**Interview Transcription**

Date: February 14th, 2013

Time: 9:00 AM – 10:20 AM

Interviewee: Dr. Deana Pennington

Interviewer: Dr. Ann Gates and Software Engineering class

The transcription below is taken from an audio-visual recording. The interviewee and interviewer were asked to review the transcription and to check for correctness. Additional information has not been added—only minor edits and clarifications. The transcriber used “…” when the audio was not clear and could not be transcribed. Brackets are used when the reviewers added clarifying words. Words or phrases that are not adding information may have been deleted.

Dr Gates: I want to thank everyone for coming here, and I want to introduce you to Dr. Diana Pennington. Dr. Pennington has a lot of expertise in environmental science she’s an ecologist and a geologist by training. She also works quite a bit in interdisciplinary research. Working across different disciplines and bringing them together is really important for this particular project. I have given her the questions that have been put together [by the class]. I’m also asking her permission to record this interview. I want to make sure that only the class has access to the interview. It’s not going to be accessed outside of the class. You saw how she acted to making it public. This is a reason why you have to let the customer know that you are recording it, okay? I’m going to start off with some general questions. We talked about the benefits: what do you see as the benefits of the system that is being proposed?

Dr. Pennington: We spent a lot of money and time putting sensors in the field. Am, a huge investment of resources that going to design and get equipment out there and it’s there for a reason. We want data, we want data consistently, we want a certain kind of data, and we want to know that that data is going to be useful to us when we do our analysis. So an important part of that is am, making sure that the data that is being collected has a certain quality, a high enough quality that we can use it, ah so quality is huge. And any sensor networks these are high-end pieces of equipment. There are failures, there are things that aren’t calibrated correctly, there’s many things that can go wrong. The other thing that happens is that sometimes things happen out in the field that we aren’t expecting. Maybe we get snow in July and we weren’t expecting that. Am, so something, an event happens there that’s unusual it might be interesting to us. We might want to do something, respond to that event in some way but we can’t if we don’t know when it happens. Am, so those are sort of the two things that are important, we need to ensure data quality, but we also need to know when something, when an event happens that might be interesting to us.

Dr. Gates: What do you see as the limitations of the current data specification tool?

Dr. Pennington: The biggest one is that it is written and designed and uh formulated for a computer scientist. It’s got the language of a computer scientist, it has the interface that is relevant to computer scientists, it talks about queries and codes and rules and that’s not the language of science. Am so uh while it does what it needs to do or it’s not usable by a scientist.

Dr. Gates: Okay, will there be different types of users? And if there are, what restrictions might be on those users with respect to using those tools?

Dr. Pennington: (With respect to using the tools) Ah, well the kinds of users that you are likely to have would either be graduate students, it could be faculty, in some cases there would be somebody designated as data manager and that person might or might not have a computer science background, usually not, they might have some train-, some informal training in information technology. Am, so, they might be completely new and might not have a clue on how to use the tool, there might not be any training available, there might not be anybody to show them how to use it. On the other hand it might be someone who’s doing it on a daily basis and is very familiar with it. Uh, and you have to design for all that entire spectrum. Am, so I think the way you reason that it becomes important is the kinds of things that you need to provide to somebody who is learning it and using it, might be tedious for somebody who is using it all the time and familiar with it. Am, so that has to be taken into consideration am, in terms of restrictions there’s probably in most cases one person who is responsible for the data and they should be able to do whatever they need to, with it. On the other hand you may have people using it who you don’t want to be able to modify, append to the rules or be able to do it, maybe make comments on it but not want them to actually edit it.

Dr. Gates: If there’s a follow up question, I am going to allow it on this one. if you have a follow up question…

Dr. Pennington: Yes?

Student: As far as users do you want someone to be able to maintain the system? As far as a public view and then also someone to maybe change the view those types of users? Maybe ah, sorry, an administrator? Yeah, do you also want that sort of layout, also?

Dr. Pennington: Yeah I would imagine somebody who’s administrating ah the data and the properties as opposed to others who might just be amm yeah there can be an administrator.

Dr. Gates: So are you going to require authorization then for users?

Dr. Pennington: Yes, there was a question in there about being UTEP, but it’s not necessarily UTEP but you do need to validate your users.

Student: Is there some kind of like uh a help session or some kind of walk thru?

Dr. Pennington: Well unless you can design it in such a way that it’s so intuitive that people can figure it out without any kind of help then you probably do need to put in some sort of wizard or training or something to get them started. Although if you can design so intuitive that you don’t need it that be great.

Dr. Gates: So when sharing data between scientists, what type of personal information will be displayed in the documentation?

Dr. Pennington: Well certainly you want to know who generated the data, what did it and who’s the original source. Ah you want contact information about them in case you have questions about the data. Ah, you might want to know the institution that there are at or where was the institution they were at when the data was collected because people move around. Ah, If there are, because people move around and projects you need to know who was the original contact but also whose the current contact. And again, their name, maybe their position, institution and some sort of contact information

Dr. Gates: So I’m going to am ask you-- that’s for the data, now let’s start talking about data properties.

