

Interview with Susan Stafford by Max Geier on Thursday, September 25, 1997.
Transcribed by Keesje Hoekstra

Susan Stafford came in 1979 to the OSU College of Forestry with a background in statistics and quickly took up leadership of the Quantitative Science Group, the computing system, and information management for the Andrews Forest, leading to establishment of the Forest Science Data Bank. She was also a pivotal leader in development of the information management community and system across the Long-Term Ecological Research (LTER) network. In 1998, she moved to Colorado State University and then University of Minnesota to take on more administrative leadership positions. Throughout her career, she helped connect the Andrews Forest program with other LTER sites and with the National Science Foundation, which funded much of the development of management systems for environmental information for programs like LTER.

Max Geier: I'm trying to remember how much I told you on the phone, so I don't know if you've gotten any communications on what I'm up to here yet?

Susan Stafford: I thought I knew. I did get this draft outline of the book proposal.

Geier: Okay, good.

Stafford: Yeah, you did. I wasn't sure exactly where I was going to fit in, but I wasn't going to worry.

Geier: I've got a series of questions I was going to ask you.

Stafford: Okay.

Geier: But if you have something you want to talk about, let's go ahead and do that.

Stafford: Can we do the slides? (Speaks with support of visuals, including organization charts.)

Geier: Yeah, why don't we start with that.

Stafford: Perfect. Okay. So, I wanted to share with you the history of how we got to where we are today. Because where we are on that Xerox [historical lists/organizational charts], there is quite an evolution from where we started in 1980. I joined [HJA-OSU-USFS] in '79, and I was brought in as a consulting statistician, and there was one programmer on site. So, we've grown substantially, but we've grown so that the organization is in tandem with the research, that it was always coupled to where we saw the research going, what we saw we needed to be addressing. I like the slides because, "do unto data before it does unto you," is kind of good. Basically, what I wanted to show you is the Quantitative Scientists Group and the Forest Science

Databank. When I think about the goals for the Quantitative Sciences Group, I think of them as two-fold, and that's what differentiates us from computing centers, perhaps. We have to facilitate research, meeting the computing and the statistical needs that our researchers have, and of our students, but we also need to anticipate. So, we get involved in the science-doing as well as the making-it-possible-to-do-the-science, because the role of science and technology is getting so intertwined today, that you need a foot in both camps to do that. So, what I wanted to do is say who we are, where we've been, and what we're doing and where we're going.

Now this is where this next list is important. This was done, probably in '93, so in the three or four years since then, this list has been replaced with this group here. What's interesting is that here, it's just a linear listing of who we've got. There is a director and a consulting statistician, myself. We have a databank manager, Gody Spycher, who is over here. We have Tom Sabin, who's an assistant statistician. He left to go and work with Gore, the people that make Goretex. And we hired Dr. Lisa Ganio in his place, so we've grown in that direction. Mark Klopsch, the network administrator, is now the [OSU] College of Forestry Computer Coordinator. Ken West was Mark's assistant, and what we were able to do there, is move Ken in by a direct appointment, because he was obviously the candidate. I did a market analysis and we were able to go through the process so that Ken took Mark's position, and then, we hired behind Ken, and that's the position that Sean San Romani has. And Barbara Marx, who's our GIS technical support programmer. She's still with us. Lisa Ganio was a spatial statistician at Point Sevren, and we did a similar move of Lisa into Tom's job when Tom moved.

I think there's great advantage in being able to promote from within if you've got the high-quality people that we've been so fortunate to attract. When we are going off in new areas, then we have open searches and we do everything as broadly represented as we can, but I don't see any point in pretending if we pretty much know that we've got the best talented individuals here, then let's be up front with that and go with it, so that's what we did with Lisa. Sharon Clark is our GIS-geographer, and Michelle Murillo was providing some UNIX support. She has left and has gone to Los Alamos Lab in New Mexico, and we have Taralynn Vendetta that we hired as a UNIX system administrator. And then, we had a whole bevy of students, so we've got pretty much all of these positions covered plus some more, because you see additional names.

The other thing that is really important to notice, is that from our inception, we've had a very close partnering with the U.S. Forest Service, and in particular the Landscape Ecology Research Work Unit. That's Fred Swanson's group. And it will become even clearer, because you'll see some grant-writing that came in a strategic time that was really able to sort of get us on the right trajectory. And these people are pretty much here as well. There's Don Henshaw, who is here, there's George Lienkaemper, who is here, there's Hazel Hammond. And John Gray was recruited into this spot. Maria Fiorello was working in remote sensing, and we have Sharon Clark and John doing some of that work now. Steve Acker is doing some data management, but Gody and Don are doing a lot of that now as well. And then we've always been fortunate enough to recruit students. So that's the evolution, just from '93-'94 to where we are now. We also have a close tie with the U.S. Forest Service Management Systems Group. Nancy Barnes

said that position is currently vacant. Carla Veach, who's in telecommunications, and Theresa Larabee.

So, you can start to see how the tentacles of QSG, the Quantitative Sciences Group, are really expanding out and extending, so that it's very difficult in most instances to know who is Forest Service and who is Forest Science. We as scientists, think that is very helpful. Where that causes some challenges is on the administrative side of trying to keep things that are supposed to be bureaucratically, kept separate, and so we have to work on that. One of the things that we've tried hard to do is break down any potential barriers that could come from those who manage the data and manage the computer network, and those whose data we are working with.

The old model of taking your data to the "high priest" in the computer center, we're trying to break all that down. Everybody within QSG, has, to some degree or another, some either biological or natural resource background, in addition to their technical ability. So that gives people that you're working with as a researcher and as a student, the sense that they understand why this is important, and they are as intrigued with your work as you are. And it really helps break down any barriers that, or, "Just give me the data don't tell me what it's all about," because people spend their lives collecting their data, and makes it easier. Another thing we do is we give a series of workshops so that we can train our users on all that which we feel they need to know about information management, and they can pick and choose. We have workshops on how to get into the network and how to remotely log in, how to run GIS, how to learn UNIX at an introductory level, so we try to do some of the teaching features as well. And then we have a help desk, which is in a centralized location.

In fact, we're just now working on the schedule, because one of our key people has left and she was doing a number of shifts on help desk, so we're scrambling to cover. But the idea there, was to provide a point of contact for users in the building who had immediate problems to come and get some help. It also was a time management ploy on the part of QSG, so that people wouldn't be coming in all the time, interrupting when there's something you're working on that you really need a big block of time to get done. So, we're constantly trying to wrestle with how do you meet the needs of our users, and how do you preserve and protect your own schedule to try to do things? One of the things I think you have to realize is that within any user community there's great diversity. So having one approach, or a one-size-fits-all kind of strategy doesn't make it. This is an area that we are continuing to improve ourselves on. But, there's a number of different kinds of folks such as extroverts and introverts. That's an important issue, especially for some of our international students who feel that it is inappropriate to ask for help, or they're burdening us, or they are imposing on us in some way. That's why having the help desk where people can come one-on-one, where you have workshops, so that people could sign up ahead of time. And then Lisa and Manuela and I teach a class winter quarter. It's basically entitled "Natural Resources Data Analysis." The point of the course is to teach students how to analyze their own data. And we teach them how to use stats so that they can take their synthesized information that they've learned in their service stat courses, and know how to get their research done.

This may seem very general for the history of the Andrews, but the Andrews and the cadre of students from the Andrews, all fit within this model, so it's not separate. You know that we're part of the LTER network? Right now there are 20 sites within the LTER, but you probably know that. Basically, what we're doing here is encouraging and fostering the idea that data management might be integrated into that whole research process. We came up with a systematic approach, it was really as much for us as our users, because it demonstrates the phases that we like to see people go through. What we want to do is avoid having students come in after the fact, with a poorly-designed experiment or poorly-documented data, and then not be able to pull the most out of that work. We start with study planning, where the PI, the investigator, the researcher, sit down and identify the objectives of the research. Come and sit with a statistician, get the statistical design approved or discuss what options there are, and then also see the data manager so that you can get your data set up from the outset, appropriate for long-term archival storage in the Forest Science Databank. Then, go out and collect the data, and that's the part that usually happens first. Work with the folks on data documenting and editing. Then comes the fun time for the analysis. That's where the class comes in because I hope that my students use, if not all, at least a portion of their data in the analysis, for the final project in the class. The wrap-up is data interpretation and synthesis, where you come back and sit with a statistician, if necessary, to help with interpretation of results, so we make sure that what's being said is statistically sound. Many of the issues that we work on the Andrews, have potentially controversial findings or audiences. It's even more critical that the statistical rigor of the approach be very sound, and whatever boundary conditions need to be attached to conclusions, are stated up front. Because you'd much rather do that yourself, than have a letter to the editor do that for you. And then, I always emphasize that we want to get the results out, synthesized and published, because if something is just sitting in a notebook or on a floppy disk, it's very helpful.

