**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.

Dr. Gates: So the granularity can go to microseconds?

Dr. Pennington: It could be measuring microseconds, it could be measuring every ten minutes, it could be measuring once per hour, once per day.

Dr. Gates: So there’s an interval associated …

Dr. Pennington: Yes, Yes there is.

Dr. Gates: So when you talk about footprint, what would you all like to know about what a footprint is? You all know how to capture a footprint?

Student: For each different sensor, would we be able to access from the individual sensor data how big the foot print is depending on the time and the scale of the actual I guess area that is recording, or would we actually have to go and kind of find out ourselves?

Dr. Pennington: Well so, there is two different questions there. One if there is an inherit resolution set that is determined by the instrument itself, but then there is also things that the scientist can’t modify, so an instrument could be capable of measuring every microsecond, but I might only have it measuring once an hour, so those are two different things.

Dr. Gates: The footprint. What about the footprint? How do you know the location?

Dr. Pennington: Well it, let’s say you know the location, and usually, let’s come back to the camera there, I would have some sort of GPS reading about the location and now there is an error associated with that GPS reading, so I don’t really know where it is within certain error accuracy. Then I also know because of the instrument is going to be carrying information, metadata, about the instrument that will tell me what its spatial resolution is. That also has an error associated with it, and then there are other things that happen. So for instance, when you are measuring environmental things like humidity or temperature, or those sorts of things and you got an instrument that has a footprint say, you know, ten kilometers squared, but you got wind and wind impacts that measurement so it is not necessarily that ten meters turns in and gets warped and so now it shifted, and the instrument is no longer in the center of the footprint. So you are not necessarily measuring evenly around that footprint area, now the group peer has done a lot of search of mathematical models of how you try to understand that footprint, but not everybody has it and in fact that is one of the areas where we work with an intellectual contribution of this people have not taken into consideration. But so those properties we might be interested in being of able of specify the footprint, the actual footprint is different than what the calibration, the calibrated instrument would state.

Student: Would we be keeping track of the changing footprint?

Dr. Pennington: Well, let’s just say that the data could be used that way. I could envision that we could design properties that alerted us when say some wind condition happen, so that we would know that footprint is changing, so yeah I could see well-developed properties that will alert you about changes in your footprint.

Student: Is there only an algorithm that could be able to tell this wind is at this speed or some sort of changes, could you be able to tell footprint from those measurement will we have to manually go out there, so how would you able to do this?

Dr. Pennington: No yeah, you can mathematically calculate it.

Dr. Gates: Any other questions before we move on? I am learning something here. We always look at footprint, upper left, lower right from a satellite. So if you are looking at satellite data, it’s constant. There is a camera that you are looking at.

Dr. Pennington: But even there you have problems because we do satellite collections, well and if you think about a camera, it’s really its spatial accuracy is only correct good at what we call naitor?, the very center of the image and essentially if you go away from that, you are looking the camera sensors are looking at an angle, so that introduces distortion and error and some of them you can correct for that if you know lots of things. But it can also, a satellite image is going through clouds, so the radiation that you are measuring actually was the sensor is being reflected and refracted.

Dr. Gates: Is that part of the metadata? Is precision and accuracy included in metadata?

Dr. Pennington: It’s always included. There is always under lab conditions, sometimes you can calculate it and correct for it but some things like atmospheric correction it is really hard to do.

Student: Just a quick question, is it of any sort of importance to visualize where these sensors are like say in a map or some sort of thing?

Dr. Pennington: Absolutely, yep.

Student: Does the system have that right now, a way to see where the sensors are?

Dr. Pennington: No, it does not. It could be. I mean we know you can generate a map, but the system is not doing that right now.

Student: Every sensor has a fixed location anyway so it is just a matter of left right.

Dr. Pennington: Right. You could certainly put all that on a map and then maybe overlay information about the proper status of the properties or anomalies that have been stated, the data is there to do that, but is not doing it right now.

Dr. Gates: Now we are going to go into sensors and data. So, what kind of sensors will we have?

