

# Duke Office of Information Technology

## Minutes

**Jan. 11, 2007**

John Board, Tammy Closs, Ken Hirsh for Dick Danner, Brian Eder, Matt Miller for Nevin Fouts, Tracy Futhey, Susan Gerbeth-Jones, Billy Herndon, Rick Hoyle, Bob Newlin for David Jamieson-Drake, Andy Keck for Roger Loyd, Dan Murphy, Tim Bounds for Caroline Nisbet, George Oberlander, Lynne O'Brien, Mike Pickett, Rafael Rodriguez, Molly Tamarkin, Christopher Timmins, Trey Turner III, Tom Wall, Robert Wolpert

Guests: Kevin Davis, Klara Jelinkova and Rob Carter, OIT; Tim Lenoir, Jenkins Chair

Start time : 4:06

### I. Google video update - Bob Price

Bob Price – When I came to Duke this deal with Google was under wraps, but everyone knew about it. We were looking at an opportunity to present Duke content through Google video. Duke has reached the level of content that Google's happy with, so we'll have a Google video page. The content, art and text are all done. We're working with designers to get the page live. We'll continue to look at getting more content. It's part of a larger conversation we need to engage in related to digital assets and their value and the reasons why we'd like to present it. I would like to come back to ITAC when the page is up.

John Board – Who will own the disk drive where the content is sitting? Bob – The content itself is hosted by Google. John – They will make backups, but would we archive the information here? Bob – We will have a repository of content that we're uploading from. So every piece of media we have on site as well.

Question – Is this content available to anyone? Bob – It's open, not authenticated.

Question – Who chooses the content? Bob – At the moment, all content is being filtered through me once faculty and other presenters have signed off on their materials.

Molly Tamarkin – Is this really under wraps, or can we discuss it? Tracy – I don't know. I think the sensitivity was about some very specific things in terms of what they deploy. Last year Stanford went ahead and announced their page, so I don't think there needs to be anything that's real secret about the fact that Google provides video. Bob – An announcement went out in the OIT Inbox already.

John – That sounds exciting and it will solve a number of problems for us in terms of access.

Question – Once the content is up, will it be available to search throughout Google?

Bob – Yes, and they'll provide a link to our content, so they'll drive some traffic to our links.

Text

## II. Storage, Backup and Recovery Service Team Report – Kevin Davis and Billy Herndon

Billy Herndon – Last July we established a service team to look at storage, backup and recovery. There were several roles involved and I was sponsor. Kevin was team lead and we had others, including financial, technical, applications, operations and support people. We asked them to focus on services and needs in the Duke community in the next 18 to 24 months. Also Mike Pickett and Ginny Cake were looking further out, at a five-year horizon. The team met numerous times with groups on campus, various departments and schools. We also met with student government organizations. In August Kevin presented back to the OIT senior leadership about 13 services that were identified. In the fall, Kevin came and did an update for ITAC. In September and October, we tried to avoid getting technology involved and just tried to understand the community's needs. Then we looked at the technology needs and collaborative efforts we had with the health system. In late November, the team provided a final report with recommendations from their perspective. Now we've identified three priorities. The goal is to deliver these three services by summer. Over two years, hopefully, we'll deliver all 13.

Kevin – The real update is the technical discussions we've had with the health system on technical feasibility, trying to get an idea of how difficult these needs would be to solve and provision. Looking at the discussions we had, we identified the priorities. All were important, some for later services. Others there's nothing out there that's being delivered.

(Kevin distributes recommendations.) The first priority is improvements in personal individual data storage. This primarily benefits undergrads and grad students, but others in the community, too. The existing personal data storage isn't enough to work with. People are dealing with bigger and bigger files, many with collaborative features.

The second priority is backup services around personally owned, personally managed computers and laptops. So much data is tied up on these devices and there really aren't easy backup capabilities. There are some good examples at other universities, including fee-based systems.

The third priority is hosted data storage. Most units have provisions for data storage, but need help with big projects needing to store terabytes and terabytes of data.

There were four themes that ran through the group: the frequency and tenor of the suggestions; opportunities to define services that would go to the broad campus community, for instance it came up with departmental IT leaders that they may not need backup now, so the ability to focus on things that touch individuals and keep us from overlapping services; we also have project teams being set up within OIT to look at services, using examples from other schools and other places. Currently we're recruiting for a senior project manager whose job would be to oversee some of these projects.

