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

Blog Title: Making data matter with Tina Lee

Interviewee: Tina Lee, CyVerse & Belmont Forum e-Infrastructures & Data Management
Interviewer: Arika Virapongse
Date: July 18, 2018

Blog highlight: "How will these countries apply that data and turn it into knowledge? Hopefully the social or transdisciplinary science component can then turn that into wisdom."

Arika: Could you describe when and how you got started working in Earth Science data and informatics?

Tina: In 2008, I joined iPlant Collaborative, which had just been funded $50.3 million to build cyberinfrastructure for plant sciences. They needed assistants to work with directors to help with the administrative work for cyberinfrastructure, community involvement, and plant science.

I was hired to work with community involvement. My job was to put together the Grand Challenge workshops. This was back when people were either in domain science or in computer science. They still use these terms today. But, I don't know of a scientist who doesn't use a computer or a computer scientist that hasn't had to learn about a domain science to make their work applicable. My job was to help bring together community members from plant sciences with computer people. We would "lock" them up at the Biosphere for 2-5 days. Ideally, what would come out of that was what cyberinfrastructure was needed in order to solve scientific biological grand challenges questions.

So, that was my introduction. At the time, we laughed because we were like: What is cyberinfrastructure? What are we saying when we say that? We all decided that we needed a 15-second elevator spiel, because if we said, "We are building a cyberinfrastructure for research," people's eyes would glaze over. The analogy that we used was: It's like an iPhone that you use to download apps--you not only use it as a phone, but also to play Sodoku. People got that.

What I felt was brilliant on the project's part was about data. It was how they brought people to the project originally. We said: We'll store your data for free. At the time, scientists' big barrier for data sharing and data in general was: "Oh my gosh, I have to carry around 500 of these disks or buy 3 hard drives. If I'm going to share those data with my collaborators, I'm going to have to FedEx it internationally to the UK." So, physically moving data was a challenge. In comparison, we were saying that contributors could store up to 2 GB for free. Two GB! With a press of a button, you could share that through the internet with your partners where ever they might be, as long as they had an iPlant account.

Nowadays, it's not just sharing with your collaborators in iPlant (which re-branded into CyVerse), but it's being a member node of DataOne, and federating CyVerse with a group out of the UK (United Kingdom). Everything has been magnified to the next level.

Now I work half time with CyVerse and half time with Belmont Forum. When iPlant re-branded into CyVerse, I took a break. When I got back into work, Lee Allison (formally at Arizona Geological Survey) had gotten both the EarthCube grant and the Belmont Forum grant, so I started on EarthCube. When EarthCube wound down, Lee said to me, "Have you ever met an infrastructure project that you didn't like?" He hired me for Belmont Forum, and we worked together until he passed away. So now Rowena Davis and I run the coordination office--not Belmont Forum itself, which is the international consortium of funding partners. In addition, when the Arizona Geological Survey got rolled into the University of Arizona, I re-joined CyVerse, where I am now officially their user engagement person.

For the Belmont Forum, we represent e-I&DM (e-Infrastructures & Data Management) as their corporate arm to help implement their open data policy. It's easy in the academic world, and at a place like ESIP (Earth Science Information Partners), to think that everybody knows about data management, has access to these cool tools, and knows how to write data management plans. Then you encounter international research where you've got certain second-world countries, the European open science cloud, and places in sub-saharan Africa where they are just happy to have a working and functional computer. So, it's a neat project in that it will also push full-path management of data. Leading some people kicking and screaming, while others are happy to lead the way. It's a nice project to be on.

Whereas, CyVerse's approach is: Here is a platform for research. Belmont Forum is larger and overarching. The really cool aspect of the Belmont Forum is that they don't just want to generate data for data's sake. They want the data to be transdisciplinary in the sense of pulling together very disparate things in a unique way--a co-creation of knowledge--and then applying it to see where that data leads. For example, Stephanie Carroll Rainie (University of Arizona) presented that indigenous people have that knowledge (at the 2018 ESIP summer meeting, 1h8m mark of the video). They might not know algorithms, but they know a sort of social algorithm and how to apply that. That is what I hope Belmont Forum will ultimately get to. It is still high level science, but it looks at: How will these countries apply that data and turn it into knowledge? Hopefully the social or transdisciplinary science component can then turn that into wisdom.

I'm glad to hear that Belmont Forum is addressing the disparity between countries worldwide.

We are in such early stages of that. The Belmont Forum really got going with providing funding for these initiatives about 3 years ago. Some of the first CRA's (Collaborative Research Actions) are just finishing up. Their model is still very new.

