Blog post: https://www.esipfed.org/esip-interviews/making-data-matter-with-denise-hills

Blog Title: Making Data Matter with Denise Hills

Interviewee: Denise Hills, Alabama State Geological Survey
Interviewer: Arika Virapongse
Date of interview: July 18, 2018

Blog highlight: "Data that we might not think are connected, but as an aggregate might help us understand more about the world around us by breaking down silos across disciplines."

Could you tell me about when and how you got started working in the field of data and informatics, particularly as it pertains to Earth Science?

Denise: I have had a long winding path between geophysics and data. I am by training a geoscientist, specifically geophysics, so I have been dealing with a lot of data and things that you would call informatics without knowing it throughout my career.

I started working with geophysical data around 1995. Then, I was out of pure research for a while, and worked as an informal science educator for a few years. Around 2006, I started working at the Alabama State Geological Survey and started using data again. In 2011, I got involved with the National Geothermal Database System through the American Association of State Geologists. This project was my first experience really working with data and informatics issues. I also became introduced to people who thought about these kinds of things.

There aren't really data managers where I work now (Alabama State Geological Survey), or anyone who specializes in this field. The only metadata that we were generating was from ArcGIS, because it forces you to do that. I wasn't in a culture to think about data. My work with the National Geothermal Database System is what really clarified for me how important it was that we did better with these things--not just for myself and my colleagues at work, but for the agency as a whole and beyond.

How did you go about gaining that learning curve?

Through ESIP! If I hadn't been a scientist, I might have been a librarian. I love libraries--their organization and structure, the information finding, and all of that. I am mostly self-taught. But what has helped me a lot are resources like the Data Management Clearing House at ESIP, and the people at ESIP. Like, Ruth Duerr is an amazing mentor and will point you in the right direction. Peter Fox is also an amazing mentor who sees the potential in people and helps them get to where they need to go. Lee Allison is someone who I like to call a Geoinformatics activist because he really helped people make connections to other people that they needed to know,

as well as make connections between people and resources & institutions to enable the work that they do.

How did you find ESIP?

I worked pretty closely with folks at the Arizona Geological Survey on the National Geothermal Database System. One of the big components of that was training courses. All of the 50 states were part of that project, and Lee Allison encouraged everyone to participate in-person or remotely at an ESIP meeting in the summer of 2012. That was my first introduction to people who actually do this for a living. I realized that they could help me solve problems at my agency. But, ESIP was really my entry into this. I feel like I've been here (at ESIP) for a long time, but it hasn't really been that long.

Thinking about your career trajectory, could you talk about the political, scientific, and technology context that pushed you towards the field of Earth Science data / informatics or made Earth Science data / informatics more important in society overall?

I work for a state agency in a state where they are trying to streamline government--trying to reduce budgets and be fiscally responsible. So we have to do more with less. A lot of my work focuses on increasing the value and use of data that we already have. Several projects that I have worked on were funded by federal agencies, and these were enabled by the fact that we have been able to generate metadata and records of that data. I work with a lot of geologic cores. It can cost a million plus to drill a geologic core, but if we already have it and know where it is (i.e., have the metadata for it), then the only additional cost is personnel time. It is a way for us to be able to continue doing important science at a much lower cost.

This approach makes us more competitive for funding. But in addition to that, our primary mission is to provide the information necessary to properly use and protect natural resources in our state. Without generating information that people can use in whatever manner they need, we can't fulfill our mission. Disaster recovery, business development, and energy development are all sectors that are impacted by the work we do. It's not a one-to-one relationship to the funding itself. It's a recognition of the importance of science to people who aren't necessarily scientists. That includes decision-makers, including the general public.

Since you began working in the field of Earth Science data and informatics, can you think of some major milestones that have shifted the field?

There are a lot:
- The cost of data storage. When I was working on my Master's thesis, it was very expensive for me to get a 100 MB zip disk. Now I have a flash drive that has 16 GB.
- Cloud computing
- Parallel processing. When I first got started, it was neither common nor easy to do parallel processing.

