Taped Interview of Dr. Pennington

Conducted by Dr. Gates and Software Engineering Students

February 19, 2013

**Dr. Gates**

Dr. Pennington has a lot of expertise environmental science shes a ontologist and a geologist by training and she also wouldn't mind interdisciplinary research so working across different disciplined and bringing them together. Which is really important for this particularly project. So i am going to go ahead and I have given her the questions that have been put together. I also am asking her permission to record this interview I want to make and am going to make sure and assure that only the class has access to the interview. No access outside of the classroom and so in asking her to do this, and you saw how she reacted to making it public this is the reason why we have to let the customer know that you are recording. Okay, so I am going to start off with some general questions. We talk about, what do you see as the benefit for the system you propose?

**Dr. Pennington**

We spent a lot money and time putting these sensors and their tools huge investment for us designing the system and putting the equipment out there and its there for a reason we want to know that the data is going to be useful when we do the analysis. An important part of that is making sure that the data that is collected it has a certain quality, a quality we can use, so quality is huge sensors network there are failures things that aren’t calibrated correctly, there are many things that can go wrong the other thing that can happen 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. And for some people an event happens out there might be unusual might be interesting to them. You may want to do something to respond to that event in some way, but we can’t if we don't know when it happens. And so we are looking for two things in really important but we also need to know when some interesting event happens and might be able to see interesting to us

**Dr. Gates**

What are some of the limitations of the current data specification tool?

**Dr. Pennington**

The biggest one is that it is written and design and formulated for a computer scientist. It got a language of computer science, its got the interface that is relevant to computer science, talks about queries and codes and rules that's not the language of science. So its does what it needs to do but it is not useable by a scientist.

**Dr. Gates**

Will there be different types of users? If so, are there restrictions to particular users?

**Dr. Pennington**

Well half the user will likely have will either be graduate students, some faculty in some cases there will be somebody designated as a data manager. And that person might or might not have a computer science background usually not they might have some type of informal training with information technology. So, they might be completely new and not have a clue on how to use the tool. There may not be any training available, there may not be anybody available to ensure the time, On the other hand it may be somebody who is doing it on a daily basis that is very familiar with it. And we have to design for all the satisfaction. So I think the reason that becomes important is the user interface you provide for somebody who is new and learning might be tedious for somebody who is using it all the time and is familiar with it. So that has to be taken into consideration. In terms of restrictions there is usually one person who responsible for the data and they should be able to do whatever they need to with it. On the other hand you may not want somebody that is using it you don’t want them to be able to modify the rules and such or make comments about it, you don’t want them to actually do anything.

**Dr. Gates**

If there’s a follow up question I am going to allow it.

**Student**

As far as users do you want somebody to be able to maintain the system. As far as publicly and then somebody to change the user like an administrator of sorts

**Dr. Pennington**

Yeah I imagine that there is somebody who be administrating the system.

**Dr. Gates**

Do users require authentication.

**Dr. Pennington**

Yes, there was a question in here about it being utep. Its not necessarly using that validation. But you do need to validate your users.

**Student**

Is there some kind of help session, or some kind of wizard to help them get started.

**Dr. Pennington**

Well unless you can design in such a way that is so intuitive that people can figure it out without any kind of help. You will probably need some sort of wizard or something to get them started. I hope you can design it so that it’s not needed that would be great.

**Dr. Gates**

When sharing data between scientists what type of personal information will be displayed in the documentation?

**Dr. Pennington**

Well certainly we want to know who generated the data, who collected, who is the original source. Contact information about them if there are any question about the data, you may want to know the institution they are at or at least the institution they were at when the data was collected because people move around. If there are because people move around and change projects we need to know who is the original contact but also who is the current contact. Maybe just get their name, their current position, institution, and some sort of contact information.

**Dr. Gates**

So I am going to ask you about something they know, we start talking about data properties.

**Dr. Pennington**

I have to say I was got confuse reading thru the question, cause there are three different kinds of data that we are interested 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, there’s the data about the properties that you develop and design, the properties themselves are a kind of data, and there’s the anomalies that are detected. So when you ask questions about the data I need you to be very specific about which of these three kinds of data you are talking about. So can you repeat the question so that I can know what kind of data we are talking about.

**Dr. Gates**

This was ask by. Which team ask that question? When sharing data between scientists what type of personal information will be displayed in the documentation?

