

Colorado State University Libraries

Conference Proceedings and Events

CI Days: Cyberinfrastructure 2010 in the Rockies

Transcription of Cyberinfrastructure 2010 in the Rockies: conference agenda, 2010

Collection: CI Days: Cyberinfrastructure 2010 in the Rockies

Title: Cyberinfrastructure 2010 in the Rockies: conference agenda

Date: 2010

File Name: CI\_Days\_2010\_Agenda.mp4

Date Transcribed: November 2024

Transcription Platform: Konch AI

BEGIN TRANSCRIPTION

[00:01 - 00:02] Jeff Bullington: Good morning.

[00:05 - 00:05] Audience: Good morning.

[00:05 - 01:59] Jeff Bullington: So it's time for us to get going. I welcome you to the day. [paper rustles] I'm glad that we have people here. And we've still got some of that other... [stutters]. Breathe, Jeff, breathe. Other people will be showing up as we move forward. A couple of housekeeping details, just like in any good airplane or movie theater, cell phones to vibrate or silent, please. [paper rustles] Hopefully, in your packet you will see the various materials for the day. A couple of things, including a building map of this floor that will point out where the breakout sessions are this afternoon, hopefully help you find restrooms, which again, just right out here to the left, to the right and to the right. Refreshments and everything else in the back. Please take care of comfort needs as as they present themselves for you during the day. My name is Jeff Bullington, and I'm one of the librarians here at Colorado State University. And we welcome you to this event about this thing called Cyber Infrastructure, which is always trying to figure out what that is. And I'm going to go over just a couple of things so that we get started and then we're going to jump right into the day. We've got lots of things to cover. We're going to do our best to stay on time at every stage. We appreciate everybody's help in keeping us on time and keeping us moving forward. [clicks tongue] I'd like to point out and introduce some of our sponsors. So we have Pat Burns, who is the, and you will raise your hand when you hear your name, who is our VP for Information Technology and dean of libraries here at Colorado State University. Alan Torrens, who is the director of the Colorado Alliance for Research Libraries, and H.J. Siegel, who is the director of IS Tech, which is the Information Science Technology committee here at CSU.

[01:59 - 03:55] Jeff Bullington: And you will learn a little bit more about IS Tech and its unique role at this institution a little bit later this morning. Other people who have helped with the organization of today include Don Pascal, who is the assistant Dean for Digital Publishing here at Colorado State University Libraries. George Makovec, [unconfirmed name] who works with the Colorado Alliance for Research Libraries. Allison Lavelle, who is one of our Science librarians and who, I bet is still out at the registration table helping take care of things. Chris Sonnier, who is another one of our assistant deans here at Colorado State and he is not here. Other people that we would like to acknowledge and thank for being here today, Tim Gallagher is our vice chair of the Fort Collins... CSU Fort Collins Faculty Council, and is attending today. Make certain, all of the alliance deans and directors who are present. And so Helen Reed from University of Northern Colorado and Maggie Farrell from the University of Wyoming, and Mary Somerville. Is Mary here yet? I know she was registered. And I think that's everybody who's a dinner director. And Jerry. Sorry, Jerry. [laughs] [audience laughs] I don't... my notes are cryptic and imperfect. So I'm doing a lot of this by memory. Other people we'd like to point out, in addition to Russ Harvey, who's one of our speakers, who's here from internet to Anna Preston, Hunsicker is here from internet, too. I don't know that she... and right back there. Great. Janet Lowe, who is on the board of directors for the Colorado Alliance of Research Libraries and a faculty member at the University of Wyoming. Ron Sega, who is our Colorado State University's resident astronaut and a faculty member in field, drawing a blank, Electrical and Computer Engineering. Thank you.

