Data Curation Profiles Symposium 9/23/2012

Panel 1: Perspectives

[00:40:45]

JC: So our first panel is a group of librarians who are practitioners, who have used the data curation profiles toolkit to understand researchers' data needs in a variety of ways. I'd like to go ahead and introduce them. They are Katie Dunn, the technology and metadata librarian at our Rensselaer Polytechnic Institute. Chris Eaker, a student of the School of Information Sciences at the University of Tennessee. Daureen Nesdill, associate librarian in the Marriott Library University of Utah. Lisa Zilinksi, she's a business librarian at the University of South Florida, and Marianne Bracke, Agricultural Sciences Information Specialist at Purdue University.

We've asked each of them to talk about their experiences with the Data Curation Profiles for about 7 to 10 minutes each and to address sort of three overarching areas: What were your goals in constructing a Data Curation Profile? What did you learn from using the Data Curation Profile? And have you or will you apply what you learned from the data curation profile? With that I will hand it over to our panel.

CE: My name is Chris Eaker and I'm a second year master's student at the University of Tennessee School of Information Sciences and also a graduate research assistant with the data curation education and research centers grant with Dr. Suzie Allard who you'll hear from in the next panel. And the purpose of that grant is to train information science students in data curation practices within the context of research center. So this past summer my uh fellow students and I spent the summer at National Center for Atmospheric Research in Boulder CO, and we were working with their information science professionals, their data management professionals and their scientists learning data curation within that type of a context. So um the data curation education and research center's grant is a partnership among the University of Tennessee School of Information Science and the University of Illinois Urbana Champaign and uh the National Center for Atmospheric Research. So as part of that I was taking a class this past spring called Foundations of data curation through the University of Illinois and Data Curation Profile Toolkit was part of the reading for that course, so I learned about it there and I was also learning about all the terminology of data curation in the class and I decided that data curation profile would be a good thing for me to try to do myself for education reasons. So I wanted to apply what I'd been learning in class and so I contacted one of the librarians at the University of Tennessee, the engineering and architecture librarian and I was able to get an introduction to a professor there in the architecture department and I worked with this professor on the Data Curation Profiles that I think you have in your folder. Um he's an architecture history professor and he is , this particular project was based on, uh it dealt with ancient Roman statues bases and inscriptions so that the data set itself was actually a set of TEI XML documents written in a special form of TEI called EpiDc for these epigraphy inscriptions and so I thought it would be interesting to do a data curation profile in within a humanity realm because the only ones I'd ever seen so far were science related and I'm sure, I know there are more in the humanities now, so it was , it was an interesting project.

And I learned from this project, I learned, you know, the DCP is, the tagline is what's the researcher willing to share, with whom, and when and so I learned that this researcher is was quite eager to share everything that he had done in this project and in fact developed a website that has some really cool visualization tools in it and um but the wider epigraphy community was not really prepared for any major sharing and you know among epigraphers in other countries and in the United States and so there was a limited amount in what he could do there. He also repeatedly stated that he was willing to share, but he thought that data had to be prepared or compiled or at least cleaned to a certain point so it is understandable for anyone else to use, so that's understandable. So you that's one thing that I learned and as a student I'm learning how to do research and how to do interviews, so I think the experience I gained from this data curation profile interview process and transcription and creating the profile itself will be really beneficial for my future career in doing research and interviews. And I'm very interested in education of data management practices so that kind of thing is uh , one thing I learned that will probably important, so it was it was a definitely a great learning experience and I'm glad that I did it.

[46:10]

DN: Hi I'm Daureen Nesdill, the data curation librarian at the J. Willard Marriot Library at the University of Utah and that is in Salt Lake City. What I do presently in my position, I was asked to put together a template for the data management plans by the Office of the Vice President of Research. And from that I developed a lib guide that included even more information about data management plans and where they came from etc. In addition, I had a page on repositories since there's plenty of repositories out there and we, at that point, were not set up to accept data. And the last page was on electronic laboratory notebooks or virtual work spaces, which is my pet project. I'm trying to institute that on our campus and I am actually working with our university IT and the Office of Vice President for Research, cause somebody has to pay for it [audience laughter]. Uh we've had a library working group for maybe four years now. I was in charge of it in the beginning and we were just looking at what is data, what is data curation, what's the roles of the librarians, what's happening on campus. That evolved into uh working through the e-science institute. Now we have a new associate dean and she has more or less taken the lead and our project right now is to look at the feasibility of adding data to our institutional repository. I tried a few times to join this workshop on the data curation profile, finally went, and one of the reasons why I wanted to was that it wasn't always about preservation, it was about data from the beginning of the research project. My background is in um medical and also ecology, I've done lab research, I've done field research, so I'm more in tune with "hey you have to know what you're doing at the beginning or you're going to lose your data."