- Increased processing speed
- Some recognition that data should be interoperable
- The idea of open data
- The idea of FAIR (Findable, Accessible, Interoperable, and Reproducible) data--what that actually looks like and how that might best support science, scientific data, and scientists

What are some of the major challenges and issues that are facing Earth Science data / informatics today?

There are many! But I don't think that they are insurmountable. They are challenges, but they are also opportunities. A lot of these are cultural and societal challenges:
- Changing people's perspective of why you might need someone who knows about data. For example, I am a geophysicist, and I specialize in this field. But that doesn't mean that someone else (who is not a geophysicist) can't use geophysics as a tool. Same thing with informatics. There are people who specialize in it, and you need to talk to them. People (who are not a geophysicist) who use geophysics must talk to geophysicists, so that they are not interpreting things incorrectly or using it in an inappropriate way. Same thing with informaticists and people who work with data as their speciality. They support us, but they are also very important on their own. So, recognizing that I may use geoinformatics, but I am not a specialist in it. Raising the appreciation of those people, and how we give credit to data, data products, and the people who do the hard work to keep the data available, usable, and interoperable.
- Changing the culture around credit. How that might be associated around tenure and promotions--inside and outside of academia. How people can get recognition for the hard work that they do.
- Having people actually release their data. Getting to the point where people can trust that people are not going to scoop you--they are going to do something different with the data. Changing that mindset.
- Enabling people who are not as privileged as we are at ESIP. We may complain that we don't have funding, but we are fairly well supported. Think about countries who don't even have a stable political system. How are they going to be advantaged or disadvantaged by requiring that data are open there?
- Getting a handle on very large datasets. Making sure that the long tail and dark data aren't ignored, and balancing between the two. Because that long tail is getting awfully big.

That reminds me of the presentation that we heard today (2018 ESIP summer meeting) about indigenous sovereignty ([Stephanie Carroll Rainie, 1h8m mark of the video](#)). About making data open but also being aware of the power dynamics that come with that.

As well as considering if all data should be open, and the ethics around using data. I don't think that is a challenge to be solved, but it is something that we constantly need to be thinking about.

With some of the work that I have been doing around FAIR data, I have had the opportunity to meet with folks in the medical science, where you can't completely anonymize data and that is a big concern. Even though those big aggregate studies can tell you a lot about the history of disease and how disease is treated, you still have to remember that there is a person behind all of that data, whose life was deeply affected. It is almost parallel to that idea of indigenous peoples and local communities. That information can impact them. It can also impact wildlife--endangered species locations, as well as unique fossil sites. More than likely, people are going to go out and find them, and it is going to be gone. How to balance between sensitive data and public good (the public right to know)?

Where do you think that Earth Science or Earth Science data / informatics is going in the near or long-term future?

The next step is to really start thinking about how this cultural change happens. What does that look like? I see a lot of positive movement in that way. I see data starting to be recognized as an important part or the main product of a publication, particularly in sub-disciplines of science and in geoscience. Data should support the publication as a primary product, not vice versa.

I am excited about some of the innovations that are starting to happen, like transdisciplinary research. Types of data that we might not think are connected, but as an aggregate might help us understand more about the world around us by breaking down silos across disciplines. You hear stories about someone coming up with something that is ground breaking in their discipline, yet the same thing was already discovered in another discipline 10 or 20 years ago. Without the cross talk, these discoveries take a long time to happen. As geoinformatics gets involved and becomes a much more established part of the geoscience research community, it is going to force some of those silos to come down. Most people I know in geoinformatics do geoinformatics and something else--often multiple other things, because they want to see the applications.

I also see geoinformatics leading to solutions. Big sky science is wonderful and I love it, even though it doesn't necessarily have a direct application right away. But I think that informatics can help us solve problems that we see. That brings us back to the transdisciplinary nature of what we are doing--geoinformatics is often looking to solve a problem. Whether that is a data structure problem, or how do I find this, or how can machine learning help me do that. Geoinformatics is usually very practical.

[Disclaimer: Any opinions or recommendations expressed in this interview are those of the interviewee and do not necessarily reflect the views of Alabama State Geological Survey, or any other organizations listed. This interview also represents an "oral history" (a recollection of history), so its value is in the personal perspectives and insights of the interviewee, rather than specific dates, years, and titles for reference.]