Geier: Does that include placement on inner and outer web sites?

Stafford: Now it does, now it does. That's a really good question. There is tremendous interest on behalf of the National Science Foundation to have as much data online as possible as quickly as possible. Some of the sites within the LTER network are making a distinction between their core data sets, the ones that are on the five major areas, primary productivity, disturbance, whatever, and the work that their graduate students are doing. We find that a lot of the work that our students are doing is really interesting and has long-term potential. So, we would like to have the graduate student data, in most instances, put in a form, so it could be in the data bank and then available over the web. We're not there yet though. We have some of our data available online. We have abstracts for data available online and things like that. There's still a great deal of discussion within the user community on what constitutes a reasonable time limit. And funding agencies are saying that what is reasonable from the funding agencies' views, and what is reasonable from someone's view who sits on their data and doesn't get their act together and doesn't get it published very quickly, are two very different things. So, we're wrestling with this. We had a meeting of the data managers from each of the LTER sites in Albuquerque in

August. One of the topics of discussion was helping NSF and the LTER community re-write what would be an appropriate data-access policy at the site level. So, that's a long answer. I'm sorry.

This is to point out that timing is everything. We've found that if you work with people at the right stage in that research process, again integrating the whole attention to data and information management at the appropriate step, rather than after the fact or trying to retro fit something, is a far more effective strategy. So that's sort of in a nutshell, who we are. The QSG, Quantitative Scientists Groups, I think of as having five major areas: the connectivity support area, the data management area, the statistical consulting, technology, and then as we've grown, this physical tracking has become more important. These are the kinds of things I see us providing in these different areas, and these are the people that are associated. Now I can go through and mark who is OSU and who is Forest Service, but from our perspective. That's not a meaningful distinction, because we really try to work together in a partnering way. So, let me give you that.

Geier: Oh, I've got that.

Stafford: Let me give you a little bit on where we've been. Like any group, we started with mainframe, which was in the late 70's, early 80's. We suffered with the same kinds of things that most people go through, delays. Then we had the evolution of PC's. Not only did we have one PC, we had tons of PC's, and it became clear that a set of PC's that weren't connected to each other, was not going to be helpful. So, we needed to figure out a way to create a local area network. And that resulted in 1987 in a grant to NSF with Phil Sollins, Fred Swanson, and Stan Gregory, who should all be familiar names to you in this scenario. We wrote a grant for an integrated science workbench for ecosystem research. We asked for online accessibility of what was currently in the Forest Science Databank. We wanted to provide high resolution graphics and computing power, so we got one whole Sun workstation. We now have 44, and a local area network [LAN] linking the PC's to the FSDB [Forest Science Data Bank]. So, basically in '87 we were successful with this and were able to make our first stab towards connectivity. Needed to get the bugs out. And a year later, with a slightly different group of folks, Fred Swanson, Tom Spies and Bill Ripple, who's in the ERSA Lab, [Remote, Environmental Remote Sensing and Application Laboratory in OSU], with that we were trying to link GIS with remote sensing. We were working on connecting the ERSA Lab to what we had in the Forest Science LAN. We added another hard disk, another Sun workstation, and what was really critical is this is where we got the seed money for Barbara Marx, who was the support programmer.

What was interesting about this is that this is the first time we were able to get money for a person, rather than just hardware and software. This was a real step in the right direction in terms of recognition on the part of NSF, that besides the stuff, we needed the people. We were able to use that one year of support as a way of leveraging other support from other projects. When this support ran out, I was able to move Barbara because she'd already proven her value to other projects. This is a schematic of what the local, the FSL LAN [Forestry Sciences Lab Local Area Network] looked like back then. This is the comic who said, "Wasn't it your idea to

improve communications,” and everyone’s with mega-phones. It became clear that we recognized that we had sort of two entities that were not really connected to us.

Geier: Yeah, I was going to ask about that.

Stafford: Yes. The Forest Service and the H.J. Andrews. In fact, this is now what we look like. The Forest Service is part of this FSL, Forestry Science Lab. So, as a result of this next grant I’ll tell you about, that was the seed money to make this connection. The Andrews is on a 56 KB line to the Andrews in Blue River. That is a very modest connection. If we are going to grow the H.J. Andrews into the vision that many of us have for it, we need to be bringing down a T-1 line out to the Andrews, but that takes a significant commitment of dollars. The rental on that line per month is \$1,000 to \$1,500. It’s substantial, and you can’t run it as a small mom-and-pop operation; you’ve got to step up. This is how we are today. We wrote a grant that provided a gateway to the Data General, a local area network for the Andrews, with some wide-area networking connections, a Sun spark station, and net-ware SQL for increased network storage.

Geier: That was 1990?

Stafford: This is 1990, and the three grants were key. The first one was a separate grant, the second two were technological supplements to existing LTER grants. And what NSF and Ecosystems Long Term Studies, or whichever program you want to associate with there, did, was recognize that if LTER was going to mandate that 20 percent of the site budget was going to go into data management, they also better make some resources available to build the infrastructure. And then it was up to individual sites to take advantage of that. That was a very insightful move on the part of NSF and the biological sciences directorate to not make this sort of happen occasionally, but really put some top-down impetus by making resources available, opening the competition so that people could go forward and compete.

Geier: Do you know who was behind that idea?

Stafford: John Brooks. John Brooks was the Division Director in Division of Environmental Biology of NSF, and David Kinsbury was the Assistant Division Director. Tom Callahan was the program manager in Ecosystem Studies. So, you’ve got the program, the division, the directorate.

Geier: Uh-Huh. (affirmative).

Stafford: And there was alignment at all three, which was really critical. As a result of that grant, we were able to do a better job connecting with the Data General [Forest Service ‘DG’ communications system], which you see in here, and which is growing now because of the IBM contract and the 615 Project. But back then they were just worried about the Data General, and connecting with net lasers through to the Andrews. I have to admit that the solution that we used with net lasers, had we made this now, we wouldn’t have chosen that. One of the things

you're always wrestling with is, that if you wait just a little bit longer, the technology will change and improve, but sometimes you can't afford to wait. So, you've got to jump, you've got to do, and then you have to be able to stand back and evaluate the decision you made, and decide, do you keep going with this? Or has it served its purpose, and you unplug it and put something else in?

Geier: What's the current concern about the net laser?

Stafford: It's older technology and it is trying to make the 56KB line sort of as a kluge. Whereas, if we bit the bullet and went with the T-1 lines, I don't believe the net lasers would be necessary. Some of the net laser problems emerge as you change versions of Novell from 3.1 to 4.1 or whatever. The net lasers were designed for the older versions, and it's always, you push the upgrades out over the 56KB line to the remote locations, they're not able to figure out what to do with it. So, you're not always able to capitalize on the improvements in the versions of some of the software, because of some of the trappings you have in place. This was the grant that we wrote a year later after we worked on connecting the things, then we decided. As we always do, as our users always decide, that whatever we have isn't enough, so we double. On one grant, in one fell swoop, we doubled the GIS capacity that we had, and we enhanced the databank. This is what it looked like then, and this is what we were able to add. But you can start to see how we grew this system in a sort of a strategic step-by-step way, and a lot of times that's forgotten when you come in now and look at as many PC's and as many Sun workstations we have.

But, it's critical to your story, because of the role that the Andrews and research on the Andrews played in allowing us to capitalize on opportunities through that Long-Term Ecological Research program and mindset at NSF to build the infrastructure here, on-site and also out at the Andrews.

Geier: Uh-huh (affirmative). So, there's a long-term planning component?