[03:55 - 05:49] Jeff Bullington: And I think that's everybody. Pat, am I forgetting anybody? Because he'll tell me. [laughs] Yeah. [audience laughs] So the next thing is we want to do a little bit of audience aerobics. And again this is the same thing just raising your hand. So the first thing is, how many people are here from university... library segments or libraries. Most of us. Okay, great. How many from campus Information Technology operations, not libraries? How many people from faculty sector or research sector? And how many people from other external organizations, not directly in a higher ed institution, but connected? Wonderful. So we've got a good mix. As of my last count, last night we had 121 people registered. We have people from, I think every single research Library Alliance Institution registered today, and we are looking forward to a good event. And I'm running early. And I don't think there's anything else I need to talk about. Again, you can review in the packets. We've got the agenda for the day. You have a list of registered attendees. One of the really important things that you'll want to visit a little bit throughout the day is there is a packet that has, for those breakout discussions that we are doing in the afternoon. We have all of the kind of starter questions that the planning committee had developed around all of those different topics. Those are going to be the basis for starting discussion this afternoon. We can go in different directions, but those give people some things to think about. We did consolidate groups down to six from the

original ten that we started with. And again, on your name tag, it should tell you which breakout you are in. And then the other really important thing is it should tell you what your lunch choice was that you selected way back when.

[05:50 - 07:11] Jeff Bullington: Because like all things like this, they're just going to bring out a pile of them and people take what is theirs and hopefully don't take somebody else's. And those are the most important things. And again, you've got a building map which should help provide direction to where those breakouts will be this afternoon. What room? We have signs outside of each of the rooms as well. Am I forgetting anything, Don? Okay, well, I suppose then we can just go on ahead and get started. And so the first discussion today is from Russ Harvey, who is from the Internet2 Consortium, and Russ is the director of the SIDs program. And again, what the SIDs program is, is that NSF funded grant opportunity that different institutions can apply for, to host a SIDs event for their institution or for their area. And we were one of nine sites throughout the nation that got a grant during this program. And Rob is going to come up and give us a brief overview about the SIDs program and other things about infrastructure, Cyber Infrastructure. So, Russ, if you want to come on up. [paper rustles] [object thuds] And then she just [inaudible]. She's just getting it- [crosstalk]

[07:11 - 07:11] Russ Harvey: Going.

[07:11 - 07:11] Jeff Bullington: And then- [object thuds]

[07:12 - 07:12] Russ Harvey: Okay.

[07:12 - 07:13] Jeff Bullington: And then you're good to [chair scrapes] go.

[07:18 - 08:07] Russ Harvey: [murmurs] So good morning. As Jeff said, there's a lot of different definitions for cyber infrastructure. And so I'm going to kind of just go through one vision that we've got. As he mentioned, the CIDs has been run by a consortium of organizations. So it's just not Internet2, Edu cores and LR, MSIC, Open Science Grid, Terra Grid, Search Grid have all been participating in this program to help out campuses with their cyber infrastructure in their planning and mainly trying to get all the parties together on the campus to start talking to each other because cyber infrastructure has grown into a large subject. Next.

[08:08 - 08:15] Jeff Bullington: Oh. Let's see. It's not working. Technology, don't we just love it?  
[object rattles]

[08:18 - 08:18] Russ Harvey: Huh. You have to hold your mouth, right?

[08:19 - 08:23] Jeff Bullington: That could be it. [laughs] Well, it was working.

[08:26 - 08:26] Nancy: Well-

[08:26 - 08:28] Jeff Bullington: Nancy, can you just-

[08:30 - 08:31] Russ Harvey: I can yell next.

[08:31 - 08:32] Jeff Bullington: It worked. I'm serious.

[08:33 - 08:33] Russ Harvey: [Russ laughs] [audience laughs]

[08:33 - 08:35] Jeff Bullington: Half an hour ago, it was working just fine.

[08:35 - 08:37] Russ Harvey: Well, you know this technology stuff. [laughs]

[08:38 - 08:38] Jeff Bullington: Yeah. [laughs]

[08:38 - 08:38] Russ Harvey: Jeff-

[08:38 - 08:38] Nancy: Try to-

[08:38 - 08:39] Man 1: Breathe, Jeff.

[08:40 - 08:40] Jeff Bullington: Breathe. Yes. [laughs]

[08:41 - 08:41] Russ Harvey: Try again.

[08:47 - 08:47] Jeff Bullington: Cyber infrastructure.

[08:47 - 08:56] Russ Harvey: Yeah. I know. That's part of the reasons why we need to work on this stuff. [laughs] Okay. So I'll just yell next.

[08:58 - 08:58] Nancy: Okay.