Molly Tamarkin – Is it fair to say the hosted storage is likely to be some fee for service?

Billy – We've looked and we think some will be just offered, and some will be for fee. We haven't gotten into that level of detail yet.

Molly – The personal network storage and personal computer backup, it seems like by making them the same priority, I just see those as similar. What's the message? Sometimes people store things centrally and that's backed up.

Kevin – A big reason is that they really don't have overlapping populations. This sort of personal file storage doesn't replace what departments are offering. This would be used more by undergrads and graduate students. The personal backup services are usually on a subscription basis where you pay by how much you store. They have version control. That's been more for faculty and staff than students.

Tracy – The priorities were not identified as the most important thing but as the most realistic or some combination of realistic and value and getting some quick credibility.

Billy – Right, the technology could be delivered quickly. Kevin – What we're hearing is that in the next 12 to 24 months, all of these are needed.

Rafael Rodriguez – What do you mean by personal backup. You're expecting a level of integrity that if it's not there, there's a mismatch between expectations and reality. Kevin

– Obviously backup strategies vary. It could be 30 days, but for students it could be seven to 14 days. This is more to guide the project team. It's not necessarily the same as a department would offer a staff member.

Question – You're talking about historical backup over time? Kevin – Exactly.

Robert Wolpert – We could take hits for two things. First, the realistic cost of operating storage space is much higher than the cost of buying a disk drive. Students tend to view offers like this as a sign that we're out of touch with the market. Basically you're selling the backup. Second, if you lose the last week of a student's work and that happens to be the paper they're working on, you can get a black eye very quickly.

Kevin – One conversation that came up a lot with students was the need for collaborative projects. They're interested in using a service like this to get five or seven people together to have access to these files.

Rafael – How do you distinguish this from other collaborations like wikis?

Molly – That need is why I was wondering why that team storage was a medium priority.

Billy – The priority isn't based on needs. Molly – Right, but that's an important need right now.

Kevin – There are services out there, and others that would provide more personal storage and more collaboration. Some solutions could bubble up to the top quickly. So yes, in some ways it is similar to a wiki.

Molly – We talked about in the absence of a file system, a wiki would become that.

George Oberlander – Do you feel confident, when we're talking about backing up personal computers and laptops, that people will be able to understand and follow directions for a complete backup. Particularly when people feel the need to put lots of data on their large drives, they need a complete backup. If there's confusion that this mini-backup could serve as a replacement for a complete background, people may not be best served by just doing the mini-backup. Kevin – A lot will come back to how the services are explained. We have to be careful. There's no expectation that this will replace departmental storage and take advantage of this, but for a general backup for everyone with a very clearly specified timeline of how backups would work for people who want it. There are definitely two very different services and we have to be clear about how they are differentiated.

John – One way is to do this is a disk clone of every computer; but the other thing is that mine and Molly's PCs are 80 percent the same.

Kevin – The issue of saving things by kind of object has been brought up over and over. We've been looking at various kinds of services. We'll need to talk about that.

Billy – One of the potential vendors goes beyond just compression. Well over 50 percent common that we need to talk about backing up.

Dan Murphy – There also are archiving sessions, where students have asked others to join a session, then they want everyone to have hard copies.

Robert – Of course there are privacy and security issues when you're gathering together hundreds of people's stuff. Kevin – That's always there. Through clients there is encryption and password protection. Tracy – Unless what you're backing up is the system files. John – It takes a really sophisticated user to know where the important data are. Billy – We understand that there has to be a real communication plan in order to clarify what we're talking about, and what the true cost is and so people understand what they're getting. Robert – And still expect understanding. Billy – Yes, and that's why we wanted to pick the simpler things. And we wanted to find some customers who are willing to work with us to get started.

John – What's your timeline to making a formal decision? Billy – It's a goal to have these top three out in summer 2007. We're trying to get project teams settled and operational costs into the budget. Hopefully in a few months we'll be back to give more detail. Kevin – This group (ITAC) will be very important in helping the project team.

### III. Duke Email spam handling options – Klara Jelinkova and Tammy Closs

Tracy Futhey – The report you have is more of what we use within OIT, but because this is a service we made a change to last fall in an abnormal way in terms of a policy change to eliminate spam, we thought it was important to come back to ITAC to review the history and set some of the context.