Dr. Pennington: You know, there is a wide variety that could be used. The commons ones are temperature, precipitation, humidity, soil moisture, and any of those could be measure it, any height above the ground, so you might be interested in ground stuff, you might be interest in ten meters above the ground, but those are the commons, CO2 levels of CO2, really just any chemical property that is out in the environment there is a sensor to measure it.

Dr. Gates: Reflectance data?

Dr. Pennington: Reflectance data absolutely, I mean well, I mean...

Dr. Gates: I am thinking just re-scope it right now and say these are the variables.

Dr. Pennington: You can try to calculate with a satellite use, you could take aerial photographs at different levels, or could be anything from space to ten meters off the ground. Dr. Tweedie has these kites and UABS that the uses, so he measures reflectance across the entire electromagnetic spectrum at all sorts of levels. I guess I maybe should explain what electromagnetic spectrum. So you have…

Dr. Gates: Start creating data dictionary. Hint hint.

Dr. Pennington: So light comes in from the sun it gets split, refracted and reflected into different ways and different wavelengths, it comes in at all sorts of wavelengths, but those things get reflected and refracted in different ways, depending on the composition of things that light hits, so your eye can detect certain wavelengths in that spectrum, but there are other wavelengths that your eye does not detect, both higher and lower, so radar for instance and thermal. Thermal is a radiation, a wavelength your eye doesn’t detect, but you feel it on the form of heat. So there’ s always spectrum. So everything, every material on earth responds differently to those wavelengths, so in satellite images or aerial photographs you can, satellite images in particular, but there are special tools that measure wavelengths beyond what your visible eye can see. So we might take a picture, but if we take a picture we only get a certain set of those wavelengths and with instrumentation we can measure these other wavelengths and those wavelengths can tell us something about what is happening in the environment.

Dr. Gates: Questions? Okay, so we have the types of sensors. Do these sensors have different data formats? Even within the same temperature, there would be different data formats depending on the instrumentation?

Dr. Pennington: Depends on, yeah, the instruments are developed by commercial industries and so each one has their own propriety format that they use. So yes, it can even though when you are measuring temperature, depending on what instrument you are using, you could have a different data format.

Dr. Gates: Is there documentation they could use to look at the formats or is that going be given?

Dr. Pennington: Well, the instrument itself has documentation provided by the vendor. In most cases what you want to do, I mean you will have to decide whether this is an assumption you want to make, but in most cases what you want to do is to take those propriety formats and convert them to something more standard like an ASCII text or some sort of binary format something that you could work with. I think you should assume that you are not going to try to work with every propriety format, that you are gonna convert the data first and do a common format that you could work on. What that common format is could be define by I mean certainly that current tool expects some sort of format, I don’t know what the format is, I didn’t do it.

Dr. Gates: So there’s a header file, right?

Dr. Pennington: Yes there will be a header file.

Dr. Gates: So there’s a header file and then the data in the header file will tell you a lot.

Dr. Pennington: The header file will tell a lot of that.

Dr. Gates: The header file will tell you what is stored, right?

Dr. Pennington: Right.

Dr. Gates: Is that clear everyone? What types of problems should be anticipated when it comes to describing different types of format and different collections of data? That might have been answered.

Dr. Pennington: Say that again.

Dr. Gates: Are there problems that we can anticipate because of these different types of format and different data collection?

Dr. Pennington: Well yeah I mean, I think I just answered that. You don’t want to work with the header G instated. You want to work on some sort of common format that you could work on.

Dr. Gates: So this question: does the system need to be able to define new file formats when new and better sensor instruments come out?

Now, we talked a little about this in an assumption that we are going to make. I don’t know if you want to talk about that.

Dr. Pennington: Well yeah, there’s always new sensors emerging and file formats change and that is true on satellite data, there is a whole history of satellite data and formats and they change. You know, obviously in order to use our system somehow we have to get the data into some sort of common format and that common format I assume is going to be defined by our system. So ultimately yes, there would be a lot of work for somebody to in order to use the system you should get the data into the format of the system expects.