**Dr. Pennington**

We already answer that one. You said something about properties

**Dr. Gates**

Oh I added you that. You were talking about sensors I added question I was going back to your three types of data I am glad you clarified that. So same question but now a bit about the data properties

**Dr. Pennington**

So now the data about the properties. Same thing we want to know who created the properties originally and if there is something that has changed that somebody else is responsible for that property we want to know who that person is, point of contact about the property but we also want to who the original owner was. And if its been modified you know it may be a relation history on the property and it might be different people involved with that so that whole providence of that property is something that I would be interested in.

**Dr. Gates**

As well as location of where it is.

**Dr. Pennington**

Location is huge. So let’s look again I am going to give you an example if I am collecting data out here in the desert the properties that I design about temperature, precipitation, what I expect in terms of precipitation. Are quite different there than what I might expect if I had the same tools in the artic. So properties have to be specific to a location, and a time period.

**Student**

Contact information should be available for any user to see or certain users.

**Dr. Pennington**

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

**Dr. Gates**

So we talked about location, any follow-up questions?

**Student**

About how many locations do you guys have?

**Dr. Pennington**

Right now we use these particular, well it depends on who you mean by you guys. Our particular research group.

**Student**

~~~ how many locations are going to be manage by the system

**Dr. Pennington**

So it’s a new system so right now it’s just us that are using it. The sensors themselves are use all over the place be all source of different people. But this particular property specification tool is develop in house. We would like it to be adopted elsewhere. So you can say that in the long run we would like to see it used world wide. But right now its used in the desert out here and in the Antarctic. So those two extremes desert and artic.

**Dr. Gates**

So I am going to queue you a little bit. Only this time, when we talked about location is it clear to you how you are going to record it.

**Dr. Pennington**

Well I think it will be important to know, I mean think location is a complicated subject. So lets say the camera is a sensor its got a specific location. But the sensor itself is recording not just in that location its got a footprint that its measuring. And that changes from sensor to sensor. So its important to know that designing a property for a sensor, what is the footprint that you’re measuring and the property applies to. Same thing is true with time. You may measure something in a specific point in time but sometimes what you are after is really a time frame a window of time. So I might look for a property that is interested not with what is happing in this point in time but how does it compare to prior time or future time.

**Dr. Gates**

So you look at time, whenever you look at anything or work on a system that has some sort of decrement you really need to understand that measurement and to what degree of accuracy that you want

**Dr. Pennington**

So generally we talked about resolution. How is it involving a very fine space and time interval or is it involving the sensor has a resolution involving a broader space and time. If you think about it, that’s sort of what pixels, you give it resolutions. You have to know what the resolution is on the instrument.

**Dr. Gates**

So, its instrument dependent.

**Dr. Pennington**

Is instrument dependent.

**Dr. Gates**

The regularity of the micro seconds?

**Dr. Pennington**

If you mind it could be might be microseconds. It could be measuring every ten minutes. It could be measuring once per day, with some interval associated.

**Dr. Gates**

So, when you talk about footprint, what would you all like to know about footprint is? No view, not a catch? You gotta dig.

**Student**

For each different sensor, would we be able to access from the individual sensor data? How big is the footprint according to the data of the actual, ah, I guess area that is recording? Or do we actually have to go out and find ourselves?

**Dr. Pennington**

Well so there’s two different questions there. One is that there is an inherit resolution set, that is determined by the instrument itself, and there’s also things that the scientist can modify. So an instrument could be capable of measuring every microsecond. But I might only have it measuring one an hour. And, so there’s two different footprints.

**Dr. Gates**

How do you know the location?

**Dr. Pennington**

Well there, you know the location, again, I mean let’s go back to the camera again. I will have a sort of GPS reading about the location, and there is an error associated with that GPS reading. So I will really know where, and the error accuracy. Then I also know because of the …there’s gonna be nformation, data about the instrument. And what it’s special resolution is. Uhm, that also has an error associated with it. And other things that happen, so for instance when you are measuring wnviromenal things like humidity or temperature, those sort of things, and you have a footprint of, let’s say 10 kilometers. But you got wind, and wind impacts that measurement. So its not necesarely those 10 meters. And now it is shifted and it is no longer on the center of the footprint. So, oh you are not necessarly measuring evenly around that footprint area. Now, our peer, has done a lot of research, we got mathematical models, and has trying to understand the footprints. But then not everybody has that, in fact that is one of the areas where we doesn’t able the contribution of this group. Most people are not taking that into consideration. And so that, properties that we might be interested and be able to specify which footprints are different then what the calibration states.