Another thing that Belmont Forum does that I really like are mid-term and end-term valorizations. It's basically a self-reflective evaluation of: How are we doing? What can we change mid-stream, a sort of course correction? Is there mission creep? Fixing things. This is nice, because many times, it's more like: Here is the money. Come back at the end of your 3-4 years and submit your report. And, hopefully, it doesn't go on a shelf. I don't believe that there is much of an opportunity to say: Here is how we can do this better. But I know that some of that does go on. For example, EarthCube is re-scoping the office. They are asking the community: How would you see that contract working with the Leadership Council? I know that there is reflection within the agency about how to make the next contract better and things like that. But, Belmont Forum is definitely interested in learning from those who have gone before. How can we do this better? Can we be more efficient? How can we use the technologies that are out there? How can we apply this? That is the part that has yet to fully fill in that circle. What is all this information about? It's got great potential.

Is there a cyberinfrastructure component to every project in Belmont Forum?

The cyberinfrastructure was its own CRA call. We call it SEI for Science-driven e-Infrastructure Innovation. That is an off-shoot of our project. Not only was there an open data policy and principles but we realized that we have to give people the tools to do open science and share data. There's also an element of capacity building--the human dimensions as well as the data planning, as well as demonstrations and best practices for using the current technology for sharing data, opening data, reusing data--the whole lifecycle.

Yesterday, the call for proposals closed, and they will do the evaluation. Evaluation metrics is our next step for the coordination office. How do you measure open data? That is a fun one too. I just love all of these projects. There are always people thinking about it, but half of it is discovering it. So you come to something like ESIP or RDA (Research Data Alliance) and ask: Who is doing open data metrics?

The evaluation metrics is an effort on our part to kickstart Belmont Forum researchers and give them the leg up. Look at CyVerse. Is there anything like that for interdisciplinary or transdisciplinary teams and can CyVerse work internationally? Would that be a good demonstrator? Who knows. That is where it is exciting. What do people end up proposing? How do we evaluate something as taking us the furthest toward open data as we can get, as well as being useful and giving us ideas about best practices? But I come toward things from a social science perspective--that is my background--and biology, originally.

It's nice to hear about transdisciplinary science of the Belmont Forum and about indigenous knowledge from some of the plenary speakers (at the 2018 ESIP summer meeting)--the "softer" social/human side of data. Because when I went to grad school, computers were taking off and everything was about what computers would enable--the next level of research and discovery. Social science always had a chip on its shoulder: We don't do hard core analytics or quantitative analyses. But what is data in the context of numbers? Just numbers. You have to put it in the

context of: How do people use or deal with that information? That is where social science fits in. And you see it in the Googles and the Amazons. They are hiring anthropologists and sociologists to tell them: How are people using data? How is that changing society? An anthropologist may not know what those numbers mean either but they can say: "These are the social systems in which data are generated and being manipulated", and understand those relationships that anthropologists and sociologists study. So there is context for people when they apply data.

Where do you think that Earth Science data, informatics, and cyberinfrastructure are going in the next 5 years, 10 years, 20 years?

Hopefully where Belmont Forum is going, which is applied. What does that data tell people in terms of enacting policies, laws, and social systems that arise to deal with climate change? It might be something like land use and land use laws. It was fascinating to hear that when Hurricane Maria hit Texas that Houston has no land use code. You own the land, you build on it. In the day and age of sea level rising, is that a good policy? Maybe, maybe not. There is lots of applications for earth data. Unfortunately, especially with the "fake news" and distrust of science, it is a little scary to think that people disregard, ignore, or purposefully diss science. As my good friend from Belgium writes: This too shall pass. Hopefully, it's just a temporary phase.

What needs to happen to move in a more applied direction?

So many of our systems have not kept up with technology. Genomics will allow us to create designer babies. Should we? That is a different question. Some of our rules and regulations don't prevent it because it's so new that we haven't developed that limitation--at least in the public mind. Should we create legislation for that? When do social institutions like churches or families decide? It used to be a decision between a doctor and a woman. Is there a social institution that will take that on? I think what needs to change, for example, is how we make our voices known. We have the technology to make one human one vote. Yet, we resort to something like the electoral college and we elect representatives. I'm not saying that is bad, but a lot of that reflects 1775. Like a freighter--our social institutions are not going to turn on a dime. Social scientists working hand in hand with the technologists could really change things faster. Like before you can really learn about a technology it's already outdated. Then there are these social and political systems that seem to take forever to change.

[Disclaimer: Any opinions or recommendations expressed in this interview are those of the interviewee and do not necessarily reflect the views of Cyverse or Belmont Forum.]