[08:58 - 12:17] Russ Harvey: So why are we having CIDs? Well, cyber infrastructure is playing an increasingly important part in, particularly in research but all aspects of of what's [chair scrapes] going on in the academic community. And so we're also finding that there are many people that are involved with cyber infrastructure. And so a lot of them haven't been talking to each other. For

example, researchers, that have been using technology for a great part, have been taking care of themselves. But what they're finding is that it's growing to the point where, you know, they would like somebody else to take care of it. It's becoming complicated and such. So, CIDs is intended to bring together these groups to start having the conversation on the campus so that we understand what the needs are for cyber infrastructure and what the rules are for everybody and we can get it working as a better system. So next. So cyber infrastructure kind of got started because NSF said technology is an important part of doing research these days. And research has been changing. What we're seeing is it's no longer the researcher sitting in his lab doing his own work by himself, what we're seeing is that it's discipline groups working on fairly large projects and these groups are often made up of people from multiple institutions. And so they create these virtual organizations, and they need a mean so that they can work together smoothly because they are at all these different locations. And so cyber infrastructure is kind of intended to try and bridge the gap to bring them together electronically, where as they would normally do it physically. So next. There's also a new requirement that's coming up for all NSF grants that every proposal will have a data management plan. They haven't said what that is yet. It's supposed to be figured out by October, but they're recognizing that the data that goes into all these research projects is just as important as the papers that are published. And so how can we maintain that data, make it available to others and not lose that, that important information. There's a webpage on where you can find more information on that. So, But it's not just research. It's also a teaching that's changing. So we're seeing network based multimedia resources. You're pulling things into the classroom, you're having real time communications with your students or remote locations and a lot of classes. We're hearing about how they're using the clickers for communicating with the class as in real time. A lot of times you're having multiple locations. Sometimes you'll have one class that is beamed out to multiple locations to allow more students to benefit from a particular resource. And so you have the online classes, and there's also team teaching where you might bring in other colleagues to help teach a class. So we're seeing CI fit into teaching as well. Next.

[12:19 - 14:37] Russ Harvey: So also in administration. So all this new technology and such is having some common threads with what's going on in administration, for example, identity management. Not only do you have to have identity management for the administration, but now you're seeing it's more important in research. And rather than create separate systems to do that. Campuses are building a single system that can be used by all. And the federated identity is starting to move out to where your identity can be used at other resources, at other universities. So the administrative side is making use of the CI as well. So next. So I want to go over a little bit of an example of a researcher and how CI plays in this fictitious researcher's life. When we talk about Jane, who's a researcher that's working on global warming, she has a lot of things that she wants to

do, collect data, do simulations and such. And so just to give you an idea of how she might use CI. I want to go through some steps that she goes through. So next. So she needs to collect information on the local climate. So she has her own research site. She collects the data and needs to get it stored someplace so she can do later analysis. Next. There's also other researchers that are doing a lot of work elsewhere, and then they want to collect their data. And so all this data collection is going on and feeds into a common repository. Next. Okay, now that they've collected all the data, they want to start doing some visualization. What does this data really mean? So Jane starts using visualization tools and doing some analysis to figure out what's going on and see what the trends are. In doing that, she starts coming up with the ideas on some simulations. So next. She starts doing some climate modeling using computer resources. She might start out with the small computers in her own lab. But then, as her needs, computing needs grows, you might move to a campus cluster or move to some of the national resources. Next.