**Student**

Will we be keeping track of the changing footprints?

**Dr. Pennington**

Ah, well let’s just say that today could be used that way. I could envision that we could design properties that could alert us, when let’s say some wind condition happened, so we know footprints are changing. So yeah I can see our groups developing properties that will alert you about changes. You look like you had a question.

**Student**

So there any algorithm that could place and check certain speed, humidity? Or an integer changes. Will you be able to tell a footprint about those measurements or do we have to go there personally?

**Dr. Pennington**

No, you can mathematically.

**Dr. Gates**

Any other questions before we move on? Sunlight, daylight… If you think about it, the camera its only facial accuracy is only correct good at what we call natal. Which is the very center of the image. As soon as you go away from it, you are looking and the camera sensors are looking in an angle which introduces distortion and error and someone can even correct for that if you know lots of things. But it can also, a satellite image is going through clouds and with the radiation is actually, refracted. So it’s part of the metadata? It preccion and accuracy are of the metadata?

**Dr. Pennington**

Sometimes you can calculate in correct form.

**Student**

Just a quick question, is it of any sort of importance to visualize where the sensor are like in a map.

**Dr. Pennington**

Absolutely.

**Student**

Does the system do that right now?

**Dr. Gates**

No, no. It doesn’t. 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 its fixed location anywhere so you can just grab it and get it.

**Dr. Pennington**

You could certainly put it in a map and maybe overlay information about anomalies that are being stated. Is there to do that, but it’s not dong that right now.

**Dr. Gates**

Ok so now am gonna go and mention the sensor’s data. So what kind of sensors is being used?

**Dr. Pennington**

Well you know there is a wide variety that we use. Temperatre, Moisture, and any of those can me measured at any height above the ground. In case you are intersed in that, but those are the common ones. C02, levels of CO2 in the air. And just any chemical, property that is out in the environment, there is a sensor to measure it. Absolutely. I’m thinking just scoping right now and the easier it was, And we can calculate with satellite and the area photographs for oil, of anything for space…

**Dr. Gates**

I’m thinking to scoop more

**Dr. Pennington**

So he measure in all sorts of levels…So you have…

**Dr. Gates**

They are creating a data dictionary.

**Dr. Pennington**

So uhm, light comes in from the sun. It gets split, and it has different kinds of wavelengths. It comes in all sorts of ways. It reflects in different ways depending on the composition of the s… it did. So you are all rightif you have wavelngths. Radars are important. Radars for example. In particular, but special tool sallow to see. So we might take a picture, we only get certain, the sides go away. With this notation we can measure these other wavelengths.

**Dr. Gates**

Questions? So we have the data sensors. Are the data formats, do these sensors have. Different data formats, let’s say within,temperature…?

**Dr. Pennington**

Yes

**Dr. Gates**

Is there documentation that you can see to use this format? Will that be given to the teams?

**Dr. Pennington**

Usually, there is information provider by a vendor. It depends on the instrument. And this is a commercial industry where each has…So there can be several data formats. In most cases what you want to do is take those proprietary formats and turn them into something more standard. Like some sort of binary format. I think you should assume that this is a tool that you are not gonna try more than one for every. A certain format is expected.

**Dr. Gates**

There is a header file. You will know more about that. What is stored. What types of problems should we anticipate when the system describes several types of formats. Are there any problems we can anticipate due to that.

**Dr. Pennington**

I think I just answered. I mean yeah if you want to get it in a certain type of format.

**Dr. Gates**

And so this question: Does the system should be able to define new file anomalies. Now we talk a little bit and about the assumption that we are gonna make.

**Dr. Pennington**

I mean there always new sensors and they change. The history of satellite data format is in constant change. And obviously in order to use our system, we need to kick in the data in some sort of common format I assume is going to be fine by our system. So ultimately, yes there will be a lot of work for somebody to be able to use this system to get expertise at the system.