Dr. Gates: So I think that we are going to agree that we will do a simulation. You are not going to be actually working with the actual instrument that is streaming the data. So we will be creating a file. They will be giving you files of the data as though real time [data is] coming in right? We will just simulate that process so we are going to assume that there is going to be a header file that is in a particular standard format. We will just make that assumption.

Dr. Pennington: So if you think about data streaming of all these sensors and you are going to assume that the data that you are getting has already been created and has already been transformed into whatever the common format is that you are gonna be using.

Dr. Gates: So whoever asked this question, whichever team asked this question—it is a good question because it helps you start scoping what the problem is, right? And what the expectations are. That was Team 1.

Okay, number 5, this is on the next page. How is sensor data going to be stored? Is there a database that can be used to make queries, reports? Or is the data kept in flat files on a file server?

Dr. Pennington: This one I sort of puzzled about because I’m not sure whether you are really asking about the sensor data itself or the data that the system we are designing collects about the sensor data. ‘Cause asking about the databases and that sort of things, yeah I mean we are gonna be giving to you this common format, but I don’t know whether it’s a database format or what that is. And your properties you’ll store those in a database or the anomaly information you can store that in a database, I would assume so, but the sensor data itself tends to be flat files as it comes off the sensors, but by the time you get it standardized in the way she is talking about, I don’t know what that would be.

Dr. Gates: So, does it make sense to you all? You’re the developers.

Student: Just to be clear, we’re gonna simulate all the data that we’re getting from the files and all that it’s gonna be real time, so we’re gonna simulate that being real time?

Dr. Gates: We’re going to simulate it being real time.

Student: when the sensors are sending off the data do they go through a third party first? And then you guys pull data from that third party?

Dr. Pennington: No. We have our own algorithms now that we’ve developed to transform the data into certain ways that we can work with it.

Dr. Gates: I’m going to ask Dr. Pennington to give you a scenario that might help you because it’s going to be coming up. Actually coming up in a little bit … you know I’m jumping .... on platforms. On platforms, we talked about web based and we talked about mobile right? And so earlier she gave me a scenario on how that might work-- how someone might use the system. Wo we’re going to go back to how you use the system and I think it might help with that question that you just answered. So we we’re talking about imagining…

Dr. Pennington: Yea well so one scenario that I can imagine, is that I’m out on the field and I get a text on my cell phone that says that there’s a problem or an anomaly with one of the sensors and I would want to be able to walk over to that sensor and try to look at, on my phone, what is the error that is being recorded, maybe look at the property that’s generating the error. Look at the data that’s being generated by the sensor, and maybe I might want to look at those three different kinds of data together, in different ways, um, so I might want to look at the data itself but have a plot…, like let’s say I did a property that said temperature should be between, you know, x and y, during this time period. So I would want to see it in a graph of temperature, threshold, that the property defines, and then I would want the anomaly to be highlighted for me. And then I might want to drill down on the property, I might want to switch to the property, I could look more closely at the property. Or I might want to look at the anomaly more closely. So there’s different things I might want to look at but like I said that there’s those three different kinds of data that I’m interested in. So I want to go over to the sensor, I want to be able to look at those three different kinds of data, maybe in different ways, and I want to be able to fix the problem, if it is a problem, right there or may be something of interest. So I need to be able to interact with all those three different kinds of data in real time, at site, on my mobile device.

Dr. Gates: So this is a use case scenario. So I’m expecting to see this much depth. Any questions?

Student: Just to clarify, you said that the three types of data that you would like to see work on right there right off hand would be the property, the anomaly, and what else I’m sorry?

Dr. Pennington: …and the data itself that’s being generated by the sensor.

Dr. Gates: Now if you were at your desk, what would you like to do?

Dr. Pennington: Same thing. It’s always those three things, now I might want to see them in different ways. I might want to see like a graph of the data where like a certain of that stuff overlaying on it. I might want to see a map view of all of the sensors and being able to maybe hover over a particular location, see something about the data that is being collected at that location. Or I might want to see the raw data. That’s a possibility.

Dr. Gates: So one of the questions that came up is: do you want it to be a stand-alone application or web-based?