Stafford: Yes. Everything we did from that very first grant in '88, we designed the system, knowing that it was always going to be growing. And that way, when we were able to go in for the subsequent grants, we were able to say, "This is where we are, this is our long-range plan. This is what we are requesting funds for to move us along this trajectory." This was the internet connection in, '90-? Let me think here. I think in '94 we had about 180 PC's, a year later we had 280 PC's, and we now have probably close to 500 PC's. That's just on the Novell side, so you can just see this exponential growth that we're going through. The Forest Science Databank history is interesting because it gives you a little perspective, starting in '73 through '80, the data were on mainframe tape, paper documentation, entry forms. Then in '80 to '84, we moved to a tape library with automated access abilities. In '84 to '88, we transitioned to the PC, because that coincided with the time that we were able to write the grants to bring the PC's, the local area network, and the FSDB up as a node on that network. Then we ported to the Novell server. Now, we are looking at client server architecture and trying to create the Forest Science Databank in such a form where the data and the metadata, primarily focusing on

the metadata, would be query-able over the web. A lot of the sites are working on that. We have our data in flat ASCII files, and we have our metadata, the data about the data, the documentation, if you will, in FoxPro primarily. So, you can start to see how we've gone.

Geier: Was there any effort to integrate this planning with what was going on in other sites?

Stafford: Oh, yes, definitely. In fact, because we were the first cohorts of sites funded in '80, when we would write our supplement grants, we accepted the responsibility of prototyping, so that others sites didn't have to do it again. We could do some of the testing of things, and then we could share the results. There's a great sense of community among the LTER data managers. We meet every year. We have our annual business meeting. Sometimes these meetings are held in conjunction with other professional meetings, like ESA, sometimes they're held with data management meetings where we will organize a contributed paper session like we did a year ago at Ecolnforma. So, it gives everybody an opportunity not only to gain visibility within the scientific community, but also to share and let your hair down. You'd say, "Boy, you know I'm really having trouble with bringing up Solera," an operating system version on the Sun workstation, and, "What are other people doing?" Some people trying to pull together the network information system are playing with many SQL's. So, we have a couple of people doing that, we have a couple of people looking at Oracle, some of these larger database systems. And we can kind of have some tech support groups within the network itself. This is allowing us, we came in early, to look in terms of some of our protocols for capturing information about data, how to manage it, how to document it and everything else. Other sites use that from us.

Geier: So, it would be accurate to say the Andrews has become or this group or the LTER group here has been the prototype group, for other long-term ecological research groups?

Stafford: Yes. Early on, we felt we were a flagship in the information management arena because we were fortunate enough to have resources. We were also fortunate enough to have this partnering going on, that allows more people to be brought into the fold. More challenges to be addressed as well, more jobs for everyone to do, but we were able to get a critical mass more quickly, and we were actually able to build a staff or a group, rather than having half of a position dedicated to data management, which is what some of the sites [LTERs] have to do.

Geier: Do you get any sense of, in relative size of your group here. This is a large group here. Is this the largest group in the LTER?

Stafford: I think so. In terms of this kind of operation, I don't know of any that is larger. We are fairly comprehensive too, in terms of tasks. I think there are groups that, probably within the GIS area, that may have more and go deeper than we do, but I don't think that they have the companion and statistical consulting strength. We are a little bit broader-based, and fairly deep, when you look at that. The only thing that's misleading, is that some of these people are not full time. They are liaisons to other projects, but these other projects have information management, and data, and connectivity, and statistical needs, so we use a person within their

project as sort of a conduit, as a liaison. That's where we've been. In terms of where we're going, this whole concept of metadata and data is an important one to us. Are you clear on metadata?

Geier: I was just going to ask, how do you define that?

Stafford: Metadata is what you would need to make sense out of the numbers that I collected on my field study. So, you would need to know why I collected it, for an abstract, was it an overall study purpose? You would want to know what were the formats of the data? You'd want to know how they were collected, you'd want to know what units they were collected in, you'd want to know whether they were rounded or not. If I had some codes, you'd want to know what those codes meant. I'd want to know what those codes meant, because two days after I use them I forget. All of that metadata, which used to be called documentation, is something that we've spent a lot of time on. And this is an area that other sites have followed our lead in as well.

Geier: Context is kind of what it sounds like.

Stafford: Exactly. So, metadata is a standard set of all pertinent information describing the data that is essential for personal, independent access. NASA has a policy that they would like their data to have adequate metadata, so someone can come in a decade later and make sense out of what they used. When I talk about the Forest Science Databank, I'm talking about the metadata as well as the data. And in the metadata, we're talking about catalogues, studies, formats, data files, the investigators, the owners of the data, if you will. That's a tricky word, we're trying to get away from data ownership because of the push to make data accessible. Project categories, locations and p-work. One of the efforts that we're currently involved in right now is taking all of the data, all of the papers that the Andrews group has created, and key-wording them. You've probably heard about that.

Geier: At the last LTER meeting I think they were talking about that. Sounds like quite a project?

Stafford: Yes. But the abstract forms, the format forms, the code forms, all of that resides in the metadata. This is an example of a form. I can show you this. Actually, I use it in my class notes here. So, this is a kind of paper form that we have. We're moving to online forms, but this is a good example where you've got the variable names. Are there blanks embedded in this variable? What's the format? Alpha numeric means that you're going to have letters as well as numbers; "i" is integer, "f" is floating point. Is it coded? Yes. What's "ni"? I've forgotten. Units, centimeters, what's the minimum, max, and whether it's missing. The minimum and max is important because we can run checks. If you're saying that basically you're not expecting a tree to grow more than two or three centimeters, and I get a number out here that's been punched in or data entered, that's seventeen, well, then we can instantly flag it, and we can go back and say "Well, was it entered wrong? Was it transcribed incorrectly in the field, or was it transcribed onto the screen incorrectly?" So, this would be an example of the variable format.

Geier: So, the scientists would use this when they put up their study, to say what they want their database to be showing?

Stafford: Yeah.

Geier: Okay.

Stafford: And you know that second block, where it said, "Sit down with the data manager?" And, "Talk about the data documentation and the database design?" This is where you would hopefully start identifying how you are going to organize your data. How are you going collect your data? And you've captured that at that point. These are the variable names, and then this is what they are. Sometimes it's very clear, sometimes it's not clear at all. So, having the accurate definition is very useful. Then here's one. If you have site 1 through 4, treatment 1 through 7, animal damage code 1 through 4, cause of death, what do those mean? So, you have to capture all that. That's what we talk about in terms of what these are, but this is a little easier to see. That's a definition form, and that's a code form. We've gone through the abstract format and the codes, and then the catalogues. Some of those are in place and some of them we're working on. But that's what you would think of as a catalogue, a listing, so that we can do that.

And with the web, it becomes even more important to have these catalogues accurate so that people from the outside know what resource we have. We have all different kinds of data. We have meteorological records. I'm sure talking with Art McKee, you've heard about the length of records there. The met [meteorological] data is really interesting, Max, because these are the data that people are always the most willing to share. We have some vegetation plot data, mortality, growth and yield data, some of these are now decades-long. You know, so when you think about the resource that those data represent, or model validation, it's incredible. And that's why it's important that, I'm not preaching to the converted, but that's why we've always felt that taking care of these data are resources in and of themselves. Because you're not going to have the funds to go out and start a new study every time you want to test a hypothesis. So, if you can use and capitalize on data that's already been collected and well cared for, you know you're using it in appropriate ways, then you can build on what you're doing and take those new research dollars, and collect the data that you haven't gotten already. So now, what we're able to do is take field reconnaissance data that we've had for 60, 80 years worth of records, and link it with satellite data which wasn't available when these studies were initiated, and you can start to sort of calibrate on the ground with what you observe from the air.

Geier: Was there a problem with taking the way that data was collected in, let's say 1951, and putting it into the same system of data that was collected in 1988?

Stafford: That's a problem. It's less of a problem than you might think, because you're not starting '97 with data that you collected in '52. The data that the databank inherited was brought into the fold as it was sort of brought along. Before we were a LTER site, we were a IBP

site, and you heard about that. They [IBP era workers] also had a directive to take care of data. The problem was, you could put data into a system, but no one could get it out. LTER was working on making that arrow work in both directions, from the data side. The most recent project, it's not LTER work, was anadromous fish habitat survey data from the '30s and '40s on the Columbia River. One of my students went back in early '90 and re-did that, and what he looked at was habitat change. Part of it was teasing out how much of it is because the protocols of collecting the data had changed, versus that the actual system had changed.

Geier: Yeah, yeah.

Stafford: So, you have to recognize that.

Geier: Has the accuracy of the measuring techniques changed?

Stafford: Yes.

Geier: I was also thinking about the format in which data is stored, starting with written records.

Stafford: Yes

Geier: And then, I don't know, punch cards?

