

Directory Group Practices Survey

Group interview 9/13/01

Interviewers:

Tom Barton (chair), University of Memphis
Michael Gettes, Georgetown University
Keith Hazelton, University of Wisconsin, Madison

Listed in order of participation:

Tom Dopirak, Carnegie Mellon University
Wes Craig, University of Michigan
Steven Carmody, Brown University
John Ballum, Brown University

Tom B

(Introduction): This is Tom Barton, there are two Toms in this call. Our objective with this call is to learn more from you folks who are...we hope are more practiced and experienced; we can tell you are. Although I do have a group-practice survey to serve as a handrail, I mean for this to be more open-ended. I don't mean for it to be constrained to the order and content of that survey. I'm just curious, has everyone had a chance to look at that at some point in time?

Michael: Has anyone not?

Wes: I haven't seen the document you are referring to.

Tom B: Wes, would you like to receive them right now?

Wes: That would be great.

Tom B: E-mail, please. Ellen will you forward him one, please?

Ellen: I will; when I initially asked you to participate in this discussion I sent it, do you still have that message? Or, would you like me to send it again?

Wes: No, I don't have that message any more.

Ellen: Ok, I will resend it.

Tom B: Thank you Ellen. In any case, that is just a guide. And I mentioned that Keith and Michael are here, also on the interviewer side of the table as it were, along with me. So, I'll hopefully keep things going and everyone should feel free to kick in and respond in some order. Ok? Fair enough?

Keith: Fair enough.

Tom B: All right. I'll guess I'll begin at the beginning, which is a question about how your group membership is represented, or perhaps to be represented; and also let that be open-ended. How did you choose to represent your groups? And you might say why is particular. Otherwise, for this question and for many, I think it is interesting to relate how the nature of planned and deployed applications in a constrained, the choices that you have available to you. So we could go around in a specified order, or we could just let the interviewees as Tom D., Steve and Wes pick it up by themselves, if you like. I'll start with Tom.

Tom D: We're not very far along in our deployment of the directory; applications there have been white pages, and computer-accounts management, and basic E-mail forwarding, and things like that. And so where we are in the management of groups falls from that a little bit from our point of view there. So, I would say today, we have what I think you described as dynamic groups, or groups that be obtained by essentially searching things that exist in the directory today, because we have, you know, college department and the person standing, and how long they've been with the University, and stuff like that. So, those are things that we currently have and we use them actually pretty lightly. We're not convinced that that quality of data is all that good yet, because we're less than six months into our active deployment – and we're still fixing things.

Various applications that we've planned on supporting the extra, they are going to force us to do a few other things; and, you know, a group object like group of names comes with open LDAP – which is the directory service that we use. And, our expectation is to use that to support class lists mostly. And then other kinds of things that we can derive from the data that we get from the University feed. And that's pretty much the nature of our plan; it's groups that we can derive from University data groups that we're going to build, and any other group that's going to be taken care of in some other way.

As far as spatial groups go, because of the way our DIT is organized we have pieces of the tree set aside for computer account logins that we keep from the major parts of the University, like computer sign ins from the FCI and PFC, and people like that. And so there's kind of a natural group membership associated with the population of those parts of the tree. So I suppose that's spatial, but I don't know quite what we are going to use it for yet. I think that's all I have to say.

Michael: Tom, this is Michael. Do you have plans for the support of either ad hoc, or individual groups that will belong to individuals?

Tom D: At the moment, that's going to show up in other ways and not necessarily at the top of the directory. It'll be more of a provisioning problem. We run the Andrew File System quite a bit and there's a service that comes with that called PTS, which is a group mechanism. And I think that we'll allow that to continue to be the ad hoc group mechanism for the Andrew World. I suspect that in the AD world – although I don't really know this for sure – there's an equivalent service that we'll allow users to use. And then there's one that exists in our web-publishing world in the Major Domo world. And so I think that our plan – at least to this point – is to provision those with the groups that we can derive from University data, and then let the users use whatever mechanism exists in those applications.

Michael: And for the record, what directory servers are you running on your campus?

Tom D: All the ones that I know about are open LDAP, but there were a few instances in Netscape.

Michael: And is there a single enterprise directory?

Tom D: Servers in my point of view.

Michael: And that's open LDAP?

Tom D: That's open LDAP. But, the business of replicating data to other not-closely associated parts of the University hasn't really begun yet. We're still trying to get good data from them.

- Michael:** When the other two schools respond, if you could also give that basic information – that'd be good.
- Tom B:** Well, I suppose we could quick follow up with you Tom, before we go on to the others with this first question. This is Tom Barton speaking. You mentioned groups – those that are being automatically created and maintained from administrative feeds. And you also described some provisions for individually maintained groups. Is there anything in between that you, perhaps haven't done but, have been thinking about that would have more of an administrative function but be manually maintained? Possibly delegated to different administrative offices or others? Not necessarily personal affiliation, if you know what I mean.
- Tom D:** I can imagine that happening, but at the moment it gets into...it's more of a political problem and a problem of trust –of letting other people master data and allowing them to clean it up – then it is a technical problem. Although there are technical issues in the fact that we do end up in LDAP and the acl scheme is not as strong as I would like it to be. So, a delegation of responsibility on a piece of the tree is not all that straight forward.
- Tom B:** Right.
- Tom D:** I'm more concerned about – if I mentioned in my write-up – I'm a little more concerned about the privacy aspects than I am about delegation at some intermediate groups.
- Keith:** This is Keith. Tom, the followed question again, maybe extended to the other folks. You listed some apps – a fairly short list at the beginning that you are directory-enabling, one way or another. Do you have kind of a queue of apps at least that you're considering directory enablement for, further down the line?
- Tom D:** Well, I mean, there are some that are sort of partially enabled right now that we'd like to become more enable. We have Blackboard --it's a big motivator. And right now we just provide Blackboard with feeds. We also run Remedy to run our help desk. And Remedy actually queries the

directory directly, but we've had some difficulties there, in the combining those two together. Majordomo mailing lists is another application we'd like to support. (I can't) remember what's on my list I just wrote; give me one sec. General web server authorization; again I think probably maybe is an extension of a work we're doing with Pub Cookie, we'd like to look at. And then the other application I mentioned is provisioning of Active Directory, and then the Andrew PTS stuff. Those are things that we're looking at now. There's a lot of little things...