Uhm I also realize that a lot of people were uh had to step outside of their comfort zones to do these data curation profiles because they weren't in the sciences and so I decided to step out of my comfort zone and do something in the humanities. On our campus, we have a researcher who has a background in the arts, the theater, and a computer science degree. He works for a center for high performance computing. So he and his wife, who is in theater, develop telemetric performances. In other words, they develop performances for the internet. They used the grid; there was as many as six sites around the world, all working together developing a performance and holding the performance at one time. So they had people in Wales, Chicago, Boston, and Utah. So I talked to him, he said he's all for it, so we had our

interview, to interview. I was a little nervous about the interview because the data stages, I was wondering, "well how does the theater, this is humanities, what would they know about data stages?" He took the questions and answered it without a problem. He didn't even have to think about it. So I was lucky about that. One of the problems I had was I've done transcriptions in the past, and I know how much work it is, but that was 10 years ago. Thankfully, there's software out there to help me with transcription, so it did not take me couple days. It only took a couple hours to transcribe, uh to transcribe the interview.

One of the interesting things about working with this person is that he had never of preserving all of this work. He has it up on the internet, but preservation was not something he was thinking about. So he's thinking about it now. He's also going to be uh one of our researchers whose data we're going to use to test our institutional repository to see if it's alright for data. And umm. I have said everything. You know I enjoyed the interview and I the only thing I would do differently would be to transcribe right after the interview instead of waiting and being nervous about transcription.

[00:50:35]

LZ: Hi, good morning, my name is Lisa Zilinski, and I am the business librarian at the University of South Florida in Lakeland. Um that is a regional campus of the University of South Florida which is located in Tampa. My role is a reference and instruction librarian, but since we are at a small campus, we tend to be jacks of all, so we—I had an opportunity to join a working group on campus. Um the campus was looking for a content management system, institutional repository, digital asset management system. They weren't quite sure, um with my experience, prior to working for the university; I was invited to join the group. Basically there was a specific need, we had a grant that was actually an in house grant between two researchers. It was an interdisciplinary project between an English department and the psychology dept. and what their goal was to define the term polytechnic over a ten year longitudinal study. We were bringing in freshman and part of the, uh their program was they were going to be making videos throughout their time at the university on what their perspective and their idea of what polytechnic meant. So they were going to track the evolution of what this term was. The researchers really wanted to stream these videos on YouTube, they wanted them to be searchable through Google, and that was as far as they really got and so um what they wanted to do with the videos. That's where I came in and well there's so much more that you can do. What are you going to do if you YouTube disappears, what are you going to do when you lose your data. So they were really really open to the idea of using some sort of repository. My co-collaborator was a library student at the University of South Florida and she had written a paper on data duration and she mentioned these profiles. And I'm like "I want to know more about that! What is this? Let's figure out if we can use these somehow." So we got together and decided that we could use this data curation profile to build a needs document, um find out what truly are the needs of these researchers. Um before we go out and build a system or buy a system. We also reached out to five other research projects on campus. We were able to create a second profile, which was a food sciences profile. Um they are actually testing food quality of MREs over extended periods of time and extreme heat. So a completely different type of data. Now were talking about large data sets, information coming from directly from instruments. So we wanted to know how

these two completely different research projects use the same repository or use the same system. So that's what we did with these profiles, was we were able to build the needs document.

The process creating these profiles, with the first group, the linguistics profile, was extremely helpful for the researchers as they hadn't even decided or thought about what their file types were going to be, how are they going to um use this information long-term, what was the retrieval, they had just expected YouTube would be there. So it was an education for the researchers as well as for us. Um, we were able to get them to really buy into what a repository could do for them ad n they were extremely open and excited about the opportunity of using a repository that they're now looking at it for other projects in the English department, which I think is also something that we don't normally see; this is usually a science area. The researchers that we had worked with were all very very excited and open and wanted to share their information. However, the one with food science, it is highly restrictive, so we have a lot more restrictions that we need to put on the data, so we're kind of working through that. Currently, our institution is going through some transitions, so we no longer need to find a repository, we now have access to BePress through our main campus, which is very very exciting because we can now start to have the discussion on moving the data, now depositing the data into BePress. So we are currently working on that. My co-collaborator and I, we took this experience that we had and we put together a poster that we will be presenting at ACES in October. So we are probably going to continue looking at how other disciplines can use profiles and use repositories.

