not afraid of tech and I like to learn a lot of the new technology that is emerging but I don't come with that strong of a background. I think both of the background and tech skills are very helpful, but I wouldn't say that they are necessarily an automatic requirement. I still will have had an impact at Purdue with the work that I've done without having indepth knowledge of science or technology. Frankly I personally find a little bit surprising. I wasn't really sure how this was going to work when I first started working at Purdue but I think the skills, perceptions and abilities that we have as a librarian really do feed into this larger environment and can be leveraged if applied appropriately to lead to greater, bigger and better things like data.

In response to the first part of your question, having a good knowledge of the health sciences is great, having good knowledge of technology is great, but really it is not just skills but an attitude that allows you to put yourself and your perceptions aside to meet the needs of the user. That really is a fundamental tenet of libraries and what we do. In a lot of ways we are the catalyst; we aren't producing the data, but we are connecting the data producer with resources or tools or something that will help address that need. Serving that catalyst role of connecting people with the information or resources they need to address their issues is still necessary and valid in this data-driven environment.

I was hired directly to look at the socio-cultural aspects of data management and data curation at the institution. In terms of other staff members, we do have somebody who has a much stronger tech background than I do who was at Purdue originally and he is in fact working and applying his skills in a number of different ways. He has created something called Databib which is a registry of data repositories – Michael Witt. And he is also the staff member working on our institutional repository called PURR which is built on software called Hub-Zero that is used to create a virtual research environment where researchers can communicate and share resources and run tools and simulations and other things. It really acts as a way to bring people together irrespective of time and distance. What it didn't do is manage data or content effectively. There is a lot of content that floats around in Hub-Zero, but it is not managed nearly to the extent that we as librarians would hope to see. Michael Witt is working now with the skills that he has to get in there with the support of developers that report to him and add data management and data curation capabilities to augment things that the Hub does. This will allow for the software to maintain the collaboration tools that are there and involve that in the repository that we're building and also add the functionality that we would normally associate with repositories.



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).