Keith: There's a big one in my mind, the calendar stuff.

Tom D: Honestly we haven't talked about that. It currently is...it's using an intermediate representation of the directory. It's actually using...right now we provision Andrew from the directory, and the calendar stuff actually takes its data from Andrew -- rather than directly from us.

Michael: But, it can be clear that your calendar environment is directory enabled. Yes?

Tom D: It is. It uses...

Michael: It may not be using the Enterprise Directory, but it's using an LDAP directory.

Tom D: It does. And it uses... it has some custom objects.

Michael: And we might call this an application-specific directory?

Tom D: That's absolutely what it is. It is moving toward not using, in fact LDAP, but using the proprietary data base that comes with it. And that's a scale issue, at the moment.

Tom B: Hmm. That was Michael, by the way, questioning. This is Tom making a comment. And how about Wes? Would you care to tackle this now? Do you need some questions?

Steve: This is Steve. I have one more question for...Tom. You said you listed several environments where you might be having groups: Andrew PTS, AD, and potentially even LDAP. Any thought on synchronizing group definitions across those environments for groups other than the ones you're maintaining with institutional feeds?

Tom D: We're not to that point, yet. I think that it'll be tricky enough to get , to do the institutional feeds. So, for example, today, I mean I don't think there is any mechanism today where we support groups within groups – unless it's inside of AD. So, it's still going to be a very flat schema of groups of essentially DNs, or, groups of DNs and mail addresses. I'm sure we'll get to it; it's just not in our immediate plans.

Michael: This is Michael. I guess one last question for Tom; and Wes and Steve – this would be for you as well. Tom, you're talking about how you're going to be taking institutional data and making it available in various ways. Are those homegrown tools? Is it product that you've bought to help make that happen? A short word on what it is that you're doing there.

Tom D: Well, they're almost exclusively homegrown tools. An awful lot of perl scripts and then a fair amount...one of the principal mechanisms we use is: most of the inbound feeds are processed with a lot of perl scripts and then a sleepy cat database. The directory itself, we've written essentially a data base trigger mechanism that processes the change log. And then that triggers off a whole slew of java and perl servelets that actually do a lot of the data integrity work. For example, if an entry gets added to the person part of our tree, that usually triggers the creation of an Andrew account of a certain class of people. And so that event sets off a trigger, which causes the servelet to run; and the servelet does much of the work. And so a lot of our applications are a venture event in that manner. And it's been a mixed blessing. Mostly due to the fact that data integrity and LDAP is not a wonderful thing. There is no real easy way to lock down different parts of the tree, so you've got to program very carefully.

Michael: If I may suggest, let's hear from Wes and if there's any one question, Tom D. can chime back in again, if need be.

Keith: That's Michael that's speaking.

Tom B: This is Tom Barton and I think we'll come back to each of the folks several times during this call, I hope. Wes, do you want to take it away? Or, do you want to be re-prompted, somehow?

Wes: I think I can just start. This is Wes Craig. We have...we used to have spatial groups in our tree, underneath the people branch of our tree. We used to have a separate branch for faculty and staff, and then additional branches beneath each of those trees to represent departments. We've eliminated that very recently. We collapsed the entire people branch of the tree to be entirely flat, and we include attributes in people's entries now that indicate their affiliation with the University -- which could theoretically be used to derive some sort of group information -- although that's not necessary. The only application that currently takes advantage of that is the white pages server; so, for instance, you could limit your search to those entries that have a particular affiliation, or a particular substring match of an affiliation.

Michael: So, Wes, this is Michael. Just to clarify, you used to have static group objects, and now you don't?

Wes: We used to have spatial information that could be used to derive, for instance, this list of people is in the College of Engineering.

Keith: Keith here, dif base and stuff, right.

Wes: Right. We've since collapsed that information, but we still maintain the affiliation's file information as attributes of person entries, directly. That's solved several problems for us, including the fact that people have -- typically have -- multiple affiliations in the University. So now you can list all of those in the person entries. We have a branch of our tree at the root called Misk, which is pretty much legacy at this point, and the new point of that is dubbed Security. And underneath that tree you have a number of groups that people can bind as according to what sort of authority they ought to have. So our acl structure in the directory is based around sub trees of the Security groups. So, this is another space where we actually have legitimate spatial group affiliations. The membership in those entries is represented through care V named attributes, where you heavily use the Kbind Umich extension to open LDAP. We use open LDAP to answer that question. We also have a groups branch. Currently

our groups branch has two sub-trees: one called System groups, and one called User groups. They both have very similar content. The System groups branch is for purposes of our email applications as higher precedence for lookups. So for instance, you would find the entry for the postmaster mail group under the System group branch; for the Root group, under the Systems group branch. If for instance a user were to create a similar group under the User group branch, the precedence would force their group to be effectively invisible to the applications that we have deployed. The User group branch is writeable by anyone who has an entry in the people branch. So we have a very rich environment, of -- for the most part, mailing lists -- both on and off campus -- run through that. On-campus user groups are represented with...the membership is represented with DN, off-campus members of groups are represented through RFC22 addresses. I think you would call that static. We can't have groups of groups; so, that is sort of dynamic, depending on how you look at it. We don't use LDAP URL's to speak of. Primarily we use DNs.