[56:12]

KD: I'm Katie Dunn, I'm the technology and metadata librarian at Rensselaer Polytechnic Institute, so we're a school of about 8,000 students. We are a PhD-granting institution, and as a smaller school, I wear a lot of hats. I manage our institutional repository, especially the metadata end of it, I work with our cataloger, I manage our link resolver, our electronic theses and dissertation process, and I generally kind of try to keep an eye on technologies that either help us do our work more efficiently, or give our users a better experience or hopefully both—that's the ideal. Um I also have a personal interest in working with research data and the administration is very supportive of me looking into this because they're aware that this is a growing area of awareness in libraries and that uh a lot of libraries are really being leaders in their campuses in starting services in these areas. Um, so when I found out about the digital curation profiles toolkit, I realized that this was probably an ideal tool for what we were trying to do, because at the beginning, as we're really in the information gathering stage in thinking about what services might be useful to provide, we don't' really have the existing services around data curation management and we thought that this was a good way to uh gather anecdotal info from our faculty members on how they create and use data. And though the digital curation profiles, they're, they're very structured, which is a good thing, because I really didn't know what the right questions were to ask and I wasn't sure what my final product would be either but the great thing about the toolkit was that it gave me a structured way to interview and a structured way to write it up that would make it comparable with other profiles in the same discipline or other disciplines. So that was really good. But the other good thing about the profile is that it's very flexible in that the questions that are in there really um, allowed the conversation to bloom in a variety of different directions and brought out topics that might have not necessarily—I might have not necessarily thought to ask about. So, that's what my

expectations of what the toolkit would do for me, turned out to be pretty much right now with the interview that I ended up doing and writing up.

So another, another reason that we wanted to use the data curation profiles toolkit was we wanted to use these interviews as a way to build relationships with researchers on our, on our campus around these issues and build awareness that the library is indeed interested in these issues and that they should come talk to us about things like this, even though we don't actually have any services for things like that yet. Um and the interviews also allow us to talk more knowledgeably to other faculty members about these things going forward because it really gave me a an in-depth understanding of what a particular research data work flow will be and I feel like after I've done 3 or 4 more of them, I'll start to see commonalities in different peoples' work flows and then even if I talk to a researcher that's not in any of those research disciplinary areas, I'll have an idea of like kind of what some of the common issues are amongst different disciplines.

So in keeping with our goals for doing the interviews, we came up with a short list of people that we knew were doing interesting work with data and who we thought would be enthusiastic about talking to us as we were doing these initial interviews so um my first interview that I did was a faculty member in physics and astronomy and she is most of her work surrounds collaborations in large uh large-scale, astronomical surveys. So most of her data is managed by these huge multi-institutional projects, like the Sloan Digital Sky Survey and the LAMOST Telescope Project in China. Um so the primary stages, or the first stages of her data are managed by these huge collaborations where the uh data management process is like uh incredibly complex, but on the other hand, she personally, although she's involved in the planning for that data management, she doesn't have to really worry about those first, the management of those first data stages. Um so we, we talked mostly about the stages of the data that she manages personally for her own research questions. And the really interesting thing, and something that I had suspicions about, but it was nice to have it confirmed by actual discussion with a faculty, was as an astrophysicist, she really has no objections to sharing anything and everything of her data and I think that's a common viewpoint in that discipline, but she's concerned about the amount of effort involved in getting the data to a point that it would be useful for other people to use and understandable. And the other interesting thing was that her thinking about that actually changed over the course of the interview, and she said, well you know that when we first started talking about this, I felt like oh man, this is going to be a lot of work to put this together, and I'm not quite sure what I would do, and she didn't think it was a realistic thing to expect to do on an ongoing basis. By the end of the interviews, without me really trying to, without me trying to influence her in that direction, she said that she thought that it would not be so arduous to do this and uh she thinks it could be potentially really useful. And the target data though she identified for sharing that would be both not too much work to share and of the most use is the data, behind the figure data in her published papers and the scripts that were used to generate the uh the figures from those data. Currently she said that in her field people will often pass the post-script files around for the figures but, and then add additional data to them or put an inset data and modify it some way. But it's really kind of a pain to work with the post-script files. But um and so it would be good for people to start from the data and if they can't run the scripts that you used to create the figures, they can at least use the scripts as a pseudo code and then identify what you

did to create the figures from that and then write your own scripts on it, and she was like, "Oh we basically have all that put together and it's kind of already documented by the paper if we're talking about the data behind figures," so she said this is the kind of stuff that she has to kind of informally dig up on her hard drive and email the people now and she said that if "well, if we could get it in there at the beginning of the process, it would just be there and I can point people at the repository."