Klara Jelinkova – We are not coming to you to tell you there is an emergency. We want to open a dialog in how we are going to work on an issue that isn't going to go away. We're getting about 2 million messages a day and they're growing by about 5 percent. The question is, how do we handle this ever-increasing load. How do we do it effectively, and not just by throwing hardware at it. Where we are seeing this spam hitting us is at the MX routing boxes and at the post office. Handling the volume of mail is always going to happen at the mail routing place. For each message, it has to throw it away, put it in junk mail folder or put it through to the post office and into people's mailboxes. We can do combinations of these three options. Currently if there's a spam rating over 96 points we discard it. The first thing we can do is increase the discard rate to 94 points. We heard from Rafael Rodriguez last time that he's put his at 25 points. Rafael Rodriguez – I had to go to 30. Klara – The other strategy is introducing quarantine. The antivirus rating happens on every message. We can discard or deliver. This delivers to quarantine. The user gets a daily summary of what's quarantined and they can go get it. The policy says they're deleted in seven days. We can set the thing to 90 points, but users still have the ability to set rating lower and send things to their junk mail folder. We also could have spam filters. They're subscription services and we could blacklist the addresses of the known spammers. There are legal challenges to this strategy from the spammers. Robert Wolpert – Also there are false positives. Klara – There's a liability there. The recommendation is to continue discarding messages with 96 or above and also to introduce quarantines for spam messages with 90 or above. Not deliver, just notify the users. Quarantine at the routing boxes and users will have to retrieve them. We could evaluate maybe three months from now, should we drop the

discard rate to 90 points, the quarantine rate to 80 points? We'll have some data for discussion then.

John Board – The amount that's between 90 and 96 is small and a large amount of spam still comes through. A tiny fraction of spam is being put into this special treatment.

Tammy Closs – This is a recommendation and we're trying to balance user's concern that we'll throw away good stuff. By putting it into a quarantine they can go check for what they want. If we want we could say we want the quarantine at 80 percent at the beginning.

Robert – Can that be user configurable? Tammy – Not the quarantine.

Rafael – The health system will be part of what gets thrown away. We already quarantine. I'm afraid we'll end up with two quarantine places and people will have a hard time figuring out where to go. Our approach is to keep it at the server and people go there to retrieve them and add them to their own blacklist or whitelist. Dan Murphy – We use this and it does give people a sense of confidence. I agree that a couple of months of data plus a couple of months of people getting used to it will be strong PR. John – Is there any quantitative data on what the false positive rate is? Rafael – When I took mine to 50 I didn't have any. When I ended up at 25, I was getting some false positives, and a lot of stuff I didn't want to get. I went back up to 30. My guess is between 25 and 40 some false positives were being caught. The amount of spam that comes in between 30 and 50 is small – 50 and above, I've never seen anything over 50 that isn't spam, and that's over years of writing the Notes quarantine.

Klara – Usually people quarantine at the server level, but people determine junk mail at the client level. People are already configured for the junk folder, so that'll be hard to take away. So we're suggesting adding a central quarantine server.

Rafael – I would be happy to say let's use one central quarantine, and I'd like to have the ability for users to set the level of quarantine they want.

John – At what level do individual blacklists and whitelists kick in? If I whitelist something at 96, it gets thrown away anyway. What is your provision for user-specified?

Tracy – I can go online and say I want low, medium or high spam treated in certain ways and also specify specific people who can get through. Rafael – Is this quarantine global? Klara – It's global. Rafael – We have done it at the individual level. It's still done on the server, but individuals set it up. There's no delivery to the client. There's an initial default, and whitelist and blacklist are individually configured at the server level.

Ken Hirsh – We do it pretty much the same way in the Law School as Rafael is talking about.

Klara – So quarantine is a good thing, but you'd like to see the quarantines at the individual level rather than the global level. Rafael – And my preference would be not to have two quarantines. We do it at post office level.

Dan – For PR, we should make sure that if there are false positives it's because of the client, not the server.

John – We could have volunteers check their spam to see what gets through at different levels, so we could say we've never seen a spam message above X.

Klara – We came today because we wanted to have a discussion as to how to proceed. You would like a central quarantine service that can be individually configured. We will figure out how to do it and we will come back.

Statement – Your concerns about Spamhaus are worth noting. We use one of the other services. In our shop there's concern about the CPU load of having to check every single message. In our experience it adds a significant amount of time to delivery.