Michael: So can you explain: you can't have groups of groups? What does that really mean?

Wes: You CAN have groups of groups.

Michael: But in your particular case?

Wes: In our particular environment, members of groups can be...they do not all have to live under the People Tree, they could also live under the Systems group branch, or the User group branch, or anything to that effect. And applications that are able to interpret the groups will chase those references down and generate whatever activity they are trying to generate -- they'll generate the final list through forward reference through all of the groups.

Keith: I am wondering about the applications that are connected and how it seems, although I don't know; it seems like that puts a little more burden the way you're approaching it on the application to know where to look and so forth. But, I may just be missing it.

- Wes:** Yes, absolutely, it's certainly a case that if you want to have something like that. On the other hand, having taken the steps, say that groups are going to be represented by DNs, you're already in the position where you have to be able to forward referencing through the data base anyhow.
- Michael:** And the applications would be responsible for end-ing, and or-ing, and not-ing of groups together to come up with whatever membership necessary?
- Wes:** It's simply additive at this point. We have no plans; for instance, having a group of people to add, plus a group of people to delete, for instance.
- Tom B:** It's Barton, Wes: A couple of follow-ups. You mentioned using a lot of the user group for mail related. Are any of the system groups also mail related?
- Wes:** The System groups are mostly mail related. In either case, there are isolated applications that use them for access control as well.
- Tom B:** Are those groups used to synchronize out to a different kind of – like a Majordomo kind of a thing for doing the mail distribution? Or what happens there? You didn't say it was for mail directly, for example.
- Wes:** We're using Mail500. And SendMail called Mail500 directly, which is then a directory-enabled mail application. As I understand it, the whole use of RFC22 mail groups and the functionality that Mail500, and I believe it's now called MailDAP in the openLDAP distribution. It's not particularly well documented. Umich would sort of like to commit ourselves to writing up an Internet draft, or something, that describes what it is. During our recent collapse of the people tree, we were also looking the possibility of reorganizing our group branches as well, to modernize object classes and that sort of thing. But, it just doesn't seem like there's anything else out there that really represents the kind of functionality that we're accustomed to.
- Michael:** Besides Mail500, can you tell us the top few applications that are directory enabled?

Wes: Well, we use directory-enabled versions of Mulberry and Eudora on campus. We're in the process of planning a deployment to have the Caller ID system for the University's phone switch communicate with the directory. That would not be in the Group branch. That would be in the people branch. We have a number of custom admin clients. We're in the process of deploying a feeding system for Active Directory. We also are planning to encode NIS-like information into...again, the people branch of the tree to eliminate the...currently we distribute the University wide password file by making it available in the AFS, which is rather cumbersome.

Keith: Are there any big off-the-shelf commercial software packages, Peoplesoft -- things like that -- that you're looking at?

Wes: We have a Peoplesoft deployment on campus and it does have some LDAP functionality. We're at the VERY early stages of discussing the way in which that can incorporate into campus. It seems to me like in a lot of environments where directory infrastructure has come after something like Peoplesoft, that it's fairly obvious how you would integrate that into your directory-enabled environments. And since we've had an LDAP based directory since the early 90's, Peoplesoft is sort of a johnny-come-lately in that regard.

Michael: Do you get into that last statement, that you guys having been running LDAP for a long time and I know that you guys also have a lot of groups, and especially individual groups. Is that correct?

Groups that I, as an individual, would be able to create?

Wes: Absolutely.

Michael: So, from all of your experience, are there a top two or three things that you've learned that were mistakes that you'd want to make sure that others don't do?

Wes: Well, I would definitely suggest that if you are going to allow random users to make directory entries that you should get your access control correct.

- Tom B.:** That was Tom Barton laughing.
- Michael:** Could you elaborate on that a little bit?
- Wes:** Well, you know that...
- Michael:** And I'm not trying to embarrass, just make useful information out of this.
- Wes:** We have, for instance, discovered objects in the User group branch, the tree that have person-type object classes – which doesn't present any problem to the system. The way the open LDAP access control works. The way the Quipu access control used to work when we ran that. It didn't really allow us very conveniently to specify, for instance, what was creatable under a particular tree – just merely that things were creatable beneath the tree. So people could create arbitrary objects. When they were created, they typically became instantly inaccessible -- because the access control that we had in place and the access that we continue to have in place permits only those objects, which ought to be there. But it does make for cruft and it does make one interested in running tools to look for cruft, and get rid of it.
- Michael:** So, just to be clear, you have set things up to allow people to create groups, but apparently people have figured out how to create more than groups, but people entries and stuff like that.
- Wes:** Right. Although, like I said, having created them, they become instantly inaccessible to anyone except for an administrator, that's crufty.
- Keith:** I'm wondering if Tom D., if what we're talking about right now is what you referring to when you talked about your kind of caution and moving to self-managed other groups.
- Tom D.** That's exactly the problem we're worried about. That, and just getting the access control that's right. Right now we run a fairly simple suite. Our directory is really only writeable by a few people. We're still trying to get, I would say comfortable with the whole process and, you know, a relatively complicated piece of software.
- Tom B.:** I would like to follow-up also with a question, if I could. Is there permission to access control issues in regard to user groups. What about

name space issues? You did mention earlier that system groups take precedence over user groups if they have the same name, I guess. But are there any other issues that arise?