Stafford: Right. We've tried to convert everything to magnetic media as fast as possible.

Geier: Is that complete?

Stafford: I don't think it's ever complete, but, what I found was the best strategy to use was to go on those data that someone was interested in working with you on. If I were to independently say, "Okay, all these data have to be converted," but no one has a pressing need for these data right now. That's like pulling teeth. But, if all of a sudden somebody realizes that this old data is really going to be critical to sort of give an initialization, of where the forest was, or where the stream was, or where whatever was 20 or 30 or 40 years ago, they will move mountains to get that data up and ready and cleaned. So, yes, there's probably stuff lurking out there, but it's by need. And need and utility sort of defines the strategy for going back and rescuing the data.

Geier: So is the biggest barrier time and people, or is it technology? I mean for the old records.

Stafford: No, time and people. It takes a lot of time. If you were to sit and talk with Gody Spycher about cleaning some of these old data, there's a long story. And that's why you really, I think you have to recognize, that the system will never be perfect. And you have to pick where you really want to put your resources and where you really want to put your efforts. It's on a computer, and I've got the data, but I can't get it out.

Geier: Yeah?

Stafford: And for me, it's technology. And people, I mean, I need somebody to do it for me.

Geier: Really?

Stafford: Stream data. I'm sure you're talking with Stan Gregory and the Stream Team, and then speak with Linda Ashkenas as well.

Geier: I probably should.

Stafford: Is she on your list?

Geier: No, I don't think so.

Stafford: A-s-h-k-e-n-a-s. She is the right arm of Stan's effort over there. She'd be another good one. And Mark Harmon, his log decomposition study and the LIDET project. Mark is an excellent example of a scientist who really pays attention to data management issues. The only way this multi-site, international, multi-national project is working, is because he's put in the time to coordinate the data management for all these litter sample bags that he will get back here and analyze. And you can imagine the nightmare, if you lose track of which one came from the tropics and which one came from Russia.

Geier: Right, yeah.

Stafford: So, that's a good example. GIS, Geographical Information Systems. That opened up a whole new arena for us in terms of spatial data. The issues of metadata with spatial data is really fascinating because it's edge detection. How do you know if you're in the middle of the cell? Yes, you're probably correct it's a value that shows that pixel is quite good. As you get out to the edge of that pixel, should it still be a 7 or should it be an 8? And of course, that depends on the granularity of the data, as well. Landscape issues are driving primarily the research focus that I see from the technology side. If we were content not to look at the landscape level, then a lot of the work we're doing with remote sensing and GIS and imaging, I don't think would be as intriguing as it is. This is an example of the 16 years between 1972 and 1988 in the central Cascades. Changes in land-use patterns, cutting patterns, you can start to see the increased fragmentation of the forest. I'm sure you've spoken with others who've talked a lot about this. This is a remote sensing image of the Andrews. I think of the Andrews as sort of a heart lying on it's side. And it's sort of somewhere over here.

Geier: A heart lying on it's side? (Laughter)

Stafford: Yeah, when you look at pictures of it, satellite imagery and things like that, it's sort of a heart lying on it's side. The satellite imagery came from a study Dick Waring did with NASA. There is an interesting point to make about collaboration. When we have good data, it's easier

for us to leverage the investments that NSF and the LTER has made with other agencies, because they know the ground on which they're working, has been very carefully worked, very carefully measured, and it's correct. So, then they're much more likely to come and do overflights and run campaigns that will help augment the data layer. So, when somebody says, "Gosh, 20% of the site budget, is an awful lot to invest in data management," and not every site is doing 20%, but it shouldn't be viewed that way. It should be 20% of one investment that allows you to leverage into a whole set of arenas that, if you weren't doing this, you could not get entry into the game.

Geier: Has it always been 20%, or when did that come up?

Stafford: That was what NSF said when they started. You know, somewhere; 10, 15, 20%. And the sites that have been more successful have tended to be closer to the upper end, and the sites that have been less successful, have been ones that say, "Well, they say that, but I don't really need to do that." Because NSF is not going to come and watch over your shoulder about where you spend their dollars. But they are going to hold you accountable to the standards that they expect, having given you 560,000 dollars a year. So, that's where you have to come out. We've used the commitment from our local LTER to pay for all of Gody Spycher and half of Ken West, so it's about a 1.5 FTE allocation. Now, as salaries increase, even though they're very snail-like at Oregon State, and the budgets have been remaining flat from NSF, those dollars haven't gone so far. What I've done is gone to information services, to get some institutional match to make up the difference between where we need to be to cover the same quantity of people versus where we were before. That was part of our match and we were successful with that.

But this is the QAQC, or Quality Assurance Quality Control program. Basically, what we're trying to do is have the metadata, which is in the system file, conform with data in the ASC II form. We use rules. If trees are not expected to shrink, trees are not expected to change species, trees are not expected to come back to life after they've been dead for 10 years. Those can be used as checks or as rules so you make sure that the data you've got in your records are correct. So, you're actually using the metadata about the data to clean the actual data, which is good.

This is a little bit out-of-date, but it gives you a summary of the current contents, and an idea of the diversity of data we've got. We've got data on aquatics, hydrology, geomorphology, vegetation, meteorologicals, terrestrial vegetation, the litter decomposition we talked about with Mark Harmon, as well as other litter, soil samples, and biodiversity inventories. On mammals and arthropods, we're not real big on that area, and we probably could do more wildlife ecology, and other data which is non-LTER, which is also housed in the Forest Science Databank. Also, data from the Forest Science Department, genetics, forest engineering, vegetation management. We've got several co-ops that are looking at vegetation management and alternatives to managing vegetation, competing vegetation, so we are the repository for those data. The other thing that's important, is to find out who is asking for your data, and to keep track of how many requests you get for your data. This is made easier now with the web,

because if you have a site you can count the number of hits. The only trouble is you can't always identify who that hit was. But it's helpful from the PR side, because this way, if you're going after more resources to make your data more accessible, you're showing there's a market for these data. If somebody doesn't give a rip that you've got 80-years-worth of successional mortality and growth data from six western states, it's going to be hard for you to make your pitch. But, if you're able to show how many different people are using your data and for the great range of uses, then you're more successful in asking for resources. This was an example of recent data requests.

This is back in '92, but what's interesting - see the met data, that's what this one says. These are all for modeled runs. They go from University of Washington, University of Montana, University of Massachusetts, colleagues at OSU, colleagues here at the Andrews, the Arctic LTER, the EPA lab, University of Oregon. You can start to see the broad range that you're able to touch with your data.

Geier: The way it is now, if someone from outside is looking for information, would they send a request for information and you would generate a report of some sort? How would that work?

Stafford: We would have our catalogue on the web, and that would show general categories of data that we have. We don't necessarily have the actual data on the web in all cases, but we'll get to that point. They would be able to look at the kinds of data that we would have and then there would be instructions for how to obtain the data.

Geier: Okay.

Stafford: Typically, they would send an e-mail, and that's on that other flow chart. That would generate a request, we would ask them some questions. We would say, "Who are you? What would you like to use these data for?" If it's data that the PI has released, then there's usually no problem. If it's data that the PI says, "Yes, I'm very willing to share, but I'd like to know more about where these data are going." Then we couple them with the PI and the PI would release it or not. We also have an acknowledgment statement that we ask be included in any publication that comes with having used these data. So that we can get some credit where credit is due.

Geier: I was going to ask that because, it strikes me as probably an evolutionary process to get scientists to release proprietary control over data.

Stafford: Yeah.

Geier: Has it always been that open?

Stafford: No. And within the network this is probably the biggest non-technological issue that we're dealing with. Because you have a whole spectrum of people. I mean, it's like those Christmas trees, you know, the introvert, the extrovert, the whatever. You've got every kind of

individual within the network. This is where those annual LTER data managers' meetings come in so handy, because you're able to talk about strategies that have worked at one site, and issues that have come up. At the Cedar Creek site, they have something that they have installed.

End of Side A, Tape 1 (of 1)

Beginning of Side B, Tape 1 (of 1)

Geier: About wore this thing [tape recorder used by Max] out when I was up at the Andrews. They had a group interview up there.

Stafford: Oh, my God.

Geier: With Roy Silen and Bob Tarrant and about six other people.

Stafford: Oh, my goodness.

Geier: And we went up to Carpenter Mountain Lookout up there.

Stafford: Yeah?