**Dr. Gates**

Now we are going to agree that we are going to do a simulation. You are not gonna be working with the actual instrument for immediate data, so we will be creating files of the data, we will be giving you files of the data as though it was real time. We’ll just simulate that process, so we can assume there will be a header file. That is the particular standard format. We’ll just make that assumption.

**Dr. Pennington**

So if you think about the data streaming in from all the sensors, you’re going to assume the data has already been transformed into what the common format is

**Dr. Gates**

So whatever team asked this question, it’s a good question because it helps you to start scoping out the problem. How will sensor data be stored, is there a database that can be used to create queries, reports or is the data kept on files.

**Dr. Pennington**

So this is where I’m not sure if you’re really asking about sensor data itself or the data the they system you’re designing collects about the sensor data. They’ll be a common format; I don’t know what that is. The sensor data itself seems to be flak files

**Dr. Gates**

Does this make sense to you all? You’re the developers

**Student**

We’re going to simulate all the data that we’re getting from the files in real time. So we’re going to simulate that in real time?

**Dr. Gates**

We’re going to simulate that

**Student**

When the sensors send out the data, do they go to a third party first and then you guys pull the data from that third party

**Dr. Pennington**

No, we have our own algorithms now that we developed that can transform the data

**Dr. Gates**

I’m gonna ask Dr. Pennington to kind of give you a scenario that might help you because it’s gonna be coming up a little bit, about … So platforms we talked about web based and we talked about mobile. So earlier she gave me scenario on how that might work or how someone might use the system. So we’re gonna go back to how to use the system.

**Dr. Pennington**

One scenario that I could imagine is that I go out to the field and I get a text on my cell phone that says that there is an anomaly with one of the sensors and I would be able to walk over to that sensor and try and look on my phone what is there that is being reported, maybe look at the property that is generating the error, look at the data that is being generated by the sensor and maybe I’ll wanna look at three different types of data together in different ways. So maybe I’ll wanna look at the data itself, like let’s say I did a property that says temperature should be between X and Y during this time period, I would wanna see a graph of the temperature, threshold the property defines and then I would want the anomaly to be highlighted for me. Maybe I’ll wanna switch between a property.. or I might wanna look at the anomaly more. So there are different things I might want to look at. There are 3 different types of data that I’m interested in. So I want to go over to the sensor, I want to be able to look at the 3 different types of data maybe in different ways and I want to be able to fix the problem right there. So I need to be able to interact with all those 3 different types of data in real time at the site on my mobile device.

**Dr. Gates**

This is a Use Case Scenario

**Student**

Just to clarify the 3 different types of data that you would like to see are the property the anomaly and what else?

**Dr. Pennington**

The data itself being generated by the sensor

**Dr. Gates**

Now if you were at your desk, what would you like to see?

**Dr. Pennington**

The same thing, it’s the same 3 things. I might want to see them in a different way. I might want to see a graph; I might want to see a map view of all the sensors like to be able to hover over a particular location, see something about the data that is being collected, or I might want to see the raw data.

**Dr. Gates**

One of the questions that came up was Do you want it to be a standalone application or web base?

**Dr. Pennington**

I want it to be web based because I could be in the artic and I could get notified that there is something wrong with my info in the desert and maybe my phone has died and I might have to go over to someone’s camp and use their computer

**Student**

Do you want any kind of predictions on future data or do you just want what we have getting in as far as display wise.

**Dr. Pennington**

I think future predictions would be great. If you could analyze the data patterns that maybe are being developed and be able to predict that would be wonderful. That would be a good use of data properties; we don’t know sometimes what those properties are going to. In most cases we are making that up as we go what are the properties So building the capacity to predict given the properties in the way that I have it designed, next hour 2 hours, 3 days and to be able to think about does that really make sense? One of the problems we have is that Scientists are designing their properties and being specified in thinking about the information science point of view. If you have a scientist and you ask them to do develop some sort of logic, They don’t have the training to anticipate the implications of what they just specified. So being able to do some sort of prediction, would help them in being able to troubleshoot their properties.

**Student**

You mentioned that you guys have some expected values, ?????

**Dr. Pennington**

Yeah, one thing you might want to do is compare cross sensors. So for instance precipitation there are certain things that happen in temp. and humidity and barometric pressure that we know about and that we can specify in a property. There may be other things that we try to do that we don’t. So when we start to develop more complex properties, it be nice to be able to say, and it ???? Here’s the way you developed your property and here is a simulation on what you might expect