Wes: Well, the people branch of the tree for the mail applications has the highest precedence, followed by the System groups and then the User groups.

Michael: How do you control the precedence?

Wes: It's effectively a search path. For instance when a message comes in to the University for Wesley.Craig, for instance, the first search that is executed is under the People branch. When my entry is found, the mail goes to me. If it was sent to Westley Craig, it wouldn't match any of the people at the University and it would go into the Systems group, where it wouldn't find anything – and perhaps it would end up in the User group space, where somebody had created a name that was very similar. That's a potential problem.

Tom B: So, Wes, this is implemented in the Mail500?

Wes: That is a Mail500ism, yeah.

Tom B: So, if I were using some other directory client, I might be able to access multiple objects with the same name and different spatial locations in your directory.

Wes: One could do that, yeah. That's quite prevalent in the system, verses User groups base where many of the System groups were created after clever users figured out that some names were particularly popular. For example, quit, exit, stop were all early User group names that ended up eventually becoming System groups.

Tom B: Just to short circuit them, is that right?

Wes: Yeah, exactly, because...well, you know, we used to have this mainframe and it had a very similar setup for creating names, although it had no precedence – which is one of the reasons why we decided to do it the way we did. There was no difference in precedence between people and groups. So, for instance, two or three people could all be Wes Craig,

simultaneously, just by adding additional names for oneself into the database. We added the precedence setup just for that reason.

Michael: Are there any other strong recommendations of things to do, or not to do that you've learned with respect to groups?

Wes I suggest setting up early on, some expiration policies.

Michael: On the User groups, or groups in general?

Wes: In the User groups space; any space that you allow people to write to. We've just changed our policy in the People branch as well, to allow alumni who care to keep their people entry around. Keeping the people entry of course is tantamount to having...enables you to have the user name @umich.edu for as long as you'd like to have that. So, we're in the process of implementing an expiration scheme in that space as well. But in the group branch, particularly, we've just been doing some statistics on the last month five times of our group, and we've found that there are thousands of groups that haven't been modified since '95 or so. It doesn't really load the system particularly, but groups like that have become span target. And, when you get a very old group, it often has references to very old users who may or may not be around or if they still are in the tree or the data base at all – they may have references to email addresses that haven't been functional in a long time; so, that causes our postmasters to be very unhappy when they get the bounced mail, etc.

Tom B: Do you have any thoughts that you folks who share the call along what lines an expiration policy might take? What shape or what form it might take?

Wes: Well, at Umich the policy that we are working on is that all groups in the tree should be modified once a year. The fact that you've modified your group indicates that you want to keep it. We're developing an app that will traverse our nightly LDAP probably on a monthly basis, and the find the list of owners of groups that have any groups that have not been modified in over a year. And then we're taking a two-year policy at the two-year mark – then we would actually remove the messages after having

effectively sent 12 messages to the owners of the group to remind them that they ought to clean up their group lists.

Michael: Tom D do you want to comment?

Tom D: I haven't thought about doing that actually.

Michael: So, it sounds like a good idea to you?

Tom D: It sounds like a good idea. It's a little different than anything that we've done up to this point. Generally we would manage fairly carefully the changes in affiliations of people. And I had imagined that as a group became...and it really amounts to a modification time. I was just thinking that if all the people actually disappear from a group, that's probably a good time to delete it. But then you run into situations where, for example, a class list where all the people do disappear, but then it's repopulated immediately and the group is still valid. So, it may be that the group not changing for a while is a better log rhythm. I'd have to think about it a little bit.

Michael: I guess it would be an interesting question as to whether classes as groups, would they be considered to be system-level groups according to Wes, and therefore might be exempt from certain expirations?

Wes: Class lists are currently created by education staff in the User group branch. We're in the process of creating a project that we are going to introduce some additional group branches between System and User: a departmental branch, and a class list branch. And we can get data from the institution for every class. Once the class lists were substantiated, we would keep them up to date as we got through accounting and whatnot from the institution. The owners of the class lists would probably be the members of the education staff – TA's or Professors, or whomever. The departments were be very similar, although as it turns out the institution doesn't have good data on who the departments are.

Tom B: You tell me you have good data on payroll, but that's not the same as departments, is it?

Wes: No

Tom B: I want to call up on a different kind of a question. Early on, you mentioned how you allow Directory Administrators, I guess, to bind as certain security groups. Do you also then audit? How does the auditing prepare for that?

Wes: Typically, well we keep an audit trail for approximately a week at this point. Actually, depending on how you look at it, we keep the modified data; we keep a complete rep log actually for approximately three weeks if you include backups. But we're not actively scanning the data for impropriety. We're pretty much just waiting for the possibility that someone would complain.

Tom B: And one more, again. I'm not sure that you addressed referential integrity, if you're handling that, if so how you're doing so with the groups?

Wes: We used to get a lot of referential problems because people would move around in the People branch, because they would move between departments or whatever. We've eliminated quite a bit of that by collapsing the directory space, so a lot of the references are now much more static and much less of a problem. However, when we do our monthly -- or somewhat more frequently than monthly updates -- we also scan the directory for dangling references and delete them.

Michael: So, given what you just said, can you talk a little bit about your update schedule?