Dr. Pennington: I want it to be web-based because I could be in the Arctic and get notified that there’s something wrong with my information coming in the dessert. And I might not have… maybe my cell phone has died and I need to go over to somebody’s camp and use their computer so it needs to be on the web. These things happen, cell phones die.

Student: You keep talking about the graph and displaying stuff do you want any kind of predictions of future data at all or do you just want what we have getting in what we typed, as far as displaying wise?

Dr. Pennington: I think future predictions would be great if we can do that, if we could analyze the data, the patterns, maybe get developed and be able to predict, that would be wonderful. I think that’s an actually, a good use of data properties is you know, we don’t know sometimes what those properties are gonna be, in fact, most cases we’re making that up as we go. What are the properties ‘cause we don’t really know very well. So building the capacity to predict, you know, given the properties and the way that I have it designed, this is what the property would predict, next hour, in two hours, three days. And being able to see, think about, does that really make sense? Does that really um… Because one of the problems that we’re gonna have, we’ve already had it, is that as the scientists start designing their properties, you know those properties are being specified in um, sort of, data base and um and thinking about them from an information scientist point of view. But the scientists don’t always have the right point of view. I guess what I’m trying to say is that if you had a scientist sit down and you asked him to develop some sort of logic, they don’t necessarily have the training to anticipate the implications of what they’ve just specified. So if you have a scientist to sit down and design a property, they don’t necessarily have the background, to anticipate, the implications of what they’ve just specified. So being able to do some sort of predictive thing and say “Ok, you’ve designed this property this way, based on what you said, here’s what we might expect”, would help them in being able to troubleshoot their properties. That was a good question, I hadn’t thought about that before but that would be an outstanding…

Student: Expanding on his question, you mentioned you guys have some expected values and that’s why you’re trying to pick the anomalies. So, adding to that let’s say you have a graph, normally it would just go from one start or beginning to where you had the last reading, so you would put the expected result and highlight them in some way maybe to show this is what’s expected and then around it have it make the comparison between the actual results and the anomalies? Is that what you’re getting at?

Dr. Pennington: Yea. ‘cause sometimes, ok so one thing that you might want to do is compare across sensors, so I don’t want to just talk about well temperature or maybe … so for instance precipitation, there are certain characteristics, patterns and habits of temperature and humidity and barometric pressure, that we know about, and that we could specify in a property, to get at, is it raining or is it not raining? But there may be other things, comparisons that we try to do… that weren’t so well known. So as we start to develop more complex properties, yea it’d be nice to be able to say now that…. And It’s also just and error-checking mechanism when you develop a property. Here’s the way that you developed your property, Now here’s the way, here’s a simulation of what we might expect. Did that answer your question?

Student: Just an idea I had so, when you expect, obviously, what you’re real data is going end up being, so let’s say uh that the real data from what you expect is constantly being more and more off each time so you have that so and so recording so you know what these expectors have been this off and have a notification for you to see so you could maybe readjust whatever algorithm that you have used..

Dr. Pennington: And I can somehow envision that you start reading these anomalies, report it, you find a pattern in the anomalies that indicates that your specification is wrong, it’s not that there’s any error or that there’s anything interesting happening. It’s that your specification is wrong.

Student: I actually had a two part question. Um, you keep talking about the properties but the way that you’re kind of explaining them is making me wonder, are these ever expanding properties that people can define for the database?

Dr. Pennington: Yes.

Student: And then the second question about that, you were saying that we have to be able to go back and actually have them specify who was the author at that time and who was the current author or whoever is the care taker for that data. So, my question is on that part would they have to have the database widely available in a database format that everyone can change and in real time view the changes?

Dr. Pennington: Maybe. I think for now it’s more local. That I could envision that this all worked out well it became widely adopted yeah there might be some sort of community registry, you know, I assume it would be a database, you know, that’s for you guys to decide not for me to specify. I don’t know how you would do it, but I could imagine there would be some sort of community registry where everybody in the community is sharing and looking at these properties.

