Michael C. Witt

I am interested in seeing the principles of library science applied to research data curation.

Michael C. Witt is an Associate Professor of Library Science at Purdue University in USA. He leads the Distributed Data Curation Center (D2C2), which is situated within the Research Data unit of the library. Witt is also the library’s liaison to the Department of Computer Science and is a practicing librarian. He holds a Master of Library Science degree from Indiana University-Indianapolis.

Why are you involved in the theme DATA?

The DATA theme is taking an integrated, interdisciplinary approach that includes the University Library as a partner and incorporates library science as a fiber that could potentially be woven into all of its five threads. I’ve been engaged in projects that have taken similar approaches at institutional, national, and international levels, and it was an honor to be invited to join this exceptional group of colleagues at the Pufendorf Institute for this effort.

What do you hope to contribute?

I hope to contribute my practical experience in developing data services at the Purdue University Libraries that includes assessing researcher needs related to data, navigating policies, collaborating across units of the organization, and defining, resourcing, and implementing solutions for research data management. Every situation is unique, but we all face many of the same, significant social and technical barriers to advancing data management and sharing.

What do you hope to get out of your stay?

I’ve tried to come to Lund without pre-conceived notions or a specific agenda. It’s clear this campus is a place for creativity and connecting with new people and ideas. It has been less than a day since I arrived, and by the time we finished our afternoon coffee, we had already brainstormed two new possible research collaborations that fit with our theme.
What are your research interests?

I am interested in seeing the principles of library science applied to research data curation. Planning and building purposeful collections of data. Organizing and describing datasets so that they can be found and understood by people in the future. Preserving and stewarding data for the long-term. Helping researchers, students, and citizen scientists find and properly reuse datasets by provisioning new tools, reference services, and data literacy.

These are all things that librarians have done for centuries with books, and if you look at the Internet and the deluge of data we’re facing in science, these skills are needed now more than ever. How can libraries contribute to new solutions for data curation and better support the research lifecycle, taking into account increased computation and new forms of scholarly communication?

What drives you?

Think about the very first computer you owned and the programs you used on it. Can you still access all of the files you created on it? Chances are, you’ve replaced that computer many times over the years, and today you are using a different operating system and set of programs. You’ve probably experienced at least one hard drive crash or a computer virus, or you’ve simply lost files in migration or have accidentally deleted or forgotten about them.

We are facing the same problem with the data that make up the scholarly record: they’re all digital now. In a sense, we are in a race to avoid what happened in the Dark Ages, when tremendous amount of recorded knowledge were lost or destroyed. Instead of fire or war, we are losing data today because of bit rot, format obsolescence, and benign neglect. We can’t let that happen.

What are you working on right now?

One project that I’m working on is re3data, building a global registry of research data repositories. We have librarians and information professionals from around the world and from different disciplines who are identifying and cataloging places where researchers can deposit their data or find others’ data to reuse. Two of our goals for the future are to improve the representation of data repositories from developing countries and smaller research domains and to incorporate the registry into a machine-actionable “data fabric” so that information about data repositories can be used and extended, not only by users, but by user-agents such as software tools and other cyberinfrastructure.

Have you ever been to Lund and/or Sweden?

This is my first time in Sweden. My family is joining me when I return in the spring, and on a personal note, I am looking forward to my children experiencing and learning about Swedish culture to broaden their view of the world.

Watch a video with researcher Michael C. Witt.