Wes: Sure. We receive data from a variety of sources, for the most part it's: institutional, payroll data, registrar data. We've recently added a data feed from the Alumni Association, and we have separate data feeds from our Dearborn campus and our Flint campus. For the most part, those feeds, in the past anyway, have come on approximately a monthly basis. The Registrar's Office occasionally comes more frequently, particularly during the beginning of the term. We take that data; we have a set of tools which we call YAM -- which stands Yet Another Munge -- in reference to the original munge that we wrote when we were getting all of our data off the mainframe. Speaking of tools, we have piece of custom code, a fairly

small piece of custom code that converts each one of those data feeds into an internal LDIF-like format. And then we have a somewhat more complicated tool that at this point has its own internal language that takes those LDIF-like update files and compares them to our backup LDIF from the server. That allows us to reduce the amount of traffic that will hit the live server to a fairly small trickle of actual changes, which is then run through the tool again, and then talks directly to the master and makes its updates live. For the most part, that's done on a monthly basis, except for the beginning of the term where it's done like a weekly or bi-weekly basis. Since the University has moved over to a Peoplesoft system, there's the strong likelihood that we'll be able to begin running this update on a weekly basis. And, we're in the process of talking with institutional MIS type organization to see if there is some possibility that they would actually run the updates, themselves, against the data – since effectively they are the data stewards, or somewhat like the data stewards. We are merely publishing on a LDAP interface.

Tom B: Wes, thanks. I think, looking at the clock it would be better for us to move along and give Steven a change to have some back and forth with us, and then we might have another round of questions for all three of you as well. Steven, do you want to be re-prompted or can you kind of pick it up from here?

Steve: There is another person from Brown on the call – John Ballem. John has recently been immersed in some of the group stuff. So, I'd like to ask him to correct any mistakes I make.

We don't store groups in LDAP, we store them in a piece of locally written software called Grouper, that's been around for five or six years. It maintains – it's not a general-purpose directory, all it does is maintain group memberships. Every group definition has a base expression, an optional base expression which could be other groups – say it could be a full Boolean expression referring to other groups. In addition, people authorized to do so could explicitly include or exclude people, principles.

Each group, in addition has six or seven kinds of acls associated with the control to see who could even know that the group exists, who could see the membership, who can edit various attributes, properties associated with the groups. Several of you know, I've indicated that we'd like to find a way to move this to LDAP. I'm not a particularly strong believer in creating and running local code. However, we found that the groups of group capability to very powerful and very useful. Tom Barton's been actively chasing how one might do that in LDAP. We currently have the usual set of applications that leverage this directory – that includes web access control. So there's the stereotypical locally developed web ISO solution that allows you to put statements in your HT access file that specifies which groups can access this particular web resource. There's a bulk mail service that is connected to some groups. We've been having three kinds of groups: one would be just the entire contents would be maintained via an institutional feed; a second kind of group would be a combination – it's base would be an institutional feed, but some set of people would be authorized to head it, or fine tune the membership adding or removing as necessary; a third category would be completely maintained by hand. We currently have – none of you have seen it so I guess I get a lot of slack here; there's a pseudo kind of dif structure in the naming. It's not enforced in a rigid way in such a thing, but it's emerged. Well, it's partially enforced. So, names are typically multiple. A series of fields separated by periods.

Michael: The name spacing in grouper – is that in any way associated with, or the same as any other name space, like the User name space, and so on?

Steven: Completely separate at this point and you would use a different kind of command or verb to refer to it.

Michael: So if you had to bring the group name space into the User name space, you may have conflicts?

Steven: Yes, it's potentially possible, but because we understand that's a long-term goal.

Michael: It's something you can resolve.

Steven: We've been careful to kind of keep a wall there. We've got, I don't know, six or ten, I guess what I would call high-level containers; it would be the first level of the tree. Some move in the group or name space. One of those is primary dumping ground for all the institutional feeds. And then, separate from that, we have a bucket that contains groups that represents big slices and dices of the entire University community. So if you wanted to email to all faculty, or all undergraduates, or the entire community, or the on-campus community, or medical students, or graduate students – all those kind of groups are in there. There's a second major container that's courses, and the idea there is that only the instructor and people authorized by the instructor will be able to see those groups; in addition the instructor will be able to edit the membership if they want. When we give it to them, the course group is based on feed from the registrar system, and the instructor cannot edit the group definition to remove that. John has recently finished writing the software that maintains this stuff. One of the things that he did that I really like is the gruels, the policies that have to do with the definition groups; and secondly how he aggregates groups into the base expression of that first bucket that I described – all faculty. And the acls that get applied to any of these groups when they get created are all specified the XML. So, when this program starts up and starts crunching, he reaches in his XML document base and uses that to control how he builds and defines all these other groups, based on that data in the institutional feeds. We started to think about departmental groups; we have a feed that can do that. We have data in the feed that would allow us to slice an academic department into pieces. So for instance for American Civilization we would know the faculty, we'd know the staff, we'd know the graduate students; we know all of those groupings separately. And our thinking is that the departments are going to require that we give a role in the department, the authority to define the base expression for the department -- which means that if department A wants to define itself as

faculty and staff, and department B wants to define itself as faculty, staff and grad students – we'll delegate the authority to the appropriate person to allow them to do that. In addition, they would be able to add and subtract individuals.

Michael: So in that particular case of a department, is the department initially populated from institutional data?

Steven: The groups that I mentioned, AmericanCiv.faculty.staff.gradstudent, would come from institutional data.

Michael: So then you would allow someone else to come in and create maybe an additional group that would add or subtract from the institutional data?

Steven: Well, they would be able to edit American Civilization.