Um so, how we intend to apply what we did with the digital curation profiles is I hope to 3 or 4 more of these and what I'd like to do is compile a report sharing what we learned on them and I'd have to clear that with the interviewers because we've um, we didn't really talk about how they'd feel about sharing it within the campus. So I need to address that with them. But what we want to do is create a report that we can um share in the libraries, with the office—the chief information officer of which we're a division, and also the office of research and other research groups on campus that are interested in data issues, and really kind of use what we learned in the interviews to kind of start a conversation about data management on our campus and steer that on a more useful direction. Um I really, and on a personal level, I really feel like this is the best method I've seen for gaining in depth knowledge about research data practices in a particular discipline and really address one of my, what I've felt is my main ability gaps in providing services in this area is the lack of knowledge from the researchers perspective about how data works. So this, I feel like, is a good way for uh librarians and other service providers to kind of address the gap between what they know and what the researchers take as like implicit knowledge about how the way they do their work.

[1:04:10]

MSB: Ok, well, good morning, I'm Marianne Stowel Bracke. I'm from Purdue University. The project I want to talk about is a little bit different than what we've heard so far. Um, I've been I've been lucky enough to be part of the data curation profiles process and at the very beginning that Scott and Jake have mentioned and so um, I never actually attended, I don't think, any of the formal workshops so I don't, I mean, you know, I kind of grew into it. But I have a couple of agronomy faculty who have been interested in um sharing their data even before I got here six years ago, so I was very fortunate to have that. And we've been working with these agronomy faculty and um trying to figure out kind of what, what their needs are. They are very open to sharing and uh they were some of the people that in in working with some of the initial documents, they were willing to participate with that. We wanted to take that further and so Jake and I started meeting not only with these faculty members, but with the people in their lab—the students who do a lot of the work. I should also mention that uh that it's kind of unique about their work is that it's, what's important about it is that it is longitudinal, that it isn't really a project that begins and ends and a new one starts. The important part of the data that their gather on crops and water quality, the value is that they have 20 years going back and they are going to keep doing it. And that's really what makes it important. In meeting with her grad students though, we all had an aha moment, the grad students, the faculty, and Jake and I, in that, in trying to share, in trying to talk about data and what people were doing and practices, we realized immediately that they weren't even necessarily communicating with each other and they had, they had differences and someone said, "You know of course this is reported as dry weight at 12%." Someone said, "No it's not, it's dry weight 15%." And someone, an international student said, "Well you know in Africa, it's 15%" and you know they all

realized, we've never sat down and talked about this. And you know, it's sort of an assumption of the profession. So we decided to take the profiles, and Jake and I interviewed six of the grad students working within the lab. You know, it's really kind of an interesting bottom up approach. Now the faculty members have a very good idea of the big picture, what they want to do, their ultimate objectives, they have ideas about sharing, they're every open to that. But the grad students are the ones running the individual experiment that add up to the bigger research. And in doing that we really go not, not just a picture of a piece of the discipline, but really of this particular lab, and how it was functioning. And the students definitely had um, deficits in some areas, like they had never thought about why they would want to share their data, but it had just never occurred to them, they just weren't that far along in their profession. Um, intellectually property issues, who owned the data, who might own the data, what happened to the data after they graduated? They were a little bit more practical, they wanted the data, they wanted to be able to write up for their, their thesis or dissertation and be done with it. So they didn't have that, that long-term understanding of where their data happened to fit. We also really got this understanding that data, the data that they're gathering in particular, it's all flowing, that um they often, students inherit the data, um from you know, another grad student and they're going to hand it off to another grad student when they leave. So there's never really a good beginning and a good end, so you really have to just jump in with this. So after looking at these interviews as well as knowing what we know about the faculty and the interviews with the faculty and having a tool where everything is pretty consistent allowed, even though I did some interviews and Jake did some interviews, it allowed— we were able to have kind of a common understanding. And we wrote up a report with a number of suggestions for how this lab might want to rethink their data, data management. And some of them were kind of practical in thinking about how they would want to save their data, or name data files, um definitely how they want to describe their data because a lot of this is handed off to either future grad students, or they have other grad students working in the lab that aren't agronomists but are perhaps ag-economists, who don't have an agronomy background, but they need the data, or modelers, people who need all this data and they need to be able to know how to work with it, without—and how could we circumvent having the only ways someone else could work with that data is for two people to sit down and have a serious lengthy conversation, very detailed about explaining how this is going to go on. But then there are also things that we suggested that were almost some cultural changes within the discipline and thinking about how, they're are already very good at training their grad students to be future agronomists, but what could they pass on that, so that you know these future grad students are going to be our future faculty and our future data providers. Hopefully this report, and um I apologize to Jake, because what's really holding up this report is me, and a little bit of a literature review to round it out, but this is this is kind of a micro-application of this tool, but it's really thorough and so you know it's not necessarily going to be something that is deductive but rather inductive. Um and it-it could be an example of how you could really get to know that. I should also add that these professors' data, parts of it are the first data that have gone into PURR. And so, um and we're working to add more of their data into PURR. And Purdue, PURR is the Purdue Research Repository and their data is really quite manageable for, for putting into something like PURR. It's not gigantic, like the astrophysicist data, it actually is very detailed, but mostly in excel type files. And a far as application, I think we are going to keep doing this. The first one we—I think a lot of people would say they probably took longer than expected either because of the interview process, the transcription process, but having done them, you