Dr. Gates: So at this point, you always want to get the “what” and not the “how.” But, having said that, what Dr. Pennington is saying is that we have to store them locally so like within this group-- within her research group--they want to keep the properties stored. Ok. So they know they need to be stored in a repository, whether that’s Excel or an SQL Database or whatever. That’s part of your analysis which you’ll do later. What she’s saying is that you’re collecting these properties and this metadata. And then there may be a point in time when we start saying ok we want to now be able to move those into a more national type of database, which the community would be building, right? That says ok-- now I’m going to take these [properties] and move them up for a broader view. So if we’re thinking about designing for change, which is going be one of the principles you learn about, we’re going to scope it, but you’re always thinking about [change]. And I said this on the first day of class, if you remember, that you start thinking about in the future we may want to do that. We want to be able to think about that because it impacts your design. So it’s a good question.

Student: Well let me also redefine something as well, for the updates of the information, do those have to be real time updates?

Dr. Pennington: Ok, which information are you talking about?

Student: The properties, yes.

Dr. Pennington: I um…..

Dr. Gates: Can I help with that maybe?

Dr. Pennington: Sure

Dr. Gates: If I’m thinking that the properties are stored someplace-- I’m a researcher and I’ve set up a sensor. So here’s a scenario: I could be grabbing them [properties] and putting them in there [i.e., in the system for monitoring sensors]. If I’m grabbing and reusing somebody else’s [property], you know you want to know that this [property] is being monitored and it’s reused. If I change it [the property] and change the parameters, then I think that’s what Dr. Pennington was talking about. We want to then store that information. I reused it, but I change the parameters.

Dr. Pennington: Yea, and you don’t want to overwrite, because it may have changed, you know, this property the way it was developed, worked real well for that camera right there. Now I’m going somewhere else, maybe way far away, and I liked that property, but it’s not quite right for this new location. So I’m gonna build on this property and I’m gonna use it here. Now, person three a year later comes in, and they’re looking at these properties and one of the things they want to be able to do is say ok, they used it this way here, and they used it this was over here. Or maybe there’s some comments about how it works and why it was changed and why that modification was made. And I can learn from that. So now I’m gonna use it someplace else, and I can say ok, here’s how it worked over here and here’s how it worked over here. Here’s how I might expected to work in this place that I can modify it accordingly. So it’s important that I know that whole trace of where that property has been and how it’s been used. So we don’t want to lose any subversion. You can think of it as versioning and it’s important to keep all that. There’s one other thing, something that you said triggered something with me that I wanted to say while I was thinking about it. I mentioned that the biggest problem with the interface right now it’s that it’s in the language of computer science. I’m flippantly using this word property. The very first time I met with Dr. Gates, about this project, that was my first question, what is a property, what are you talking about? So I’m using the word property easily because we’ve been working on this now for two years. But that’s not a term that the scientists would call this thing of properties, they’re thinking about properties of the thing that you’re measuring and you guys are talking about data properties. You know, the data properties are related to the properties of the thing on the ground. But I had to clarify that, I didn’t know, what she was talking about.

Dr. Gates: A data dictionary moment.

Student: What do you mean by data properties, are those like rules or how would you define them?

Dr. Pennington: How would I define data properties?

Student: yeah, are they the rules that you when once you go into the sensor and you pull in that information, is data property the rule you set up to be able to find what’s useful in that information, or how is that set up?

Dr. Pennington: I don’t understand your question, so I’ll give you an example to see if that clarifies for you. I have a sensor I’m measuring temperature of something the air or the ground or something, the temperature, the actual temperature measurement is a property of what ever it is I’m measuring, the entity I’m measuring. So I have the measurement and the actual entity and it’s a property of the entity. But when I’m doing data properties, the way were talking about it, I’m going to say during this time period at this place, temperature should range from 10 to 30. So now your talking about the property of the data stream, not the property of the thing I’m measuring. That’s a subtle difference but its one that’s not immediately obvious to the scientist when your talking about it. Did that help at all.

Student: So for that you’re talking more data constraints versus data actual data of information

Dr. Pennington: Right, yeah, yeah

Dr. Gates: So the anomalies come. Its like predicting what you think that the data--how it would be behaving, or what the measurements are in relationship to [the data].