[14:38 - 17:35] Russ Harvey: Okay. She's come up with a model now and she wants to understand what this model means. And so there's the visualization that goes into a better understanding what her model does. So all the computing, the visualization and CI resources go into helping her do that. Next. She's working with a team that's trying to understand the environment. It's a big problem that takes lots of people to work together. So she wants to collaborate with all her friends. And so she does that in multiple ways with video conferencing as an example of the access grid where people can all talk together. There's also the shared applications. There's all kinds of ways that they can communicate, trying to make it like they're all in the same room, even though they are at various locations. So next. [clicks tongue] So that gives you an idea of what she's trying to do. But what is this CI stuff? Well, you can start thinking about it like a car where someday you want the car just to take you to where you want to go, and you'll tell it and it'll go there. Well, we aren't quite there yet. So you need to understand what the how the car works to some degree. You need to know there's an engine, and you got to put fuel in, all these various things, but it's likewise with CI right now, is that, it doesn't quite do everything automatically for you. So you have to understand a little bit about what's going on under the hood. So I'm going to go through a little bit of what Jane has done and talk and talk about the parts that that are required. [paper rustles] So next. Okay. So remember first Jane wanted to collect information, collect her data. So she has these remote instruments. She wants to tell the instruments to start collecting this particular data. Okay. That would be fine except for... how do the instruments know it's Jane that's doing that? So next. What we need to add in is management of security and access. So you end up with a system that says it can verify. Yes, it's Jane. The instruments can say, "I trust the management system." So I agree that this is Jane. Likewise, when it starts sending the information out, they need to know that it's Jane, that it's going to. [taps] So, the access and management of security, it ends up being an important part of all aspects of CI. Next. So

now that the data has been collected, she wants to get it into a repository. So she tells the collection to send it to a particular data set. You know, it might be on her campus. It might be a national repository.

[17:35 - 19:38] Russ Harvey: You know, it could be a number of locations. It has to go through the security aspects again, because the repository wants to know that it's Jane and that all these things are okay. So I won't mention security again, but remember that it's a part of every operation. So we get the data into the data repository. Next. And she wants to start doing some analysis. So then the computation starts connecting into the repository to do the analysis. So she ends up understanding the data. And then next. She wants to see what what was happening. Remember the visualization? So now, the computation has been done. It stores some results into the data sets, and then it can go into the visualization side to understand what this data is really telling. Next. Okay. She wants to do her simulation now, and so that will be done in combination with data on the data set, the computation and the visualization, all working together to understand her model and see what's working right and what's working wrong, and how it compares to the actual data that she's created. Next. Okay, so we've gotten into a lot of the computing and technical side of things, but there's also the human side. [paper rustles] So Jane wants to do some publishing. So she needs to reach out and use CI in order to do that publishing. She wants to use those collaboration tools so she can talk to each other. She wants to do outreach in education. She wants to tell the K-12 kids about what she's learned about the environment. So all those things end up feeding into the human side of CI. If she needs help, there needs to be the help desk available. So next. So, okay. [laughs]

[19:38 - 21:37] Russ Harvey: So Jane is going after a new grant. And what she has to do is start working with policy and funding agencies. And so they fit into the CI picture, even though they aren't part of the technology, they are an important part. And depending on how you draw your bubble around around CI, you might consider them CI or not. But, you know, it is an important part. So Jane goes after her new grant. So, next. So this kind of gives you one picture of the various components of cyber infrastructure. But there's still one thing that's missing. Next. There's the network. The network ties all these things together. Now, so for the most part, researchers and people that use the cyber infrastructure don't want to know about the network. They just expect it to work. But you need to understand that it is there and what to do if things don't seem to be working correctly. So, but the network is really what makes it, allows you to work at these various locations seamlessly or try to make it seamlessly. So in addition to all the parts of cyber infrastructure, there's also all the different players. And so this kind of gives you a picture of different groups that are involved. And I'm sure you can find yourself in these groups up here. So it's the researchers, there's the campus IT, the security folks and network folks, supercomputer sites, grid organizations, discipline groups, the policy

collections organizations. And so all these have to start working together to create a seamless CI. And so that's part of what CIDs is about, is to try and bring together all these folks to start understanding what the rules are, what the needs are, and how we can improve cyber infrastructure.

[21:38 - 23:37] Russ Harvey: Okay, so there is other ways you can look at cyber [inaudible]. For example, here's looking at it from the instructor point of view. So you can redraw these diagrams to whatever seems to meet your needs. You got instructors, you got students and and and such. But still a lot of baseline infrastructure is the stuff they are going to be using. Okay, a little bit about some of the components. A lot of you might think, mentioned that network is something that you probably don't want to care about. But, it's important to you. So just so that you understand it, when we talk about network, it's not just one network. It's actually multiple networks connected together. So you have got national backbones that that connect various regional networks together. You have regional networks that connect to your campus network and then eventually connects to your desktop, your resources and such. So when you start thinking about the network, think about the fact that it's your campus, it's your region, it's your national backbone, and then connection to your other situation that are in other countries as well. So the international connections. Computing resources can come from multiple locations. It can come from your own desktop, it can come from shared resources that might be shared clusters on campuses, maybe regional shared resources, or national centres like like TeraGrid and the Open Science Grid. There are going to be private and and clusters that can be shared clusters. So there is lot of different ways to cut computing. When it's starting to come into the picture now is cloud computing, where it is entirely someone else's commercial entity that's providing the resources. So we need to figure out how and when to use the appropriate resources.