really do get better. You get better at asking question, maybe knowing when that doesn't have to be followed up on and that does. And you do see patterns, even though, and I work with the college of agronomy so I work with social sciences as well as applied and life sciences and even with landscape architecture, some of the arts and humanities. Even though their very disparate, you do see some things that are common and I think when you start getting to things like storage and intellectual property, um those really aren't all that different for people. They are back up, you know that's a big black hole for a lot of people. And um it just is. And I think, you know, we're going to continue to do something like this and I'm not entirely sure where it goes, but I've got lots of faculty that are just, that don't know it yet, but are in line with, for being approached.

[1:14:00]

JC: So while we're getting the battery exchanged I'm going to slip in here and borrow the microphone for a few minutes. Thank you all very much that was really informative. I do have a couple of questions for you before turn it over to the larger group. First I'd like to connect with Chuck's key note. He brought up some very interesting points I'd like you to address.

So he mentioned that we kind of struggled with our definition of data, so we talk about data curation, what is it we're actually talking about? Most of the large meetings that I go to, that at least it pops up and if not first thing, fairly close to the first thing of what's the scope? What's the actually, the meaning of data? And I'm wondering, did you encounter that kind of need to define data in your sort of micro interactions with researchers, when you say "I want to talk to you about your data" did they know what you were talking about? Were they able to describe that fairly easily?

DN: Okay, I didn't ask him if wanted to talk about data with me, I was asking about research in general, so about the whole process. And no he didn't a problem with the word data, we mainly knew what it was.

LZ: We had two separate profiles and the food science profile, the researcher was pretty confident on her what data was and was pretty confident the difference between the different data sets. Umm, the linguistics profile was a different story, and I kind of did the same approach, I didn't go up and say "I'd talk about your data, let's talk about your data." It was, "I'd like to talk about your project. I understand that you have some files and some videos that you want to make available. um during the process we started tracking your results and that's when we started talking about excel files. So it is, again, a little bit of an education, but we didn't necessarily come up with a definition.

KD: In talking with our astrophysicist, I didn't really have that issue because it's a pretty traditional data driven discipline. In soliciting data for our little pilot repository that we're trying to get off the ground, I did kind of have to reframe how I was asking about things. Like I had a professor in architecture and acoustics who ended up depositing several videos and um MATLAB scripts with us that they use to analyze their data. And initially when I was talking to him, he wasn't sure if we would want the videos or want the MATLAB scripts. So I had to make sure when I was saying data that I followed by saying, "well we mean any of like the information that supports your research or helps people refine or understand

it" so that they were sure that we were thinking of it really broadly at this point it's worth we were thinking about how to work with these things.

MSB: I would just add that with the agronomy folks um, they really didn't have a problem um, they kind of knew what their data was, but in the interview getting them to talk about the data product that was really difficult. They wanted to talk about their science and kind of the results um, but when we were trying to ask the questions that said your data is—starts out her and goes through this stage and that stage and you want to deposit something—I think that maybe that's something that's a little bit of a flaw on our part that that's how we think and we're going to go through a process, but um trying to get them to think about, you know at this stage, "what kind of files do you have?" and really backing it up to that basic level—that that was where the challenge was with the people I interviewed. But I think they understood um what data was.

CE: We talked about, when we started out the interview and I saw we want to talk to you about your data and he said, "well, all I have is XML/TEI files if that's ok with you." And I said yeah, that's your data so let's talk about it. He was eager to talk about those files once he realized that was okay with me.

[1:18: 28]

JC: Another question that I have for you is that conducting these interviews, while certainly librarians have sort of a base to work off from the reference interview, this is sort of a reference interview on steroids, it's really intense, it's really in depth, it's an area that we haven't been used to working in. How did it feel to go in and conduct those interviews? How did you prepare yourself to do them? Are there things that we could do to help others get that same level of preparation to do these things?

CE: Considering I'm still a student, I haven't done any reference interviews, but um I um learned about them. But I read through the profile several times and the questionnaire several times just to get a good understanding of the types of questions we'd be covering throughout the interview process. And then I read at least three complete profiles that I downloaded form the website to see what kind of information they obtained and how they prepared the final product. So it took some, you know some extensive preparation just to prepare myself so I felt comfortable going in with what I was going to be talking about.