Dr. Pennington: And its comparing the data, you know you’re you’re the data, the data is now the entity of interest, as apposed to the thing your measuring being entity of interest. A subtle but important distinction.

Student: Going back to what Adam was saying earlier, about trends or anything else you want to be able to predict that this just constraints all in properties to constrains that mention anomalies.

Dr. Pennington: um yeah, yeah, yeah

Student: So when defining, the user are able to define the data properties

Dr. Pennington: the users are the ones defining the data properties, yes

Student: and when defining them you have to take the weather into consideration?

Dr. Pennington: Something is not right about your question. The weather is what you’re measuring, in most cases. You’re measuring temperature, precipitation, and wind velocity. Those are what your measuring, and collecting data about. So the properties are about what you expect.

Dr. Gates: What you expected, how you expect weather to behave. And when you see a spike, you want to see when that spike occurs-- maybe the temperature raises, and you’re not expecting that. So its April and you don’t except to have a snow storm, and when that happens you want to be able to capture that moment because it may be telling you something.

Dr. Pennington: it may look like a snow storm, and maybe it is a snow storm, or maybe one of the instruments has gone haywire.

Student: but if its not defined in the properties, within the limits, then it’s considered an anomaly.

Dr. Pennington: right, right

Student: so this might be on the scope of design, but do you think it would be helpful in that case to be able to synchronize weather data?

Dr. Pennington: that’s a good idea, if you had some other source of information that you could use to [track]

Dr. Gates: That’s number 7 on the second page. The question is will we have access to check for predefined values, historical values, and time constraints.

Dr. Pennington: yeah so you could imagine, instead of me specifying the temperature range that I expect from a particular day maybe I could access the weather and climate data base and just say compare this value to that data source over there. I hadn’t thought about that, but yeah that would be helpful.

Student: I want to ask something on performance in terms of speed, for example would it be preferred to just have something that’s not going to be very flashy, but gets you the very essentials in a very simplistic manner, just quickly or is it ok if it takes those extra seconds more but gives you a big a more, not flashy but more precise range of data, like what’s more important when your viewing this, that you get a lot even though it takes some time?

Dr. Pennington: well I think its always a good design to provide the things that you cant provide fast as fast as fast you can, and you crank all the others in the background. That’s always a design decision. Nobody likes to wait, for anything. Some things take time.

Student: I’m just asking how imperative, if its just a slight annoyance

Dr. Gates: How important is it? Is performance important?

Dr. Pennington: Performance is always important. But, if I’m looking at trade off.

Dr. Gates: He’s looking at a trade off, and I think that’s part of analysis. [Note: what follows should be ignored--When you look at analysis and then you start saying I’m going to add these bells and whistles. That’s part of the analysis part.]

Dr. Pennington: I mean it would be easier if you came with a real concrete sort of would you rather have us provide this information right up front, and this is five minutes later, or we can provide this in two minutes if I don’t do this. If you give me something concrete like that its an easier choice to make.

Dr. Gates: I’m going to jump to question 10. If an anomaly has occurred because of a faulty sensor or a sensor needs to be recalibrated, how should that be reported? I think you kind of talked a little bit about anomalies. Is there anything else you want to add?

Dr. Pennington: I think/guess you need provide some choices to the user. In some cases you they may want instant notification on their mobile. Its sort of, I donno if you guys work with the airlines at all. But they have all these choices about if the flight is delayed do you want to be notified an hour in advance or do you want to know the day before or do you want to know by phone or by email. When I comes to contact, that’s a very individual sort of choice so you need to provide those choices.

Student: I need to go back to the data property thing a bit, but from what I’ve read I thought the data property also sort of do calculations between other sensors.

Dr. Pennington: could be, yeah might be. Part of understanding the failure of a sensor is knowing how it is supposed to be. If this sensor is showing lower and lower and lower values, then maybe this sensor is supposed to be showing higher and higher and higher, there correlated. Maybe inversely correlated or they maybe regular. But there’s a correlation that were expecting, if the readings are not correlated like they’re supposed to be then that could be an indication that there is something wrong with the sensor. So you may be, almost certainly in many instances will be comparing across different sensors.