[23:38 - 25:49] Russ Harvey: Storage is similar in that... storage can be at multiple locations. You need to figure out where the appropriate location is. You know, it could be in your lab, it could be in a shared facility on campus, or it could be regional. A lot of times it depends on where you're doing a computing. Do you want to have the the storage close to where you're doing the computing? You need to have it close to the researchers that are working with it. So a lot of different decisions go into that. Okay. So far we've talked about a lot about, the hardware side of things, and how do you pull all that together. Well, the researcher doesn't want to know about all the details that's going on, so it's nice to be able to try and hide those as much as possible. So there are gateways or science gateways that make it easier to interface with all these resources. A lot of times, a discipline will put together their own gateway for their researchers, and it's put in a view that's easily understood by that particular discipline. So it's just mode that they work in. Common software, being able to use common software. And then probably one that's, to me is really important, are the common data

formats and schemas. A lot of times I've seen researchers that will start a project and they say, "Okay, we're going to start collecting all this data. What format are we going to use?" Oh, well, we'll just make it up, which works fine for them as long as they, you know, they're just using internally. But then all of a sudden they want to start sharing it with somebody else. And it's maybe not quite in the format that they need. And so, you know, worrying about translations, if that's possible. So one of the things I've always told the researchers is go see the librarians. Librarians know how to put this stuff together. Performance and monitoring tools. And the cyber infrastructure today doesn't always work the way we'd like it to. So how do we know what's going on? How can we fix it? And who can we tell what the problem is when we have a problem? So these kind of things need to be built into cyber infrastructure not added in as an afterthought.

[25:55 - 29:20] Russ Harvey: [remote clicks] There we go. So I wanted to do.. I have this vision for cyber infrastructure. And these are the things that I think cyber infrastructure should do. So number one is to have a computation and storage, easy to use and so that you can easily go from your own lab to a campus resource to a national resource, as your requirements, your power requirements go up. That's not particularly easy these days. A lot of times when you go from one realm to another, you end up having to recode a lot. You have to structure things for the new environment. So if we had a common way that you can do your work in each location that would help that migration. Data repositories in formats and locations that are easy. I mean, a lot of you are out there, librarians, you understand this issue. How can we make the repositories easy for people to contribute to? I was talking this morning with some folks about how much data is actually lost from research because a researcher will collect all his data, do his research, write his paper, and the actual data just sits in his lab until someone throws away the computer. So how can we keep that data from being lost? Tools so that people can easily create their analysis and simulation environments. So right now, there are tools around, but you really have to become an expert in particular set of software. And if you want to try something different, then you have to become an expert in that new software. If we could come up with some kind of easy user interface that people could [thuds] you know, maybe a graphic interface where you can just go and start plugging boxes into each other for different functions. You know, something like that would make it easier for people to use cyber infrastructure without necessarily having to become experts in it. Okay. And then the collaboration tools. I think it's an area that still needs a lot of work. I mean, I'm sure most of you have used video conferencing at some point, and to be honest, it's often painful, is you have to have a technical person there to get things set up. Your meeting starts ten minutes late because they had to get things working, you know. A lot of people aren't going to tolerate that, and understandably so we need to make these tools easier to use and more reliable. So I wanted to save some time for some discussion and find out from you some of these questions. What makes cyber infrastructure hard for you to use? What could make it

easier? And what could make it more useful? [chair scrapes] So what... just jump out there. Who has had a problem with cyber infrastructure that they think needs to be solved? [chair scrapes]

[29:24 - 30:10] Man 1: Before I will tell you, I'm very old. [chair scraping] And so I used to sit around with people the way [microphone feedback] [inaudible]. [chair scraping] And when you put your bullets at the end was about the video [door locks] conferencing. It'd be nice if there were... I guess there are some tools that I think I'm not familiar with me yet. I'm used to make hair on a whiteboard where I can throw a picture wherein... yeah. Where I could have a whiteboard and meet with people in a room to brainstorm and I could draw a diagram. They can go over and write on that same diagram and change something. And I think tools like that, that can be used easily as a part of video conferencing would help make that part of the cyber infrastructure easier for me to adapt to.