DN: I went—I attended a couple of telematic performances which were interesting. I attended one where I was in the audience that was being filmed. And then I uh the second one I went to was streamed into the library so I saw it from a different perspective. I know that uh, the PI has a website, anotherlanguage.org, so I went there and looked at some of the previous performances and he's got a lot of information on that website, so that's basically how I educated myself. And that's more or less the same way I educate myself before going out and talking to a new researcher in my department is finding out a little bit about what they've published, what they've done, and then go in and ask a few questions.

LZ: So prior to working for USF, I actually worked for Price Waterhouse Coopers and I dealt with large amounts of data. So I've had some experience already in some of these data type interviews. What we did with our researchers working with my co-collaborator, we had pre-interview meeting whether it was

in person, which we did meet with one researcher in person, or communicated through the phone, or via email. With our food sciences profiles was way outside of my area of expertise, and what we ended up doing was after the interview we still had these questions, these clarifying questions that we needed answered, so we actually did a lab visit. She took us through all of the equipment, we met postdocs and grad students that were sitting there and they were really able to tell us more about their data. And I think talking to the researcher before you actually sit down to um do the interview is very important because you'll uncover things that you may need to ask in the interview during the first conversation.

KD: I felt pretty prepared for it, it helped that I went to uh one of the workshops, but given that's, that was kind of a limited term thing and he only did 12 of them, I was trying to think of like what part—[audience laughter]—only! I only went to one, but they don't have the opportunity to do that anymore, so I was like what did I get out of that that other people could get that experience without actually going to the workshops? And I was thinking that the most valuable parts of it were talking to my peers that were thinking about doing these interviews or had done these interviews, and getting a chance to kind of informally air some of my doubts and concerns about the process and talk about how they were going to address those. And also to watch some of the videos of Marianne working with her faculty member to see what the, what it actually going to be like, and that way I could kind of mentally rehearse going through this process. And also like Chris said, reading the profiles ahead of time and knowing what kind of information other people had gotten from their faculty members. I had kind of a mental idea of how the process would go the toolkit told me approximately how long it would take which I found really really useful and that was right on for us. So I was thinking that in order for other people going onward to feel prepared going on to their first interviews, I think that doing uh being able to watch those videos is helpful and I think Jake had said they're working on making those available on the web. And I also was wondering if maybe to get that kind of um airing of concerns and bouncing ideas off of people, I wonder if like an informal mentoring process might work, like if you were someone who had never done an interview before and you could be matched up with somebody that had already done some, then you could really um I think I would have found that reassuring if I wasn't able to attend the workshop and talk to people about those kind of ideas then. So that would be a really you know low-effort way of trying to give people more confidence in going to the interviews. Another thing that I really found helpful was—I had to communicate with the researcher that we weren't talking about her data process in general, it was very much attached to a specific, a specific research project and a specific dataset and what modification and what changes it went through in that process. And a good way to frame that was to um was to select a specific paper and ask can we talk about the research you went through in this particular paper, and she turned out to say, "oh that's not a good one let's talk about this one instead." But at least we had a starting point then and I was able to read the paper and have more confidence that I would know what she was talking about when I was going in there. And now having done that once I feel like I wouldn't have to—if we were going to do another interview based on another project, I wouldn't feel as much like I needed to read the paper ahead of time, so we could maybe do one on a project that maybe had not yet been published, I would have more confidence since I already have that background information. But for the first interview with a faculty member I really feel like it's useful to do one on something that's been published already so you could read that documentation.

MSB: If I had any idea that those videos were going to be so widely used, I would have rehearsed or something because actually the faculty member that I worked with is, I know very well and we have a very good relationship and you know you turn the cameras on and suddenly it's like "uhhh."

KD: Oh you were so good!

MSB: But um, I guess, I would echo a couple of things. In that particular case, I knew her work very well and uh you know, I got to go out to the actual field station and look at their work. But in other instances where you don't know their work, even picking a paper and reading that ahead of time does give you a starting point even if they decide to work off something else. One thing I think I would recommend, even though this is not my style and how I like to work, but you know sometimes it's good for you, umm and I don't know if Jake will go into this at all, but there's another grant that we're kind of using, another variation of the curation profiles and even though a group of us worked on changing them a little bit again and again and you know you wordsmith and things like that, when I actually did the interview, it felt very foreign and I realized, I wish I had practiced. I wish I had sat down and had someone play a researcher and just read the questions out loud and realized how they sound in an interview process, not just in an editing process. And I know people are loathe to do role playing, I—I myself hate to do it, but um, I don't think it's a bad idea. I think, you guys were saying that as well, you get to kind of informally air, you know, "you know I'm not sure what that question means either" or "you know like, if I were put on the spot to define repository, what would I say?" and um, so that would be something that I would , I might recommend, that you know, even if you just wanted to do it by yourself in your office and read things out loud and hear how they sound as opposed to reading them, I think that that makes a big difference.