Geier: I went through six tapes.

Stafford: Oh, my goodness.

Geier: Yeah, and about three sets of batteries. (Chuckle)

Stafford: Oh my God. That's a new definition for long winded? Huh?

Geier: (Chuckle)

Stafford: The Cedar Creek site have what they call the "data pledge." What someone needs to do is they download this data pledge, and it says "I promise to follow the ethical guidelines" of whatever, and then they sign. If it is data that are, I think, under three years old, then the PI is automatically included as a co-author on whatever publication comes out. If it's between three- and-a-half and something else, they consult with the PI on the use of the data. And if it's older than whatever, then it's free access. They were talking about, rather than having a bible, you know where you put the pledge or whatever, having a little icon that was Darwin's "Origin of the Species" that you put your hand on.

Geier: (Laughter)

Stafford: Clarence Laymen was the source of that story, he's the data manager at Cedar Creek, and it's great. We have a directive from NSF that all data collected on NSF funds, especially within the LTER project, be made available with a minimum of restrictions, in two years. What we're doing within the data management community, is trying to identify which data could easily fall in that category right now, and which data would it be appropriate to have some sort of exemption or exception clause. When we talked with Scott Collins from NSF about this, he said, "That seems reasonable. If two years is a slightly funny number and it needs to be tinkered with a little bit, help us tinker with it! But, don't make all data at a site an exception, because that's not going to fly." I think a sort of sociological evolution has gone along also. You have a site on the East Coast, Hubbard Brook, that's very well established with "statesmen scientists" who say, "That's absurd! I can't possibly glean everything I need to glean from my data in two years. I won't do that." Well, NSF is going to be at a point soon, where they're going to make that ruling have some teeth, whether it's withholding renewal funds, or whether it's increasing the sternness of reviews that go on their record. I don't know. So, anyway, it's kind of an interesting process.

So, that's who we are, where we've been, where we're going, and what we're doing, now where are we going. One of the things that has been very clear from the onset of this, is that the kind of data that we deal with, scientific data, are very different than what public software houses have been designed to deal with. When we talk about database issues, and you talk to someone who runs or writes programs for American Express, or Visa, or Bank One, or United Airlines; they're dealing with transactional data. How you're connecting from Portland to Chicago to Detroit to wherever, that's very important, and after you've made that trip it's no longer important. That's a very different kind of beast, different from the kinds of data that we're dealing with.

Commercial data tend not to have much in terms of longevity, whereas, the value of our data increases the older it is, if the metadata is of good quality. We are going to do things to those data that you would never do to the balances in your checking account. You're not going to run Fourier analysis on what you've got in your checking account. Scales, whether it was over large scales or whether it's very tight small granularities, for these landscape-level kinds of things with images, and reconciling one data set over another. It's just a whole different kind of bailiwick to get into. And, as a result, the software that we've been able to use for this has never really met our needs specifically, and so some of these things have to be retrofitted, and force fitted into something that it really was not designed for. As a result, that's given rise to computation ecology programs at NSF and other places so that algorithms and software and approaches and procedures and test beds could be developed that are really targeted for scientific data versus an off-the-shelf commercial application. So, that's been an interesting thing.

Geier: So, part of what you've been involved in is the development of software?

Stafford: To a certain degree. What we did here is, we don't have a stable of software engineers. So, it is a lot easier for us to contract out, hire consultants, from a subsidiary of

Microsoft or whomever, and have them come in and do an overall assessment, and then help us design what kind of fixes we need. And that's what we did this last winter. Gody and Don had gone up to Beaverton and had taken some training, including server-training. The people that were running it were very impressive. They were from UDP, United Data Processing or something. They started chatting with these folks, and they said, "We could come down and spend a couple weeks with you, and do a problem analysis." That was really helpful, because we were able to lay out where our catalogues were, what kind of metadata we had, and what kind of data we had, what kind of connectivity and platform problems we had, because some of our data are UNIX. All of the spatial data reside on UNIX machines, and most of the databank files are on PC's. So, you end up some platform independent solutions, rather than something that's only going to work in one environment or the other. So, we've done a little bit of that.

The internet has just gone ballistic, you know. What it's done from our side of the equation is that it has increased user expectations for accessibility, both recipients and generators of data, and federal agencies. It has shortened that timeline to make data accessible, just overnight. You can't think of non-web-based solutions. You can have a broader way of solving the problem that includes some non-web applications, but you will be dead in the water if you don't include that portion of the community that's just moving in the direction of the web. This was a schematic that we played with a few years ago. It's looking at the data management system, the DBMS, looking at the FSDB, the metadata, the data. We've used it more as a talking document than anything else, but we played a little bit with some of these things. Visualization software. This came out of CORAL. This was a group that started in University of New Mexico, under John Razer. He's gone now and become a company or corporation unto himself. But a lot of the visualization software that the San Diego super computer center is doing with their Monterey Bay project. Things like that are really interesting. John Helley came to the meeting in Albuquerque and showed what Stuart Gage from Kellogg Biological Station had done with taking images of the Kellogg site, and then with different sun angles over time over seasons, showing the greening of large areas of land. So, you could start to think you could link that to some climate models, and how you would tinker with photo periods and things would do under different warming scenarios. It was really fascinating. That visualization is an area that we need to do more of because we underestimate our abilities to visualize something in two dimensions sometimes. This was what Warren Cohen was doing on the modeler's project. I don't know if you're talking with Warren. That's a NASA project that is looking at 14 different LTER sites, and they are developing methodology and protocol for three biosphere variables. Net primary productivity, decomposition rates, and, I don't know, precipitation or something. It's a good example of looking at how LTER sites have been able to partner with NASA and other large agencies, to make more out of the investment that each individually is making.

Geier: I haven't talked to Warren, although I noticed when I was down at the Andrews today that almost everybody down there, all the students are working on his projects.

Stafford: Yes, exactly. And this was something that he was using. He used this visualization software to show how his model was taking five different data layers, to compute these different variables. We're allowing the data to be represented, so that whoever is looking at it

can get something out of it. People are going to pull different things from it. That's why documentation is so important. Because for somebody animal damage codes are meaningless, but for somebody else that's looking at impact of migration patterns on the landscape, that would be terribly important. This is an example of a diagram for landscape-level synthesis through major modeling efforts that use all the data layers, so all of the circles are modeled. All of the flat things with broken-off edges are actual data layers. This diagram was called the Scotch Diagram, because that was what helped the thinking along one evening as they were working on it.

Geier: Was the Scotch. (Laughter).

Stafford: What was interesting was looking at how data could be not only input, but output, and it could be both. The data coming from a model, could be input into another model. And where you think about the implication for that is the generation of computation of errors. If the data in a model is found to be incorrect and then it's propagated through five or seven different models, what is that overall implication? Another thing is, sometimes in this process, even though you try to make all the right decisions, it may be that you realize that a model run that you made, that you then based a number of other things on, was initialized incorrectly, or that the calibration was off. So, what you need to do is the metadata for this has to be able to track what different things were set at, so that you could actually go back to February 17th, and retrace where you are to September 25th, and that requires a different level of collecting data. Some of the software packages are good because it will keep a log of everything that was done, but you might like to have that synthesized in some way rather than having to painfully go back so many days. But you can start to see some of the implications, and, and clearly this is something that scientific data has to deal with people that work with scientific data rather than, you know, United Airlines reservations. Then the challenges. This integration of GIS and remote sensing user interfaces, visualization software which we've already talked about. Distributive analytical environments where I could be working and someone could be working, and we could be looking at the same data and we could be sort of looking at them together. Better sampling resolution and standardization. We'd like to have transparent computing environments, we're getting better at that. The fact that we have people here working on Unix software that are Sun Unix and then IBM Unix, and we have PC's and we have PC's that are Windowed and PC's that are NT systems. We have Windows '95, we don't have Windows '95, and we have some Mac PC's.

Geier: What does this term transparent mean then?

Stafford: That you as a user are not constrained by the environment in which you're working in. One of the things that has been bounced around is this notion of "National Environmental Data Archives." We've played with the concept of trying to develop a network information system for the LTER. We've recognized that at each of the 20 sites there is expertise on site, which is probably in the best position to be in a daily contact with the data. But, it would be nice to have a single point-of-entry that would then fan out and find data at each of the sites or whatever sites are appropriate, to deliver it back to whomever is morphing it. So, what we're doing is developing DTON, which is Distributed Table of Contents. Each of the 20 sites [LTER]

creates a standardized table of contents, and it's distributed because it's across the network. Then, if a query comes in, they look at that aggregated table of contents, and the data might be at the Andrews, it might be Coweeta in Georgia, or it might be at North Temperate Lakes in Wisconsin. It doesn't matter, you can go and find it. And we'd like that model better than everything all in one place, because everything all in one place implies a certain amount of custodial maintenance that probably isn't going to be done as effectively as it will be at each of the individual sites, because the people who collected the data are most interested, and it is their prized possession.