**Student**

???

**Dr. Pennington**

Yeah, like a sort of a revision that you might find. Start getting these anomalies reported, you find a pattern to the anomalies and you find out the specification is wrong.

**Student**

You keep talking about properties but the way you keep explain them make me wonder; are these ever expanding properties that we can define for the database?

**Dr. Pennington**

Yes

**Student**

Ok and the 2nd question is; You keep saying we have to have them go back and actually have them specify who is the author and caretaker 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 in real time and view the changes

**Dr. Pennington**

So I think maybe, I think for now it’s more local but I can envision that if it came out well and it became widely adopted. I assume there would be a database, but that is for you guys to decide. I can imagine a community registry where everyone is looking at the property

**Dr. Gates**

At this point we are looking at the what and not the how, but having said that what Dr. Pennington is saying is we have to locally, within her research group, they want to keep properties stored. So we know that we have to keep them in a repository, whether it’s excel or SQL database, that’s part of your analysis. What she’s saying is that you’re collecting these properties and metadata and then there might be a point in time where we start saying we’ll be able to move them to a national database that says we’ll take them into a broader view. We’re thinking about designing for change

**Student**

For the updates of the info do those have to be real time updates

**Dr. Pennington**

Which info you talking about properties?

**Student**

Yes

**Dr. Pennington**

um??

**Dr. Gates**

Can I help with that question? I’m thinking is that the properties are stored someplace, I’m a researcher and I set up a sensor, I could be grabbing them and putting them in there. And if I’m grabbing and using somebody else’s you want to know that this is being monitored but it’s reused. I might go and make changes and change the parameters then that’s what Dr. Pennington was talking about. You want to store that info. I reused it but I changed the parameters

**Dr. Pennington**

Yeah you don’t want to override. This property is built for that camera right there. Now I’m going somewhere far away and I like that property but it’s not quite right for this new location. So I’m going to build on this property. So person 3 comes in a year later and is looking at these properties and the want to be able to do is they used it this way here, and maybe there are some comments on what worked and what changed and I can learn from that. So now I’m going to use it someplace else.

**Dr. Gates**

I’m ah I’m ah researcher I’ve set up a sensor I may be I may be, so here’s a scenario I could be grabbing them and putting them in there and if I’m grabbing and reusing somebody else’s you wanna know that this is being monitored but its reused I may go and change it and change the parameters. Then I think that’s what Dr. Pennington was talking about you want to then store that information I reuse it but I changed it.

**Dr. Pennington**

Yeah and you don’t want to override because, this property and the way it was developed worked well for that camera right there and now I’m going somewhere, else maybe way far away and I liked that property but its not quite right for this new location. So I’m gonna build on this property I’m gonna use it here, now a person three years later comes in and they’re looking at these properties and they wanna be able, one of the things they wanna do is be able to is say ‘ok they used it this way here and they used it this way over here.’ Or maybe there are some comments about how it worked and why it was changed and and why that modification was made, and I can learn from that and so now I’m gonna use this someplace else so I can say ‘ok well so here’s how it worked here and here’s how it worked over here and here’s how I might expect it to work in this place,’ and I can modify it so its important that I know the whole trace of where that property has been and how its been used. Um so we don’t lose any version so you can think of it as versioning umm and its important to keep all of that, and there’s one other thing, something that you said triggered something that I wanted to say while I’m thinking of it. Umm I mentioned that the biggest problem with the interface right now is that its in the language of computer science, you know I’m flippiedly using the 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 scientist the scientist think properties they’re thinking of the proprieties of the thing they’re measuring, and you guys are talking data properties and those the data properties are related to the properties of the thing on the ground but I had to clarify that I didn’t know which I was talking about.

**Dr. Gates**

That’s a data dictionary moment.

**Student**

Are the data properties like rules? Or how would you define them?

**Dr. Pennington**

How would I define data properties?

**Student**

Yeah are they the rules that you once you go into the sensor and you pull information, does the data properties you setup to find what’s useful in that information or how is that set up?