[30:11 - 30:19] Russ Harvey: So a multi-location whiteboard would be a plus. One in the backpack there.

[30:19 - 30:27] Jeff Bullington: In order to keep everything pulled together for the filming of the session, I'm going to have to run around with the mic or figure out another way to get to the question more.

[30:29 - 30:58] Man 2: So the technology is ever advancing very quickly, but one of the things that we struggle with all the time is the policy and the legal space. You mentioned cloud computing. How do we deal with that if we're going offsite with some of these things? And how has policy making policies, making laws kept up with the pace of making changes in technology? I think you know that policy making pieces is something that's very critical in all of this.

[30:59 - 31:18] Russ Harvey: Yeah, I agree policy... a lot of times people start thinking about policy for technology and think it has to be something that's entirely new and different, but a lot of times you can adapt the existing policy [clears throat] to the cyber infrastructure environment. But you know, but it needs to be done.

[31:21 - 32:08] Woman 1: Hi. One of the issues that I think that you hit upon was the data that is lost when it sits on somebody's hard drive, and eventually the hard drive is thrown out. I'm wondering about whether the NSF, new NSF dictum will really make a difference with this? And the reuse of data is another issue, um, once it's gathered and how... and, you know, the format is an issue with reuse of data. And I'm just thinking about if libraries get into the role of being a repository for data for like NSF granted projects. How to deal with this sort of thing like format.

[32:08 - 32:25] Russ Harvey: Yeah. Yeah. You think about that libraries have always been the repository for the papers after they've been written. You know, that's where your thesis goes. But now the data is as important part of that thesis. So, you know, it makes sense that the libraries would worry about that as well.

[32:29 - 33:02] Man 3: In a Kansas cyber infrastructure project that we have from from the F score program at NSF. One of our biggest challenges has been getting the research community and the IT community on the same track, which is to say, sometimes the CI challenges that the researchers have are rather boring to the IT people. And we have a real disconnect. It's been quite a challenge.

[33:04 - 33:31] Russ Harvey: Yeah. One thing that we did at UC Davis, my home campus is to have a person that ends up being the liaison to the researchers, to try and understand what is it that they're actually trying to accomplish and how technology can be applied to that. So you kind of need a bridge person in there so that the researchers don't have to become the experts in technology, but yet they can be able to start using the technology.

[33:38 - 34:16] Woman 2: Well, I think one thing that is a big part of it and sort of hard to figure out is the economics and the sustainability of it, because there are a lot of tools that are great but not everyone can afford them. And then even if you can afford to get them once, you have to maintain them. And how much should be supplied by private companies who then own the software, you know, and the code. And how many places have people who can really work with the open source stuff. I mean, all of that is a big piece.

[34:17 - 34:25] Russ Harvey: Yeah. How often have you seen great code developed by a grad student? And then when that grad student goes elsewhere, you know, it's essentially abandoned, so.

[34:27 - 35:04] Man 4: So a problem that's emerging is what I call the data diarrhea problem. [laughs] We don't yet have the infrastructure to store large files, to backup large files, [man coughs] to transport large files. I've had a physicist knocking on my door saying, "I got this petabyte file that I need to transfer from Colorado State University to Stanford Linear Accelerator Center. How do I do that?" And I recommend a station wagon full of thumb drives [laughs] heading west. So we just have way, way TMI, too much information [laughs] these days and it's a challenge. [woman laughs]

[35:10 - 35:31] Russ Harvey: Other comments? Well, there are a lot of these things where things will be tried to address today. And so hopefully by the end of today, we'll have some action items that will allow us to, or allow you to make some progress on these issues.

[35:32 - 35:32] Jeff Bullington: Thank you very much, Russ. [applause]

END TRANSCRIPTION