Dr. Gates: that’s a good question and we’ll be digging into that more on Tuesday.

Student: so that’s a capability that’s already in place with the..

Dr. Gates: Yes

Student: I have a question that kind of builds on what he was talking about. How be able to distinguish between what would be an anomaly and what would be considered a faulty sensor?

Dr. Pennington: I mean that would have to be the scientist would do the analysis on that. I think it’s going to be hard to tell just from the data. Unless the data flat lines, then you know that’s a problem with the sensor. But if it’s a calibration problem that’s much harder to detect, I think in most unless it’s just a flat out failure its going to take some analysis from the scientist to try to understand that.

Dr. Gates: Just to give you a scenario. There’s so much information that streaming. Depending if you’re doing this, if it’s on constantly and part of the tool. It is just to help the scientist to look at those places where unusual things are happening and then to be able to analyze it. So it’s not sophisticated enough at this time to be able to say with certainty that this is a problem that has occurred.

Dr. Pennington: Just think about if I’m working with satellite data, I might have Petta-Bytes of satellite data coming in on a weekly bases, I just want something to call my attention to something I need to look at.

Dr. Gates: And that’s the purpose here.

Dr. Gates: Ok. We’re going to go to data analysis real quick. We have about ten minutes. Are there types of data analysis needed other than trends and predictions of future data? I think we talked a little bit about that. I don’t know if there is anything else to add.

Dr. Pennington: I would say that at least initially we want to keep the analysis pretty simple. I mean I think that one of the uses I see with this is that as we start developing that, you know when we can get, as data starts coming in and we start analyzing how they are in this way, how they are responding across sensors and though time and through space. We can use data properties as a way to call our attention to patterns that are accruing. So we can analyze the data that we collect about anomalies, we can analyze those in a way to help us better understand what we’re seeing. So that sort of analysis I see happing down the road, but I don’t see it in this system right now. Unless something, you develop a different system to do that analysis. I think maybe just trends and the simple things is the place to start.

Dr. Gates: There is one about colors, but I think that’s getting into the design.

Dr. Pennington: colors are important. If I’m looking at precipitation I want it to be blue, it don’t want it to be red, or black. Color is always important.

Dr. Gates: Do you want to be able to download any of the graphs?

Dr. Pennington: Yeah absolutely

Dr. Gates: And you want to display the graphs in real time or near real time?

Dr. Pennington: yeah, I want to be able to look at the graphs at real time, and if I like something I want to be able to download it and print it out, and put it in a publication.

Dr. Gates: What operating system would you like it to run on?

Dr. Pennington: all of them

Dr. Gates: And mobile?

Dr. Pennington: all of them, scientist are like everybody else, we have our favorite tools and we don’t want to switch just to use a particular piece of software.

Dr. Gates: Do you know of other systems that perform similarly to this?

Dr. Pennington: nope, not like this

Dr. Gates: So there are entrepreneurship opportunities?

Dr. Pennington: yes, entrepreneurship opportunities.

Dr. Gates: Other topics that you think are important that you think we need to discuss?

Dr. Pennington: No, not at the moment.

Dr. Gates: I’ll open it up for other questions

Student: Just to clarify, so are we showing/displaying anomalies compared to the data properties, I guess comparing the data properties to the actually data. And to see if there is a spike or anything, and once we have the anomalies, the scientists determine if those anomalies are errors or actual.

Dr. Pennington: yes, absolutely. And so it would be nice when you build this storage, however you store the anomaly information. It might be nice to include in that a way for the scientist to flag it, once they look at it and decide whether it was an error or whether it was not an error. To flag it as what it was. I think that would be important information to collect.

Student: Can you kind of explain again what a foot print is?