Geier: That would be the integration of the data generation and data management from that somehow.

Stafford: Right, exactly. The other thing that's interesting is that within the LTERs we have taken very different approaches. We are much more bottom-up, research-oriented than top-down. A few of us went to China and talked to the CERN Group, the Chinese Ecological Research Network (not the big European physics lab). That was a very different concept for the Chinese to deal with, because they were of mindset that there would be a computer center and there would be the high priest in the computer center and there would be these directives, and everyone would do everything according to these directives. Whereas from the LTER, you have 20 different sites and each is competed for independently. If they are not doing excellent site science, they will not be renewed, and then they have an obligation to work at this next higher level, which is in a coordinated network fashion. The data management sort of parallels that as opposed to having one site that's in charge of all the data, that's not the way it works.

This last thing here is the training of the new cadre of scientists. This is something that a number of us have been trying to do. We've been trying to develop a curriculum in ecoinformatics, or environmental data management. If you take the 20 data managers from the 20 sites, and you sit them down at a table and you ask them, "Now, what did you study?" Every path to that table would be different. But the content of what they covered has great similarities. So, what would be nice is if, instead of being a computer person that was interested in ecology or an ecology person that was interested in computers, you could actually take a bona fide course and become an environmental information manager. That would be much more direct. What we'd like to do is use the pockets of expertise that are represented across the LTER network to allow some internships. So, people would come here and they might take some in-depth statistics classes in planning and design, they might go to one of the other sites like Sevilleta that has a beautiful facility at the Sevilleta field station with dozens of Sun workstations, and they could learn Unix and learn a lot about their technical abilities. Then, if you were interested in tropical research, you'd go down to the Puerto Rico LTER site, and do an internship there, where your thesis topic or whatever, would be based on managing tropical data. Someone else might be interested in the Arctic, someone else might be interested in the Antarctic, and you've used the diversity of the network to the advantage of the learning experience.

Geier: Is there a typical path that brings people into this group that you're working with here?

Stafford: No. Most have some background in ecology or biological science or natural resources, and there's a strong computer component of some sort, or an analytical component of some sort.

Geier: Formal degree in that?

Stafford: Yes.

Geier: Oh, okay.

Stafford: It is less likely to find pure computer scientists than it is to find a hybrid, probably because if you're going to get a Ph.D. in computer science, or even a bachelors in computer science, you're going to be picked up by Microsoft or Tektronix at four times the salary than someone is going to pay you at an LTER site. But you're not going to be working with biological systems, you're not going to have the opportunity to do field work, and you're not going to travel and do the kinds of things that are intriguing to biologists and ecologists.

Geier: Did you do field work that much?

Stafford: My undergraduate was in biology and was a lot of math. My masters was in quantitative ecology, and I did a computer model study of commercial fishery stocks on the Great Lakes. There was a model that had been put out by some folks in, I want to say Green Bay, even though I was not up there. I tested that model with Ontario's data. So, my field work was going to Sandusky, Ohio, and getting the records of walleye and yellow pike and everything else. For my Ph.D. I shifted and I went into applied statistics, and there I worked on developing a model for determining land values of forest and vacant lands in three counties in upstate New York. The theme was the application of mathematical models to a question or an issue important outside of the math. I'm not a theorist, that's not what I'm intrigued by. I like to solve problems. I like to figure out how to bring an organizational arrangement to something to make it facilitate and expedite what we're all about. When I started with my one programmer back in '79, and then we kind of grew the concept of the Forest Science Data Bank, we obviously had to be tied in with connectivity issues and the technology issues. We increased our scope through the kinds of statistical course work that we're able to teach. So, no, I don't do a lot of field-based research.

Geier: What attracted you out here, because most of your work was in the Northeast, wasn't it?

Stafford: That's right. There were two states I wanted to work in. One was North Carolina, and one was Oregon. My aunt lived in British Columbia, and I used to come out in the summer. I used to be my cousin's receptionist in her therapy office. Then, in the afternoons I would volunteer as a craft teacher for the Boys and Girls Club, and we'd go and take hikes and do all these fun things. So, I ended up thinking that this would be a nice area of the world to live in.

My major professor came in one day, three days before this assistant professor tenure track consulting statistician position closed, in a town called Corvallis that I'd practically never heard of. I quickly faxed a letter that said I was interested, and if they were still open, I could send my materials. John Gordon was department head at the time and he said, "Oh, by all means send your stuff." Then I got a call, this was like on a Tuesday, I got a call at home on Friday, and he said, "Well, we're very interested and we'd like you to come out Monday." I said, "Oh, a week from Monday." He said, "No, Monday." Well, we compromised, and I came out on Thursday or whatever. I lost my luggage, I nearly missed the plane, and I did not have a reservation. So, it was a four-day interview, in a shirt that was on my back, quite literally, and I was offered the job before United ever found my luggage in San Francisco, and I've stayed ever since. And it's been kind of a growing opportunity.

Geier: A four-day interview?

Stafford: It was four days because at that point it was a joint appointment between forest science and statistics [OSU departments]. Through the interview it became clear that it would be better if you had your grounding in one department, and then, you have a courtesy appointment in the other. But I was worried about visibility, because, if I was here, helping students and faculty, I wouldn't be over in the stat department, and, if I was there, teaching and working, I wouldn't be here. So, I could be doing really fine work, but because as a consulting statistician you need to be accessible, whatever I was doing elsewhere, even of high quality, would become a little bit of a liability. So, we worked it out that I would have a full-time position in forestry, and I would have a nice courtesy arrangement in statistics. We kept that, and it worked fine.

Geier: Did you get a tour of the whole LTER site when you were doing this work?

Stafford: Yes. Most of the LTER meetings. Jumped from one LTER site to the next, and I haven't been to all of them, but I've been to a number of them. I have not been to the Arctic site, and I haven't been to the field sites for McMurdo Dry Valley and Palmer Station. They are out of University of Reno and Santa Barbara, so they've got some continental U.S. location. But, a lot of the other sites I've been to.

Geier: Who was the major professor that kind of put you onto this site here?

Stafford: Bill Steitler.

Geier: Bill Steitler.

Stafford: He is not LTER, but he was a statistician in the College of Environmental Science and Forestry [State University of New York, Syracuse], and he just came in and threw this thing on my desk quite literally one day. He said, "Here, take look at this." All my compatriots, all my classmates were unemployed for at least a year, so I had a hundred resumes copied. I was going to finish up and then I was going to tour Europe. I was going to do all these wild things,

you know. I ended up with 99 resumes left and this great opportunity to come out to Oregon State.

Geier: Your doctorate is from SUNY, right?

Stafford: Yes. At SUNY, and my undergraduate was at Syracuse. That's a private school.

Geier: I was curious when you first started working here, I don't know maybe I'm interrupting your thought?

Stafford: No, let me just wrap it up. I think there's a distinction between perfection and perspective, and we've talked on that a little bit. We've taken this, going from point A to point B, you can't get there from here kind of thing. Sometimes you have to take a slightly different approach, this is the bottom-up versus the top-down theory. They're chasing ketchup, thinking that it was blood.

Geier: (Chuckle)

Stafford: And there's going to be some of that. If you're on the front edge, as Gody likes to say, "On the bleeding edge," you end up making some false calls, but you have to do that. You can't go forward if you're scared not to do something that might, given the knowledge you'll have in a year from now, you'd choose something differently, but you have to go with what you've got. Expectations. Especially on the level of metadata, people have very different views on what constitutes completeness and documentation and things like that. Students love to spend three months or three field seasons collecting data, and come in on a Friday and expect that they're going to have it all analyzed by Monday morning. Well, give the analysis its due. That really is part of the fun, as well. If you knew how it was all going to turn out, you wouldn't have done the experiment in the first place. Some things are not quite the bargain that you think they're going to be. That's a nest that's upside down here.

Geier: Okay.