[1:27:37]

JC: So I'd like to open this up for questions for the audience. Does anyone have a question for our panel? If you do, please step to the mic.

SB: [in audible] … would you say yes, I'd do the exact same thing?

MSB: I – I think I would probably do the same thing again. I actually think it was pretty useful. Now I'm comfortable with the tool, the tool's lengthy, it's in depth, but I think I would do it the same way.

KD: Um I'd do it; I would also do it same way. I think I didn't make a lot of mistakes that I would have if the toolkit wasn't so explicit about you know the best way to do an interview, and you know, "bring your two recorders" and uh "this is how long it's going to take." The one thing that I would have done is –I echo Daureen in saying that after you do the interview, get started on that transcription and then do the profile right away. I feel like there's some things, just a few things, that would have been easier to clarify sooner after the interview that were lost. So it's easy among other responsibilities to let something like this sit on the back of your desk, but if you can keep chipping away at it every day, the transcriptions don't take that long at all.

LZ: Um I think for me, for the most part, I would have done it the same way. However, with the food science profile, I almost would have gone to the lab first and then conducted the interview and built the profile, especially since uh seeing the instruments work, um, made a lot more sense when we were talking about what is the data that's collected. You can actually see an instrument work and she showed us the actual reports that came from these instruments. And I think that seeing that all these light bulbs, "oh that's what that meant! When we were doing the interview, I could have asked, this question." So I was lucky to have the opportunity to ask follow up questions at the lab visit. But I think I would have done the lab visit first.

DN: I did do a second interview, she has since left the university so we're not going to be using it for anything, but it was a good practice because what I did was ask another librarian to come with me and so it was two of us interviewing and this person got the experience from my previous experience. And this research was not in theater it was in medical research.

CE: I think if I did another one, or if I had to do this one again, I would learn a little more about this researcher's particular project that we were going to be talking about. I didn't know what epigraphy was going in and but I learned it throughout the process of the interview, but it would have been nice to have a little better understanding of not only epigraphy in a broader sense, but a particular project that he worked on, so I could ask more probing questions and extract more information. So.

[1:31:07]

[question] CW: So my name is Charles Watkins and I am a colleague of Jake and Scott's and Purdue Libraries, and uh you clearly put a lot of work into the data curation profile preparation, what are your hopes and expectations in how you are going to publish that work? How you're going to get the kind of credit that you would hope to get as library faculty members?

MSB: Okay, Um, well as I sort of hinted at, because we had done such a thorough report, so I think Jake and I are going to turn the thorough report into a case study for preparation. And I um, so I –I think in our case that's a pretty clear route.

KD: I'm actually not sure about publication; I did go through trough the IRB process at our institution so I am cleared to use it outside of our institution, um if I choose to. But I'm not that experienced in publication, and it's something that I would like to get into, but I'm a little unsure of what the form, um how to make what I've learned from this into a form that would be publishable. But that's a general challenge I have with everything I do in my job, so.

LZ: Um we have already put together a poster presentation for ACES in a month, um talking about the process that we went through in creating the needs document and creating the matrix of the document of the different systems we looked at and how we compared those two to give a good methodology on how can you use these profiles to choose an IR or a digital asset management system.

DN: I actually never thought of publishing this first interview I did. I did not go through IRB because we basically want to use the data to see if we can add data to our institutional repository, but I see where it

can be part of a larger paper, where it helped us make our institution repos—pos, I can't say that word—institutional repository data ready.

CE: As I mentioned earlier, my goal in doing the data curation profile was mainly educational, for my own benefit. So I didn't really consider any future use of it. But I think it's, in all my free time, I hope to maybe do some more, so, because they are time consuming, but they're definitely worth the time, and I think in the future wherever I end up in at my job, there's definitely a use there in learning about researchers uh projects and their specific research habits.

[1:34:23]

JC: So I'll follow up with a question from Charles's question. I think there's two things going on from listening to your responses. One, there's the how can we leverage what we've learned from doing the profiles into some kind of publication or presentation or poster? And Marianne and I are doing that as you mentioned and Lisa has done that as well. And then there's also the publication of the profile itself that I'd be curious to know, given the amount of work it takes to produce a profile, are there things that we could do or should do to encourage people to share their profile with others through our website or other means, or incentives or ways to make it easier or more beneficial for you that you can think of?