**Dr. Pennington**

Well so what’s the I’m not sure I understand your question, so I’m just gonna try and give you an example and see if that clarifies for you. I have a sensor I’m measuring temperature of something of the air, of the ground, of something so the temperature of the actual temperature measurement is a property of whatever it is I’m measuring. The entity I’m measuring so I have the measurement and I have the actual entity and it’s a property of the entity, but when I’m doing data properties for the way we’re talking about it I’m gonna say ‘this time period at this place the temperature should range from ten to thirty.’ Now you’re talking about the property of the data stream not the property of the thing I’m measuring, and that’s a subtle difference but its one that’s not immediately obvious to the scientist when talking about it, did that help at all maybe?

**Student**

Maybe maybe not.

**Student**

You’re talking more data constraints versus actual data information.

**Dr. Pennington**

Right.

**Dr. Gates**

Anomalies come based on predicting what you think the data how you think the data will be behaving what the measurements.

**Dr. Pennington**

Yeah its comparing relationship and its comparing the data that you’re looking at the data, the data is now the entity of interest as opposed to the thing you’re measuring being the entity a subtle but important distinction.

**Student**

Going back to a question that Adam was saying earlier about transit, anything else you wanna be able to do just constraints at the moment appropriate constrains and anomalies? (garbled cannot be fully understood)

**Dr. Pennington**

Yes.

**Student**

So the users are able to defining the data properties?

**Dr. Pennington**

The users are the ones who would define the data properties yes.

**Student**

And when defining them you have to take the weather in to consideration?

**Dr. Pennington**

uhh sorry I’m having a moment of cognitive dissonance here, something is not quire right about your question. Um the weather is what you’re measuring in most cases you’re measuring temperature, precipitation, and wind velocity and those are what you’re measuring and you’re collecting data about that’s what the properties are about what you expect in that.

**Dr. Gates**

What you expected how do you expect weather to behave? And when you see a spike you wanna know when that spike occurs like maybe the temp raises and you’re not anticipating that, so its uh April and you don’t expect to have a snowstorm and when that happens you wanna be able to capture that moment, because it could tell you something.

**Dr. Pennington**

I mean it would look like a snowstorm or maybe maybe it is a snowstorm or maybe its one of the instruments has gone haywire.

**Student**

So I’m sorry, but if its not defined in the properties within the limits then its considered an anomaly?

**Dr. Pennington**

Right.

**Student**

Ok so do you think it would be helpful in that case to be able to somehow synchronize weather data?

**Dr. Pennington**

Yeah yeah that’s a good idea if you knew or had some other source of information that you could use too.

**Dr. Gates**

That’s number seven I think, question number seven on the second page is

‘Will we have access to database to check for predefined values historical values and time constraints?’

**Dr. Pennington**

Yeah so you can imagine instead of me specifying the temperature ranges that I expect from one particular value, I can access the weather and climate databases and say ‘compare this value to that database over that data source over there,’ and I hadn’t thought about doing that but that but yeah that would be helpful.

**Student**

How important is performance in terms of speed, for example would it be preferred to just have something that’s not gonna be very flashy I’m just gonna get you the very essentials in a very simplistic matter just quickly, or is it ok if it takes those extra seconds more to give you a very big but more not flashy but more precise range of data. Like what’s more important when you’re using this thing do you get a lot even if it takes some time or do you just?

**Dr. Pennington**

I think it’s always a good design decision to provide the things you can provide fast as fast as you can, and then you cordincall the others in the background. That’s always a good design decision, nobody likes to wait, for anything but somethings take time.

**Student**

I’m asking like how imperative is it just a slight annoyance or is it gonna hamper you?

**Dr. Gates**

How important is it? Is performance important basically?

**Dr. Pennington**

Performance is always important but if I.

**Dr. Gates**

Well your looking at tradeoffs, he’s looking at tradeoffs and I, that’s part of the analysis, I mean when you look at analysis and you start saying ‘if I’m gonna add these bells and whistles,’ that’s part of the analysis part so I think you wanna separate them.

**Dr. Pennington**

I mean it would be easier if you had a real concrete sort of would you rather have us this information right upfront and then this is five-minuets later, or would you rather or we can provide this in two minuets and if we don’t do this if you give me something concrete like that its an easier choice to make than.

**Dr. Gates**

And you’ll be doing that later on so, I’m gonna jump to question ten if an anomaly has occurred because of a faulty sensor where the sensor needs to be recalibrated how should that be recorded, I think you kinda talked little bit about anomalies is there anything else you wanna add?