Michael: If they go in and they edit American Civilization, then what becomes the system of record with respect to the grouping?

Steven: Well, American Civilization would be the thing that would be used by the calendaring system by Web access control by bulk mail. And American Civilization would be defined – as some Boolean expression that refers to the other things.

Keith: How does the persistence of that, like I'm imagining that the institutional data overwrites this at some period and you have to reapply?

Steven: That's where groups and groups really comes in, because the only thing that gets overwritten is the base. The edits -- the includes and excludes, if you will -- do not get overwritten.

Keith: Gotcha

Steven: We did the same thing with course groups.

Keith: Did you have some way of telling that a particular member of a group came from another source? That it was an edit as opposed to it came from institutional data?

Steven: There is no command you can send to grouper that would give you that answer. You could write a client that could issue a series of commands to grouper and you could figure that out yourself.

Michael: Ok.

Steven: We've also started some discussion about a container that would include faculty committees, for instance. Right now, as I said, we're in transition from grouper to LDAP. So, one of the criteria we're using is we might give someone the authority, the ability to edit groups if the return on investment – what it costs to claim the person and the value. In other words they are responsible for a lot of groups if the ratio is correct. So, it turns out there is one person who does all the faculty training. They already maintain web pages, that was the membership that it might make sense to train them to manage group memberships and then those groups could be used in the various applications that I listed. We haven't even gotten to the point where the production people are thinking about personal groups.

Keith: Steven, a question about that – you've repeated a couple of times, thinking about that at least – moving to LDAP. I'm wondering about the drivers, because infrastructure work is hardly ever fun, or regarded as critical there; so there must be some driver there.

Steven: Using locally written code as a key component of infrastructure, in my mind, is an exposure – an extreme exposure. And the guy who wrote this left five years ago. He only lives up the road in Summerville and he's real friendly; he responds to my email (laughter). But John Ballem, who is on the call, has recently had to immerse himself... and we've kind of cranked up our usage of this thing. John has had to immerse himself in how this thing works. And so John walks out of the building after this call and gets hit by a 757, sorry, so that's just in my mind a viable support factor.

Michael: This question is really for you, Tom, and Wes. Can you give us the characterization of the number of groups, the scalability issues? I know in your situation because you seem to be doing a lot of additional edits on groups and so on, this may be difficult. But, how many groups are institutionally maintained, or user maintained? Just to give people an idea of the magnitude of the problem.

Wes: We have an estimate of the number of the classes that are taught each term, nearly all of which have groups in the directly of various sorts. That's something on the order of 10,000. In addition to that, anecdotally, we know that every hall and every dormitory has a mailing list for students. Pretty much every student organization on campus has their mailing list. Every department pretty much has their mailing list as well. In total the current set of user-maintained groups in the university – all of our groups are use-maintained groups, so other than anecdotally, it's not possible to count which actually ought to be system groups -- there's, (as of) last weekend or so, 83,000.

Michael: Now, of those 83,000, how many of those are – and I really just made the comment, but – how many of those are from institutional data?

Wes: Zero. I mean automatically created?

Michael: Automatically created.

Wes: Zero are created automatically.

Michael: Ok, I just wanted to clarify on that.

Wes: We have policies in place for, for instance, when someone leaves the university, and they own groups that are used for university business, the process by which one changes the ownership of effectively a university-owned group. At this point there is no assignment, however, we are planning to have a department tree and a class list tree. The class list tree will actually be fed from university sources, in particular the registrar's office data, and at least initially populated and kept up to date as drop/adds happen with the professor or the instructor having some right of access to the group. The departmental list, which as I said, we don't really have any good institutional data on what represents a department. In fact, our departmental phone book approached us to get good data – rather than the other way around. (laughter) Apparently, the paper copy of the phone book continues to be published by having little slips of paper that are taped into place and sent into the photo type center. It's a big phone book. There is no intention of having that departmental list contained, for

instance – members of the department. The intention for that is to be a contact point for the outside universe. So, if I wanted to send mail to the Department of Chemistry, that would be the department there, and in all likelihood the email to Department of Chemistry would be sent to the secretary of that department, or something to that effect, or some other contact. We have not plans at this point to try to do institutional lists for email, or for access control, that are maintained institutionally. I'm not sure that that would be a good idea for this institution, in any case. And I don't think our institutional data is that accurate. Certainly, it's not that responsive even if it is that accurate.

Michael: Tom D.?

Steven: I guess if I could make one comment there, one of the things we're starting to see as the user community relies on the directories more and more, is back pressure – not by us, but by the community – to get the data cleaned up.

Tom B: We're having the same thing.

Steven: And that's going to be more effective than any screaming, or jumping up and down that we do.

Tom D: We've having that effect plus the opposite effect, which is those who are extremely sensitive to their privacy and masking their personal data as best as they can. So we discovered a tremendous amount of on-campus addresses and phone numbers, in fact, were merely addresses and phone numbers of the department. And so, there is a large number of people that...we actually don't know how to reach them. It's a department-by-department thing.

Michael: Can you comment on the numbers of groups and the magnitude of the problem? I realize that you don't have many LDAP groups, but you do have Andrew and other systems.

Tom D: Well, that's a tough one because I don't think the Andrew groups have ever been cleaned up in the 15-16 years that it's been in service. I mean...

I would certainly guess that there's thousands of groups; and, most of them are probably in active. And, I would say, 90% of them are incorrect.

Michael: And, do you have any institutionally maintained groups? Whether they be class verses...