MSB: I—I overheard um a couple people talking yesterday and they said that they were thinking at the reception about whether or not this should be a publication and they're like, well I'd certainly put it on my CV and I thought, "Oh! I'd never thought of that." And um, so I think that modeling some of this, you know, like showing it as a publication in sort of this trusted data curation profile repository, and then maybe having suggested ways to cite it. Or making them, and we talked maybe about having a database rather than just PDFs.

KD: Um, I uh, my profile's kind of in a preliminary state, I haven' enriched it with quotes from the researcher as much as I'd like to, but once I do that , I do plan to run it past her before making it available on the website. And I think probably getting her feedback is going to be the biggest barrier to me, because uh, researchers are very busy and it might take a while before I get her feedback. I don't know, I fully intend to put it up on the website. I think the citation information and like here's how you can list it on your CV would be good. I had planned on putting it on my CV, even though it's not really a traditional paper or whatever.

LZ: Um I think a community practice would be something that would be very beneficial that would list how to cite, having a database of profiles, um having ways to cross reference same discipline or possibly finding out if there's people are using the same type of data sets to make it a m ore robust community, I think that would definitely be a benefit to using them.

DN: I'm going to find a little more information that I didn't have in my original profile, and then add it to the Purdue site. Um one of the things we haven't talked about is, I guess we have a panel on it later, I forgot about that, is to start talking to library schools about these profiles, and um somehow educate more librarians, but I don't know that takes money, right?

CE: I've already, or mine's already on the DCP website, but I think that uh, I think the idea of the citation, although I don't know what people would want it for but if they do, I think it'd be great to have a citation, some sort of standard citation that would indicate this is how you cite a data curation profile.

[1:37:59]

JC: I will mention that uh Scott and I are working with Charles to look at how do we publish the DCPs in a more structured manner through the institutional repository. So we're looking at creating a series, a publication series of data curation profiles, which I think should address some of the things like citation format and stuff. So it is in the works and we are working on that. Other questions from the audience? Yeah, Suzie?

SA: I think I'll stay low so the camera can see over me. Suzie Allard from the University of Tennessee. One thing all of you, you know you've reviewed, you've mentioned reviewing other profiles and looking at them, so your experts in reading these as well, what do you see the value of reading a profile from a different institution for your own institution to where other people might be learning from your institution for their institution?

MSB: I'll crib an answer, from what I heard from yesterday, um, that, uhm, that even if you're not using, that if you've never used it before and you're writing—especially when you do the write up, you know you've done the interview, but looking at how other people have uh synthesized the information and the way you write up the profile to the questions are not perfectly matched like it's question one equals section one, it's kind of like question one equals section 4.2 and umm, seeing even if they're not the same, seeing how other people how other people have done it and taking the different parts of the information and synthesize some I think is useful, just from a very practical point of view.

KD: I think that the breadth is going to be a big benefit because part of what I'm looking to get out of doing these interviews is, learning how people use data in a variety of disciplines and I—I 'm one person and I can't do 12 or 15 of these, but I can read 12 or 15 that other people have done. Doing them on your own campus does allow you to build relationships with those researchers and potentially um recruit content if you at that point. But I still feel at that point, other institutions, particularly ones that are in similar disciplines to the researchers you're working with, it still gives you that breadth that breadth that's so important.

LZ: I have to agree that uh reading, actually with both of they said, as one looking at how other people have put these profiles together, kind of helped guide our process. There were sections where we weren't fully clear on what really belongs there or how much belongs or if something belongs somewhere else, and just looking to see how other institutions have done them, it's okay this is making sense. But also to see, we did an interdisciplinary and then we did a food science one, so seeing how other institutions handed something interdisciplinary was something that helped us in putting our profile together, and I think in the long run it will allow us to have um deeper conversations about data with our own institution by seeing what's happening in others.

DN: I guess I'll have to echo what everybody else said, but I don't think the institution means as much as the disciplinary or multidisciplinary focus of whatever profile. It also allows, you know if you are expecting to go out and interview somebody, it also helps to read some of the profiles and see exactly what other people have already picked up from this similar discipline.

CE: I believe there's a real benefit in comparing research practices or the, you know whatever we cover in the profile from one discipline in one institution versus the same discipline at another institution to see how they compare and then you know, I think when the profiles reach critical mass, we have a lot of them, then we can say okay, physics researchers generally think this way and you know they're pretty much consistent across the board or maybe not. So we', I think it would be interesting to see that information through and you can find that through the profiles.

JC: Okay, so we have reached 10:30, so we please join me in thanking our participant panel.

SB: We're going to break here, so we'll meet back at 10:45.