**Dr. Pennington**

Well I think you know I guess you should proved some choices to the user in some cases they may want instant notification on their mobile. I mean its sorta like you know, I don’t know if you guys worked with the airlines at all they have all these choices about your ‘flight is delayed do you want to be notified an hour in advance or do you wanna know a day before,’ and ‘how do you wanna know? Do you wanna know by phone? Do you wanna know by email?’ When it comes to contact that’s a very individual sort of choice you need to provide those choices.

**Student**

it wasn’t a follow up I hate to go back to the data property I have been, but from what I’ve read the data property I thought the data property is also some calculations some sensors?

**Dr. Pennington**

Part of understanding a failure of a sensor is knowing how if this sensor has been showing lower and lower and lower values, and maybe this sensor is suppose to be showing higher and higher and higher you know they’re correlated maybe inversely, but there’s a correlation we’re expecting that correlation if the readings are not correlated like they’re suppose to then that can be an indication that there is something wrong with a sensor. So maybe almost certainly in many use cases will be comparing across many sensors.

**Dr. Gates**

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

**Student**

I’m sorry I have a question that builds more on what he was talking about, how would you be able to distinguish between what would be considered an anomaly and what would be considered a faulty sensor?

**Dr. Pennington**

Well yeah that would have to be the scientist would to do the analysis on that, I think its gonna be hard to tell just from the data unless you know, the data flat lines then you know that’s a problem with the sensor, but if it’s a calibration problems that’s much harder to detect so I think in most cases unless its just like a flat-out failure its gonna take some analysis by the scientists to try and understand that.

**Dr. Gates**

Just to give you a scenario there is so much information that’s streaming right? Depending you know if you’re doing this or if its on constantly and part of the tool is just to help the scientist to look at those places where unusual things are happening and then be able to analyze so its not sophisticated at this time to be able to say with certainty that this is a problem that has occurred.

**Dr. Pennington**

I mean just think about if we were working with satellite data, I might have petabytes of satellite data coming in on a weekly basis I just want something to call my attention to the places I need to look at.

**Dr. Gates**

And that’s the purpose here, ok we’re gonna go to data analysis real quick we have about ten minuets ‘Are there types of analysis needed other than transect predictions?’ I think we talked a little about that I don’t know if there’s anything else to add.

**Dr. Pennington**

I would say at least initially we want to keep the analysis pretty simple, I mean I think one of the uses I see in this is that as we start developing pat, you know 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 in three space. We can use data properties as a way to call our attention to patters that are occurring, so we can analyze the data that we collect about anomalies we analyze those and better understand what we’re seeing so that sort of analysis I see happening down the road but I don’t see this since this system right now you develop a different system to do that analysis I think maybe just trends and simple things to start.

**Dr. Gates ..continued**

There’s one about colors and stuff like that, I think that’s getting into design. Colors are [inaudible].

**Dr. Pennington**

Colors are important. You know if I’m looking at precipitation I wanted to be blue you don’t wanted 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. Are they in real time?

**Dr. Pennington**

Yeah, I want to be able to look at the graphs in real time. And if I like something I want to be able to downloaded it and print it out. [inaudible] application.

**Dr. Gates**

What operating system you would like to run on.

**Dr. Pennington**

All of them.

**Dr. Gates**

Mobile? All of them.

**Dr. Pennington**

You know scientists 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 work similarly to this?

**Dr. Pennington**

Nope not like this.

**Dr. Gates**

[Inaudible] entrepreneurship.

**Dr. Pennington**

Entrepreneurship opportunities.

**Dr. Gates**

Okay, good. Other topics that have been, that you think are important that we haven’t discussed.

**Dr. Pennington**

No, not that I know of.

**Dr. Gates**

I’ll open it up for other questions.

**Student**

So are we gonna show you, display anomalies compare to the data properties. I guess comparing the data properties to the actual data and to see if there’s a spike or anything or and the once we have those anomalies the scientist determine if those anomalies are errors?

**Dr. Pennington**

Absolutely, sure. [Pause] I guess it will be nice when you built this storage, however you store the anomaly information. It might be nice to include in that a way for the scientists to flag it and once they look at it and decide whether it was an error or whether it was a not an error we have to flag it as what it was. I think that will be an important information to collect.

**Student**

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