Stafford: It says, "No wonder it was such a deal, if all the eggs fall out." Everything is changing fast, and you have to really recognize that. You can't just hunker down. You've got to be able to be flexible, you've got to plan for growth, and you've got to plan for lots of opportunities and lots of challenges. But opportunities really balance higher than the liabilities of the challenges. I think you also have to recognize what you can be and what you can't be. Put your money on that part of the technology you really need. You don't really need a sledgehammer to kill a fly. But, if you are moving into interactive models and a lot of visualization types of things, you can't be bounded by a 56KB transmission range. So, recognize when you need to be the turbo, and when you can be the VW Bug. Don't confuse the two. I always end with this one; the genius of the future lies not in technology, but in the ability to manage it. We get that confused sometimes.

Geier: Yeah.

Stafford: So, that's it.

Geier: I like your use of Gary Larson.

Stafford: (Laughter)

Geier: I did my Ph.D. at Washington State.

Stafford: Oh.

Geier: He's an alum of that school.

Stafford: Is he really?

Geier: Yeah, and the year I graduated was the centennial year, and they commissioned him to do a centennial cartoon. So, my diploma had a Gary Larson cartoon in it (Laughter).

Stafford: Oh my gosh! Oh my gosh!

Geier: I've always had a soft spot for him.

Stafford: Yeah, yeah.

Geier: I want to ask you a few questions about the beginning of your involvement here with the Andrews, and your impressions of the group at the time you started. When you came out for that interview, did they take you down to the Andrews, or if it was more of a campus-site interview.

Stafford: I never saw the Andrews, no.

Geier: Is that right?

Stafford: We weren't an LTER site at that time. Dick Waring, and Jerry Franklin were involved with the Andrews at that point, and they had a parting of the ways. But Dick had recommended and Jerry had confirmed, that it would be good to get me more involved with the kinds of data and statistics and information management types of things. I just started to have responsibilities for supervising and managing people that worked on different components of the Andrews.

Geier: I was curious. You came in here about the time that the Forest Science Department [OSU College of Forestry] moved into this building over here?

Stafford: Yes.

Geier: I wonder if you would talk a little bit about your perception as a faculty member of the impact of that move over here.

Stafford: We were in the Forest Research Lab [on Western Boulevard] for a few years, three or four maybe. Then we moved over here. When we first entertained the opportunity to move, it was because the Forest Service was shrinking, and they were quite worried. Bob Tarrant was involved in this, and you've spoken with him, that they would get neighbors in this building, but they weren't sure who they would be. It seemed it would make a lot more sense if the neighbors were kindred spirits and sort of research-collaborators. So, the whole move was based on really very good reasoning in terms of bringing people who view the world similarly together and are interested in working on things together, together in one place. The problem was that John Gordon left and went to Yale, and part of his negotiations were that it would be one contiguous space that would be open, and we would just pick up and move. Well, it didn't work that way. There were pockets of spaces. So, there was a little bit of space here and there was a little bit of space here and there was a little bit of space here. In retrospect, I don't think that was all that bad, because it allowed some interaction with people that, had you just been an island all by yourself, in one corner of the building, it would have been easy, not to associate with. It's sort of like, if you're trying to learn a language, and if you don't immerse yourself in the French speaking part of town and you're living with the Americans, you're not going to speak French, you're all going to speak your own language. So, that was good. I think that was an opportunity.

Geier: What you're saying, it was by accident, not by design, just kind of happenstance?

Stafford: Yeah. There was tremendous resistance to change then. This is what Bob Tarrant noticed, and he said, "Okay, we're going to move and this is going to be fun, this is going to be exciting, we're going to make this work." Bob Tarrant was the interim department head between John Gordon and Logan Norris. We were adaptable, so we were just going to go with the flow. The tone could easily have been Forest Service versus OSU, if you're not careful, and in some camps there was this, "Who's this tucking their nose under the tent? We're very content here and thank you very much. Just leave us alone." Those of us that had services or were providing functions everyone found useful, like statistical consulting, were assimilated quite readily. So, I didn't find it to be a problem. I think it's a little awkward to have our mailboxes so far away. If someone sends me something hardcopy, it could be days before I get down there. Whereas, if it's e-mail or phone or something like that, it's a lot quicker. But it's good exercise as well.

Geier: I was curious about the integration of the Forest Science faculty with the rest of the university after that move was made. Was there a difference there?

Stafford: No, I don't think so. I spent a year in the Provost's Office [OSU] as a faculty associate. What struck me about that was how isolated the College of Forestry was from the rest of the

campus. It didn't matter whether we were on the corner of 30th or whether we were over here, we are a group unto ourselves. We are funded independently, primarily, and we are funded generously in the eyes of many of our other colleagues across campus. I would maintain that we work very hard for the funding that we get, that we're not getting stuff handed on a silver platter. But, I didn't realize how isolated I felt at times being over here, versus someplace else. When I was in the Provost's office and I spent half a week a time, 50/50 split, I would be right in the hub. And because other departments are sort of around that central core, you'd be aware of things. You'd get a *Barometer* [OSU campus newspaper]. Here, we don't get the *Barometer*. It's a different kind of mindset. We're maxed out though. We're all busy, we're all running around like chickens with our heads cut off, so we're certainly not losing anything. It's just that our focus has been on what we are about. I see my colleagues in meetings and at symposiums and workshops and at talks you're invited to give across the country, more than I see them here.

Geier: Someone told me that the building, when it was designed, it was to keep people separated. Do you find it functions that way?

Stafford: Yes. Our QSG group is pretty much down this hallway and around that corner. I work to develop a sense of community within our group. We've had two retreats now, two annual retreats. We do a few social events, like a holiday party at our house in December. We've had a whole rash of babies, so we've had baby showers. We've sort of blended the life within and the life outside of work. I think you have to recognize that there are two parts to that. But, it would be very easy to come in and not talk to somebody all day. Those of us that like talking to people aren't going to abide by that, so we're like that extrovert tree. But it is easy to be isolated.

Geier: You were here for a long period of transition of the LTER group or the Andrews group. I was curious if you sense a change over time in the degree of integration of the group. What is your perception of when it seemed the most integrated, or maybe it's not so much now?

Stafford: It was very different when Jerry Franklin left. Jerry was a very strong figurehead for the group, but also for the national LTER. I maintain that Fred Swanson has done an excellent job recognizing the diversity of talents and personalities within the group. I don't know what is a truly realistic level of integration? I think you have an extremely talented, extremely diverse, extremely busy group of individuals that come together, not for the money, because there's hardly any money that gets passed to the salaries to most folks, so they're committed to the theme that the project is working on. There are going to be differences, and there've been people that have come into the fold, and there are people that have left. But, in terms of what manifests the group working perfectly, I don't know. I think you have to --

Are there things that aren't happening? From my perspective, if I need to bring a group together to talk about data access policies, we can make that happen. I find that I have to take the responsibility for that. If I'm going to have to make sure that I'm at the meeting, so that I have a voice at the table. That's not just going to be stretched out to me. If I was of the kind

where I easily felt like I was left out of something that could be a problem. But I'm involved in so many things that, if someone else wants to take care of something, I look at that as relieving a burden, not excluding me. If I feel strongly about something, then I'll go right in, and march in and say, "Look, these are the things that really have to be addressed." The web is a good example. I feel very strongly that we have to have a web position, because we are stretched so thinly trying to support the infrastructure that we have and then this web comes in. And, with the Forest Service and the need for certain fire walls for their corporate data, that the Forest Service absolutely maintains, I mean it's not a choice, this is how it has to happen. We can't be glib about these things, and we need to do these things in a smart way. It's my place, I think, that when I feel strongly or I see the writing on the wall, to make sure that that becomes an agenda item.

There are good friendships within the group. There are individuals that socialize together, they ride bikes together, they go for walks together, they do similar recreational kinds of things together. I'm not in that group. I don't feel left out. That's a matter of choice in what we do. I'm very active in a whole other set of things. I think that the group is composed of very good people, and good people are just by design spread very thinly, so you end up having to pick and choose. And I think that these choices that people make can be misinterpreted as lack of interest some times. For me, it's more a scheduling problem than anything else. I think that I do more on the LTER network level with the other data managers. I chair the data managers' committee, I chair the task force for the steering committee of the national group of data managers. I feel very connected that way. Whether or not I'm going to get in and arm wrestle over budget determination at the site level, I don't, no, I mean, I didn't. I could have chosen to get really upset when the data management allocation was trimmed. But, I saw that as an opportunity to go over and ask for institutional match, because quite frankly, OSU needs to do something in the institutional match area. I spent a year at NSF, and I wouldn't have been asked to go to NSF by the division director, if I didn't have a connection and a visibility within the LTER program.