Tammy – We looked at it and recognized that sometimes Duke gets blacklisted and we didn't want to blacklist ourselves from ourselves. We didn't feel like we were ready to implement that. Some of the other services might be worth pursuing.

Ken – We have to build into this a process where people can appeal to you.

John – My graph of spam messages plunged after Dec. 17 and it hasn't gone back up.

Tracy – That'll be next week.

Molly – Is there a tool that dynamically blocks something once it's been identified, so that origin is blocked for X number of hours or days? The addresses are changed so quickly, so can that be done dynamically?

Klara – One of problems with blacklisting is that legitimate people and messages get blacklisted. I think it's not for the Duke services. I think there are better solutions to this problem.

#### **IV. Update from attendees to the Common Solutions Group meeting – Mike Pickett**

Mike Pickett – The Common Solutions Group includes CIOs and some tech staff in a three-day meeting. The first day and a half is workshops. Then we break into the official CSG meeting and there are policy and logistics discussions. We met at the University of Southern California. The three workshops were on research computing, managed storage and collaborative tools. They reinforced that most of the universities around the country are struggling with the same things we are talking about. Research computing – we talked about power facilities, data center issues, costs, business models for working with schools and faculty, where high-performance computing is going. (shows some examples)

Tammy Closs – A couple of surveys were presented. They're on the CSG website at [Stonesoup.org](http://Stonesoup.org). Research computing was overall, but there is lots of documentation there, including two surveys.

Mike – In terms of governance, 78 percent of schools said they think having a governance oversight group was a good thing. Top issues for data center facilities, need upgrades, central control isn't desired by researchers, 15 percent felt they were serving the needs of the campus, most folks felt like their networks were good enough for what they needed. Also, the fastest areas of growth – staff monitoring facilities and power would be main model for providing research support; central computing provides a place to put the computers. Less than half believed central services would include support technology for grant development. Factors influencing central data center use – cost, control is becoming the distant second. Used to be a lot of need for machine-hugging, but as cost goes up, that goes down. Addressing regulatory change – 27 percent are prepared today for this pending change (grants are requiring us to take care of data. How long until this falls onto the schools?). 54 percent are starting discussion; 19 percent aren't even thinking about it. (Discussion on staff for research computing in central computing group, and large budget as implemented at Indiana University. Also discussion on discipline-specific needs for research and backup.) We talked about collaborative suites – which ones do universities have installed? Sakai,

Oracle, Sharepoint, others. Also talked about collaborative tools – email, file sharing, calendaring, podcasting, wikis, surveys, institutional repository. Tracy – This is what people consider to be in category or what people are providing? Mike – This is what people are providing. The thing that was interesting was that we realized how wide that description of collaborative tools is; it is something that changes radically, and it isn't unusual for central organizations to bring up a tool and find out it's out of date. A lot of other tools, wikis, are popular. Seeing small pieces of wikis, collaborative writing, Robert – Molly pointed out that if we do that before we offer storage people will use it for storage. But so what?

Klara – One thing not reflected in the graph when you think about collaborations, they are collaborations on a national level. When you have multi-institutional grants you use email because it's easiest where you can keep you identity.

Mike – We also had a discussion about Open AFS and the call for collaborative contributions. It moved into, where is AFS going? Several schools said they're coming off of it. Duke is dithering, managers are watching and wondering. It served a purpose, but other tools may be pushing it a little bit. There was some discussion about how new features of Open AFS get prioritized. We didn't leave with any particular direction.

Tammy – We do have a group that's coming together. We've hired Chris Callum from UNC, and one of his first goals is to talk about file service systems and how we need to talk about that going forward.

Mike – We also talked about digital rights management as a policy issue. Basically we heard, be conservative in what you limit and liberal in what you allow. Ensure access for several generations, for cell phones, iPods, Zooms. Digital rights management is just as primitive as our discussions on how to use these media.

## **VI. Other Business**

John Board – Later in the semester the provost will visit us, so if you have specific items you want to raise, give him notice.

Rafael Rodriguez – The start of Daylight Savings Time has changed, so start doing an inventory of how that will affect your computing systems. It starts March 11. Molly Tamarkin – FYI, OIT has a nice project team looking at this, as a resource for what may need patching.

End time : 5:26

Duke Office of Information Technology - [www.oit.duke.edu](http://www.oit.duke.edu) - (919) 684-2200 - [help@oit.duke.edu](mailto:help@oit.duke.edu)