**Dr. Pennington**

Let me draw a picture, that’s probably the easiest thing. [Inaudible]. So I have some sort of instrument here that’s collecting information and so if it’s located, maybe its located right here on this tower. So here’s the ground. Now the instrument itself, let’s say that it has a special footprint, we called this we’ll say this is one (1) kilometer. I don’t know what it is, maybe it’s [inaudible]. The metadata is going to tell you it’s a got a one (1) kilometer resolution. But maybe what you’re measuring is a; so base on this what is, okay this instrument is collecting information around this area right here base on what the metadata says.

**Dr. Gates**

Well what if the wind is comes; what if there is a strong wind coming this way. Do all of this particles that you’re measuring are being blown.

**Dr. Pennington**

So essentially what ends up happing is you end up collecting information here. Does that make sense? Does that help?

**Student**

Not about this but skipping back to when you said the integration being able to share and all that, how far do you want to be as far as social like do you want have some kind of facebook, twitter or all that crazy chat or you’re looking at do it all in house internet browser and keep all the information inside there [inaudible].

**Dr. Pennington**

Well, it’s an interesting question. Most scientists detest facebook and twitter and those sorts of things. Primarily because they don’t see a point to it. They don’t want to chit chat with each other about…they don’t want to know what they’re doing right now.

**Dr. Gates**

They got their own problems to deal with.

**Dr. Pennington**

They don’t care what you’re doing right now. Unless it’s in this context of the work there’re trying to do.

So having said that if you think of a use case where it will be important for them to have some sort of social networking then it’s important to include it. If it’s just to so that I can be regularly informed of what you’re eating I don’t want it. So if you were going to put something like that in, I think it will be important to scope it in a way that it’s directly tied to what the work you’re trying to get done.

**Student**

[Inaudible] The facebook and all that sort of stuff it’s just idea for us how you want it be within that system. As I’m saying you want to have it set so the scientist when they share it to each other all of them see it and they have that kind of hang out all that discussion within that progress itself instead of trying to bring in facebook and all that.

**Dr. Gates**

I wouldn’t put in facebook, but if you want it..

**Dr. Pennington**

But that kind of thing like if I’m 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 have some discussion about it. Yeah.

**Student**

Does the system have some way to tell, what how like if the weather sort of adjusted the usual boundaries of the sensor.

**Dr. Pennington**

The reason this will be kind of important is when I’m designing my properties. I might want to say if the wind is under 10 miles/km, I do miles per hour. 10 miles per hour then I expect something this other data to be here in this range. On the other hand if the wind is you know over 30 miles an hour then I’m like I expect something different. So it’s all about expect settings, articulating what your expectations are.

**Student**

There’re predefined?

**Dr. Pennington**

We have a predefine way. 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 it detects anomaly do you want to per say just throw alerts up on the system send out emails, send out text messages.

**Dr. Pennington**

It’s go back to what I said about sort of the airline thing, 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 any sort of way or is this all confine to just within scientists?

**Dr. Pennington**

Well, so there’s an interesting move towards what they calling citizen scientists. Where you’re trying to involved people who are non-scientist in your daily collection efforts. And I can imagine that somewhere down the road somebody is gonna come up with some clever way to involve other people in this process, so it could be that it would end up being in the public. I might…so here is a scenario, maybe I set up sensors down in the plaza in downtown El Paso and I’m measuring something and there is some sort of anomaly and I want to engage whoever’s in the vicinity to go down there and check it out for me.

**Dr. Gates**

Sounds like a proposal.

**Dr. Pennington**

A proposal.. [inaudible] Citizen sciences are a really hot topic, because there’s uh, you know data collection takes a lot of time and effort. And if you can get people who are interested in whatever you’re studying to help you with it.

**Dr. Gates**

Good question. Will bring you in. Any other question? So let’s thank our speaker for [inaudible]. I want to thank everyone…

**Student**

I just have one last question, I know it said in the public [inaudible] that will be provided some source code but as far as the interface I don’ know it was mentioned or not and I was wondering if we gonna be able to see it or have someone show us [inaudible].

**Dr. Pennington**

I think that’s…

**Dr. Gates**

So I been thinking about that. In what I’m worried about is what Dr. Pennington talked about earlier is that she really wants it to be science centric. So I was going to show you the interface before but I’m worried that it will take away the creativity. So what I’m gonna tried to do is work with Dr. Salano to provide the information in a way, in a more general way that will help you come up with a new interface environment cause what they need in Tucson. So that’s gonna be a big part of the analysis that we do.

**Dr. Pennington**

I want to thank everyone.