Dr. Pennington: let me draw a picture that might be the easiest thing. So I have some sort of instrument here that’s collecting information. So maybe its located right here on this tower, so here is the ground. Now the instrument itself may say that it has a special foot print will call this, lets say this is 1 km. I don’t know what it is, that’s an awful big footprint, well we will leave it. So that’s what it says, then meta-data will say it has a 1km resolution. But maybe what your measuring is. Based on this what it would say is ok this instrument is collecting information, around this area right here, based on what the metadata says. What if there is a strong wind coming [from the left], then all these particles that you’re measuring, are being blown. So essentially what ends up happening is you end up collecting information here [to the right of tower]. Does that make sense, does that help?

Student: Skipping back to the information being able to be shared, how far do you want it to be as far as social? Do you want to have facebook or twitter [integrated into the system], or all in house within that browser and keep all the information inside there?

Dr. Pennington: Well, its an interesting question. Most scientists detest Facebook and Twitter. Primarily because they don’t see a point to it, they don’t want to chit chat with each other. They don’t want to know what you’re doing right now. They’ve got their own problems to deal with, they don’t care what youre doing right now. Unless its’ in the context of the work they’re trying to do. Having said that, if you think of a use case where it would be important for them to have some sort of social networking then it is important to include it. If it’s just so that I can be regularly informed what you’re eating, I don’t want it. So if you were going to put something like that in, I think it would be important to scope it in a way that it directly ties to the work [they’re] trying to get done.

Student: That’s what I meant, the facebook and all those networking sites is just an idea of how social you want to be within that system. If you want to keep it within that system, do you want to have it set up, so that when those scientists share it to each other, all of them see it, and have that kind of hang out and discussion within that browser itself, instead of trying to bring in facebook.

Dr. Pennington: Could be, yeah. I mean, I wouldn’t put it on facebook. But that kind of thing, like if im in the field and something happens, I might want to notify other people in my research group what’s going on. And I might want some discussion about it, yeah.

Student: Does the system have some way to tell, [for example] if the weather adjusted the usual boundaries of the sensor?

Dr. Pennington: The reason this would become important is when I’m designing my properties. I might want to say wind: if the wind is under 10 miles per hour, then I expect this other data to be in this range. On the other hand, if the wind is over 30 miles per hours, then I might expect something different. So its all about articulating what your expectations are in a predefined way. But like I said in most cases we don’t really know what to expect so part of this is going to be using this process to help us understand how things happen.

Student: How would you want the system or users to communicate through our system. For example, if you detect an anomaly, do you want it to throw alerts, send out emails, send out text messages?

Dr. Pennington: I think you need to provide all those possibilities, and let each user specify how they want to be contacted and when, and in what way.

Student: Would you want any of this preliminary analysis available to the general public in some sort of way, or is this all confined to just the scientists.

Dr. Pennington: There is an interesting move towards what they’re calling citizen scientists, where you try to involve people who are non-scientists in your data collection efforts. I can imagine, somewhere down the road, somebody is going to come up with some clever way to involve other people in this process. It could end up being public. Here’s a scenario, maybe I set up sensors down in the plaza in downtown El Paso and I’m measuring something and there’s some sort of anomaly and I want to engage whoever is in the vicinity to go down there and check it out for me.

Dr. Gates: Sounds like a proposal. That’s a good question.

Dr. Pennington: Citizen scientists is a really hot topic because data collection takes a lot of time and effort and if you can get people that are interested in whatever it is that you’re studying, to help you with it then [great].

Dr. Gates: Good questions. Any other questions? So let’s thank our speaker. I want to thank everyone.

Student: I have one last question. I know it said in the project overview that we would be provided with the source code, but as far as the interface, but I was wondering if we would be able to see it, or have someone show us how to use it?

Dr. Pennington: I think that’s [up to] Dr. Gates.

Dr. Gates: I’ve been thinking about that and what I’m worried about is what Dr. Pennington talked about earlier. She really wants it to be science centric. I was going to show you the interface before, but I’m worried that it’s going to take away the creativity. So what I’m going to do is work with Dr. Salamah to provide the information in a more general way that will help you then come up with a new interface and language for presenting it to the scientists. That’s going to be a big part of the analysis that we do. I want to thank everyone.