Tom D: They're groups that we create from the institutional feeds and can use tend to those...and there really are not that many...but they're really not used that often. People tend to create mailing lists, and things like that, kind of as they need them. We don't tend to have a lot of campus-wide mailing lists. They tend to belong to individual clubs or small groups of people who maintain them themselves. The phone books – while we have a University phone book, it's really pretty inaccurate.

Michael: You said earlier that you had web-enabled services, so do you have web pages, that all of the faculty can get to and nobody else?

Tom D: We have web pages that requires, for example, Andrew authentication, and they've actually downloaded all of the employee-Andrew IDs into the web server. So, it's been done on an application-by-application basis, mostly. We have services that use Apache web server that we run, that use our configuration, that can use Andrew pts groups for access lists. But, not everyone uses those. People tend to take their own feed from the administrative systems and build their own authorization.

Michael: If I may, Wes – same question.

Wes: What's the question?

Michael: Do you have web-enabled services that, for example, that require authentication and therefore only faculty can to them.

Wes: Yeah, I mean, specifically for the directory? No.

Michael: So you do that, but not directory-enabled?

Wes: The directory has a web-enabled application, however, it's not specific to the faculty. Any faculty, staff, or student can authenticate to the directory.

Michael: I think the question is a little different. Do you offer web pages that you want to restrict that web page to just the faculty and you make use of the data in the directory to help control that access?

- Wes:** No, in fact we serve our web pages – our institutional web pages – out of AFS, so the service that we use for managing access to that is pts.
- Michael:** Ok.
- Wes:** However, we do have several institutional groups that are managed automatically that are even better than the pts system, for instance. We have a group called University:Members, which theoretically contains the complete membership of the university.
- Keith:** Are there, on this group of interviewees, portal efforts?
- Wes:** It depends on how they integrate.
- Michael:** Can I ask one favor? That we get Steven to answer the previous one, and then go on to that one? Because I think that's a great question.
- Steven:** Do we have pages or services that are controlled...or, where access can be controlled by groups in the directory?
- Michael:** There's that one. And then in grouper – how many groups are there and how many of those groups are institutionally maintained?
- Steven:** I'll answer the first one. I'll ask John to answer the second.
- Tom B:** Ok.
- Steven:** Yes, the web access-control system could. We have a thousand web masters on the central web service --- 800 or something like that, and any of them could create an ht access file and point it any group in grouper, and that would be used to enforce access control. There is a fair amount of user education involved and getting people to understand this. So perhaps the marketing side of our organization needs to do a little work there.
- Tom B:** I wonder if I could, before you answer the second question.
- Steven:** I was going to slip one more piece there. The web server is currently configured so that if something is IP address-restricted to the campus, and I access it from home, instead of getting rejected I get bounced into a proxy mechanism, and if I can authenticate with web SSO mechanism AND I'm a member of the group – campus community – then I can still

get in. That may be right now the best example of leveraging the directory for access control.

Tom B: Can grouper, I'm wondering, there's at least two different types of group-related questions that an application might want to ask. One is to give the list of whole membership list of a particular group and the other is to know if a given principle X in a given group, Y.... Does grouper support both of those types of paradigms?

John: Yes, Actually, I'm working on my server to find out the answer to the second question which is the total number of groups we currently have today, but can you repeat the question?

Michael: Tom, I might add that there would be a 3rd.

Tom B: Go ahead.

Michael: The 3rd would be tell me all the groups that someone is in.

Tom B: Yes. The question is what kind of group membership questions can grouper support of its applications or its clients?

And the three questions that we've thought of so far are ~ Give me the entire membership list of a given group. ~ Tell if a given principle acts as a member of a given groupY ~ And tell me all the group memberships of the principle acts.

John: Yes, to 1 and 2, which is, it's member functionality to do with like a web authentication-type. A group membership, obviously has to do with a bulk mailing or a major-domo type of thing. And the 3^d one is not in the API for the application. It's done via right now from one of our own locally-built perl modules--that ---there is a brut force to find all the persons in the groups to...

Steven: I thought there was a command to send it. There is certainly something on the web interface that will list all of the groups that I'm a member of.

John: I believe it's brut force.

Steven: Oh, ok.

Tom B: Ok, thanks. And then are we at the point of having the answers to all of these questions?

Michael: And then the number of groups.

John: Oh, the number of groups. We currently have about 8,000 groups. Of those, about 3,400 of those are base groups built on nightly feeds. The rest of those groups are a combination of ...each nightly feed has a group that is generated from it. Nightly feeds are not human editable. The groups that are generated from those – the drive groups – have delegation of administration, and it's probably currently right now about a half a dozen groups that have been made up by some administrator to include like departments and things like that.

Michael: So, is that to say that most of those 8,000 groups are from institutional data and only a handful are ad hoc maintained?

John: Yes.

Michael: Ok

John: And they are either directly from the feed itself, or derived as a result of that from the feed.

Keith: I'm asking Tom, the moderator, if portal questions fits into what time we have left – and the portal question is simply, If you have a major portal effort in some stage of thinking to design to implementation...

(turn to B side of tape)

directory -- the group enablement stuff with relation to that?

Steve: We have no portal left.

Keith: Lucky you.

Tom D: The HR group has gone off and done their own employee portal. And, since they have the employee data system, they run it on their own group thing, and they've used it outside contractor-implemented. There's a student portal being talked about amongst the administration, but it's still mostly being talked about now. I don't know quite how it would be done. My guess is that tool will be done with an outside group.

Michael: Wes?