Geier: When did you go to NSF?

Stafford: I spent the calendar year of '94 as a Division Director in Biological Instrumentation Resources. [BIR]

Geier: When you went there, did you have an agenda or purpose?

Stafford: Yes. In fact, I wrote a paper on it, I'll give you a copy.

Geier: Oh, Okay.

Stafford: I wanted to build the visibility of that division, because that's where a lot of their training programs are and a lot of the infrastructure development, the database activities program, the computational biology program, instrumentation development, shared instrument program, all the Novell programs that NSF and the Biological Sciences Director was trying to get up and running, had a home in BIR. For me, BIR was the best kept secret in the BIO

Directorate, and I loved getting on that soapbox, and going on site visits and seeing what NCEAS [National Center for Ecological Analysis and Synthesis] was doing with Mosaic at the time, and what Carnegie Mellon was doing with visualization software. You know it was a fascinating experience to me. Wonderful. Then from the administrative side, I realized that institutional match is something one has to get out there and hustle for. I e-mailed Fred when they were wrestling with some themes in the synthesis areas of LTER 4, and he asked, "Who's thinking about the match?" I said, "Okay, let me see what I can do." To me, that's how a family group should work. Those that are doing something and they are central to the core, and if it's working, they will reach out, they'll get input, and then they'll take what steps they need, and then they'll go with that. I don't expect to have everything I say converted into a policy. But, I certainly want to have a place at the table when a policy is going to be set up that I either am impacted by or I have to help implement, or I see as being shortsighted for the long-term benefit of the site and the network.

Geier: If I understand what you're saying, the level of integration of the group depends a lot on the individual. In other words, how integrated you are into the group depends on how interested you are in becoming integrated?

Stafford: I think so.

Geier: I have a question here about you being at NSF. Do you see that as unusual for people who are involved in the LTER program to go to NSF, or is that something fairly typical?

Stafford: More and more are going to NSF. When I was there, Jim Gosz was Division Director in Environmental Biology, the division that LTER is run out of. And he was a former P.I. on the Sevilleta LTER. Right now Bruce Hayden from the Virginia Coastal Reserve site is Division Director of DEB Environmental Biology. Gus Shaver is a program manager there, and he's from the Arctic site. Scott Collins is permanent, Gus and Bruce are rotators, as with Jim and myself.

Geier: Jim?

Stafford: Jim Gosz.

Geier: Yeah, okay.

Stafford: Scott Collins is now permanent NSF program officer, and he was from the Konza LTER site. The other way of looking at it is that LTER individuals have a high degree of commitment to service their community, and NSF is just a fascinating place to work. It gave me insights that were unbelievable. I knew I was going to learn things, but had no comprehension of how much I was going to learn. And I think I was able to offer something, but it was a truly incredible experience. I was going back-and-forth because my family stayed here. So, I worked out an arrangement where I was there for three weeks and I was home for a week, and I was there for three weeks and I was home for a week, and I did this the entire 12 months. We had that all worked out ahead of time.

Geier: That's gotta be tough. That schedule. Well I was curious, because I know Jerry Franklin was there kind of early on.

Stafford: Yes.

Geier: Has anybody else from this LTER site served at NSF? That you know of.

Stafford: That's a good question. I don't know.

Geier: No other names have come up. I was curious.

Stafford: No, no. I don't think so.

Geier: Okay, all right. Well, a little bit of shifting of gears here.

Stafford: Yes.

Geier: I want to get perceptions at the time you joined the LTER here, about your understanding of what the purpose of the research group was, and what the purpose of the Andrews Experimental Forest was. More generally, how a university and an experimental forest were integrated at this location.

Stafford: I remember the early LTER meetings were rather unstructured, for me. Maybe because as a statistician or whatever, you start with your hypothesis, then you go to the givens, and then you kind of make your plan. It took me awhile to realize that things got done, it was a highly productive group, but if you were going to chart how these things occurred, that mapping would be a lot more erratic and a lot less linear than what you might imagine. In terms of the relationship with the Andrews and the university, the college and the department, I think I took that for granted when I started. I have a much better sense of some of the stresses involved with this now, after coming back from NSF, after seeing other similar political and apolitical decisions made for supporting off-site locations; the transmission rates of lines being run-down there, the capabilities a site can have versus the liabilities a site has to endure. I think I have a much better grip of just how difficult it is to maintain an off-site, world-class, premiere research facility, in light of budget cuts and personnel demands and everything else. But at the time, since I'd never really been based for research down there, nor been on-site support and the application of methodology that solves questions and problems. I don't think I was impacted by that as much.

I think the new facilities, the bunkhouses and everything, are incredible. It's an incredible place. The old tacky trailers that looked like they were going to collapse and people were going to die because someone would drop a match or something, those were scary. I was glad that they bit the bullet and did that, but it changes the complexion of that location. It will never go back to what it was before.

Geier: You were talking before about the integration of data management with research early in the process when that's taking place. How necessary is it for a data manager to visit the site where the data's being collected? Is that common?

Stafford: Yes. Don and Gody, who work most closely with data, do that a lot, Don, primarily on the met station. Did you meet Fred Bierlmaier when you were down there? He has scraped up a local area network for the met stations. He can get downloaded on the web in almost real time with some delay, the flows. So, when we were having these flood events, we could get on the web and we could see that it was about to happen. Don has been out there a lot. So yes, Don, who I see as being in charge of the wet data, and Gody, who's in charge of the dry data for the field plots, get field experience. Gody actually has a Ph.D. in soils, and has worked on the Andrews. Don has a Master's in Statistics from Oregon State, so they have a grounding. I like going out on field trips, because when you're lecturing, it gives you a better sense of how to develop a context for these examples that you want to give.

Geier: I was also curious, you've been talking quite a bit about your concern for remaining visible, and you mentioned students quite a bit in your discussion here, which is actually unusual from some of the people I've talked to in Forest Science. I was curious first of all how often students are involved in that early level of coordinating data management with data collection.

Stafford: Through all students who take my class, and this is a chapter in the book, so they get a dose of it. Now, if I'm not on their committee, or I'm not part of their advisory group, I can't make sure that they do the right thing. But it's not for lack of having been told that these are things that should be attended to. So, it's interesting that this session was more student-oriented.

Geier: Well, some people have told me that one thing they don't like about the groups is the lack of opportunity to teach. People like to talk to the students; it just depends on the person, I guess.

Stafford: I don't think that's LTER. I think that's because we are a graduate-level department. A lot of times people are talking about teaching at the undergraduate level. The number of courses taught at the graduate level is smaller, but that's a fine comment for someone to make. But they need to be asking themselves, have they sat down and put together a course syllabus and then trotted it past Logan [Norris - department chair] and said, "Can I run this as an experimental course?" That's how this class got started.

Geier: Oh, really? When did you start teaching this class? You might have mentioned it before.

Stafford: No. I started it probably in '80, the first year I was here. When I first came out I was teaching the Forest Engineering Institute program. It was six weeks in statistics and operations research. And then, a year or two later, '81, somewhere in there, early '80s, we got started.

The class size was six, seven, eight. Now, we have 40-50 students, which is too many.
(Laughter)

Geier: I believe that.

Stafford: So -- [pause].

Geier: I want to ask about your role as a university professor working with this interagency group. Much of the work involves short-term funding. What are the advantages and disadvantages for a faculty member in that kind of situation?

Stafford: The advantage of having it short-term is that, if you're on the wrong track, you're not on it for very long. It gives you an opportunity to pull the plug on something, if it's not working. That's not usually where we're at, though. Usually we're trying to make something work and we've succeeded with just a little bit of resources, and now we have to leverage it into something else. I think you can complain about it or you can just recognize that that's the way this is, and that in the big scheme of things, we're incredibly fortunate for the amount of resources that do flow our way. And with that comes the recognition and the responsibility to maintain a quality of work, so that, as these resources get tighter and tighter, we're still at the top of the heap. I think that it's no surprise that the Forest Science Department is the top grant getting department in the university, and we have this investment in technological infrastructure that allows the research to be of high quality and statistically sound, processed quickly and managed well.

End of Side B, Tape 1 (of 1)

End of Interview