Dr. Pennington: So I have to say I got confused reading thru your questions because there are three different kinds of data that we are interested in here. And it was not clear to me sometimes which kind of data you were talking about. So there’s the data that’s coming off the sensors, for your checking, there’s the data about properties you’ve developed and designed so data properties, the properties themselves are kind of data, and then there’s the anomalies that are detected. So I need you when you’re asking questions about data I need you to be very specific about which of those three kinds of data you’re talking about, so can you repeat the question and be clear what data you’re asking about.

Dr. Gates: OK. This was asked by, oh I didn’t write it down, which team asked that question? Okay you did.

Student: Repeat the question.

Dr. Gates: When sharing data between scientists, what type of personal information will be displayed in the documentation?

Dr. Pennington: No that was the one, we already answered that, there was another, you said something about, the properties.

Dr. Gates: Oh I added that. I said you were talking about sensors. I was going back to your three types of data. I’m glad you clarified. Okay so the same question, but now with data properties.

Dr. Pennington: So that data about the properties? Same, we want to know who created the properties originally and if there is somebody that’s changed and somebody else is responsible for that property you want to know who that person is, so you need to current contact about the property but you also want to know who originated it or how, who originated it. And if its been modified, there may be a revision history on the property and it might be different people involved with that revision. That whole providence of that whole property is something that I would be interested in.

Dr. Gates: As well as location of the property?

Dr. Pennington: Location is huge. Let me give you an example. If I’m collecting data out here in the desert the properties that I design about temperature and precipitation, what I expect in terms of temperature and precipitation are quite different there then what I might expect if I have the same tool in the artic. So properties are specific to their location and a time period.

Student: I have a question, that personal information, contact information should it be available to, for any user to see or just for certain users?

Dr. Pennington: Well I think that any user who has permission to view a property should be able to see the providence of that property. So permission really comes into play with who is allowed to see this particular property. And that needs to be specified by who’s responsible for the property. I might have a property that I’m working on that am I want private for some reason. Maybe I’m doing some research and I’m trying to keep it private because if I release the property my competitors are going to know how or what data I’m collecting so I might not want to share the property with anybody. Or I might want to share it with a few people. Or I might want it to be public. But anyone who has permission to review the property should also be able to see the contact information.

Student: So they get to say who can see it, basically?

Dr. Pennington: Yeah, Yeah.

Dr. Gates: Okay, so she talked about location, any questions on, any follow up questions about that?

Student: About how many locations do you guys have?

Dr. Pennington: Right now we use these particular, well it depends on what you mean by you guys, okay? Our particular research group.

Student: I mean the sensors, how many sensors location is the system going to be managing?

Dr. Pennington: Well I, I mean it’s a new system, right now it’s just us that are using it. But the sensors themselves are used all over the place by all sorts of different people. But this particular Properties Specification tool we’ve developed in house we would like it to be more adoptive elsewhere so you can say that in the long run what we would like to see is to be used worldwide. But right now it’s used in the desert here an up in the arctic. So those two extremes, desert, arctic is what we have to deal with.

Student: Well the arctic is a desert.

Dr. Pennington: It is a desert, that’s why we are up there. (Laughing). It’s a cold desert as opposed to a hot desert, that’s why we are doing research in those two places.

Dr. Gates: So I’m going to cue you a little bit, only this time. When we talk about location, is that clear to you how you’re going to record location?”

Student: Geo tag.

Dr. Gates: hmm?

Student: I’m sure we would use some sort of longitude, latitude for every specific -

Dr. Gates: Can you ask a question around that? This is Jeopardy… (Laughing)

Student: When, I guess ah tag a property with its location, what other information is required?

Dr. Pennington: Well I think it would be important to know, I mean location is a complicated subject. So let’s say the camera there was a sensor, it’s got a specific location, but the sensor itself is recording not just at that location its got a footprint that it’s measuring. And that changes from sensor to sensor. So it’s important to know if you’re designing a property for a sensor, what is the footprint that you’re measuring and that the property applies to. The same thing is true in time. You may measure anything at a specific point and time but sometimes what you’re after is a time frame, a window of time. So I might be looking at, I might report a property that is interested not at just what’s happening in this point in time but how does it compare to prior times or future times.

Dr. Gates: So when you look at times, or whenever you look at anything. [For example,] whenever you’re working on a system that has some sort of measurement, you really need to understand that measurement into what degree of accuracy you need.

Dr. Pennington: So generally we talk about resolution and how is it resolving it to a very fine space in time interval or is it resolving, the sensor has some resolution of what its measuring or is it a broader space in time. You can think about it that sort of like pixel sizes you know you get different resolution and you have to know which resolution is of the instrument.

Dr. Gates: So it’s instrument dependent?

Dr. Pennington: It’s instrument dependent.