Wes: This is Wes Craig from the University of Michigan. We have a student portal which is in...I believe it's referred to as Extended Pilot, which I

think it is advertised on campus for everyone to use. It has its own white pages client, and it uses the directory to determine...it has its own mail client as well. It uses the directory to determine where the user reads their mail, as well. I don't think it uses...I don't think it has any sort of access control, or anything like that. We're also in the process of ...as we change our web authentication mechanism, we're contemplating having the main page -- that you get to after you've authenticated -- use the directory to determine which from a set of pages you'll be directed to. Using the affiliations, you could call that directory enabled, depending on your definition of directory enabled. Certainly not whizbang directory enabled, but it would presume that the directory was authoritative for things like affiliation.

Keith: Right, then in addition to which pages do you use here in our case -- this is Keith again -- which tabs are on your portal page? There's the: Can the portal be a proxy for you towards backing things into use groups to help with that stuff? But, it sounds like we're pushing the limits, that the portal stuff just isn't that far along in its connections to the directory at this stage.

Tom B: Sorry, I dropped out of the conference somehow, and just got connected back in a moment ago. I gather we're talking about portal?

Keith: I've think we've kind of wrapped that, so it's swimming back to you.

Tom B: Ok, very good. I notice that we don't have many minutes left -- 7 or 8 minutes, perhaps, of the planned time. I wonder if we may have one other kind of a round on a closing of questions that is more future oriented -- so here's some. Let me ask you a question about which no one knows anything, including myself. So, there are no wrong answers here. I'm curious if you track roles, whatever that means? And does something like role-based access control -- is that a meaningful phrase to you? And, have you been thinking about such things, and if so what the heck is it? Go ahead Tom, if you would.

Tom D: Everything I know about role based-access control, Keith Hazelton taught me, late one night.

Keith: Oh God, there you go. Now you're really in trouble.

Tom D: We haven't really thought about it much at all. I understand the need for it. I just don't even begin to know how to do the political work along with the technical work.

Tom B: Let me back up a second. I'm not really going to ask from your experience, even your implementation planning. But, what's it mean? If it's going to back to the existential of the foundation, what do you think it should be – if ever you would go there?

Steve: Can I dive in for a sec? We build every night, John Softer rebuilds every night all of the base definitions for our course groups. And the feed contains the students; the feed contains the instructions. And, I mentioned earlier that I have this rule base that tells this program what to do. So, when he's building groups, he builds a group that contains the students, and then he builds a group that contains the instructor or instructors. And, the group that contains the instructors then is given certain permission visa via the course group, the one that contains the students. And so in some rough sense, this may align with your question of roles. You know, the policy base says that the people in these roles, for each course, get specific rights with respect to the course that they're teaching. And we're talking about a similar kind of model with departments. The departmental administrator would have certain rights with respect to the departmental group. Thirdly, with the first container I mentioned that contains vast slices of the community, my understanding of the model – I don't think they actually rolled it out yet – is that we have some type of hierarchical structure in the main space there. So, graduate schools set of groups, for instance, we would delegate to some set of people. We'd delegate to a role in the grad school, and that role might, in fact, be mapped in the end to two or three people, the authority to manage the grad school groups and do whatever they want with them. And, so the process of specifying

who's acting in that role becomes editing the membership of graduate school.admin. But, I think it aligns...the implementation, I think, aligns, as I understand it, with the RBAC model in that we're trying...

Keith: RBAC, by the way, being the Role Based Access Control.

Steven: We use the group membership to tell us who's authorized to act in a particular role. And then John's Policy Base, if you will, is used to specify what rights and privileges people in those net role have. And, we can change the rights and privileges by editing the policy base.

Tom B: But that's all contained to the group's world at this point, anyways. Is that correct?

Steven: Yes.

Michael: Do we hear from Wes?

Steven: But, the next step, of course is once you get to directory-enabled applications, managing the membership in these groups is in fact a form of provisioning. It depends on whether or not I can use the calendar service.

Tom B: Exactly, so you relate it back to provisioning. That's a key word I kind of heard from you. Ok, that's great. Wes?

Wes: The security branch of our tree, effectively allows sort of a primitive role based access control. The various branches of the tree have objects in them, typically for instance there's a consulting branch, and in the consulting branch are the various consulting groups on campus and the person buying the...according to which department they have access through. We do our access control based on the branch that you're in. So, for instance, all the consulting objects get access certain semi-private attributes. We have a postmaster object in there that campus postmasters can get access to. We have a manager tree that allows each department on campus to have unlimited manager access to the people branch or the group branch. We're not really using the directory for any access control, certainly not in any sort of large application to do access control outside of the LDAP database. For the most part, all the access control that the database is responsible for is within the data base itself.

- Tom B:** Tom D, did you have a chance to say you piece on that all together?
- Tom D:** Yeah, I think I did, didn't I?
- Tom B:** Ok, I wasn't quite sure. You did say every you learned was from Keith, and I thought I'd give him a second chance with that question.
- Michael:** I just want to go back and double check what I heard from Wes. That whatever role-based access you're doing or contemplating, is oriented towards the management of the directory and that applications aren't really making use of it? Is that true?
- Wes:** Pretty much. I mean there are isolated applications, for instance, our mail servers use the directory to extract the list of people that are allowed certain supervisory access to that system. There are a number of isolated cases where that's happening around the University, and that information is encoded in groups. As far as what you would describe as role-based access control, the only place...that's only in a very primitive form in our directory, and it's pretty much orientated towards access control of the directory itself.
- Tom B:** I was just going to say that we are just about at the end of our time. I thought I'd give everyone an opportunity. If there's something else they wanted to say that we just haven't touched on and they feel is terribly relevant, this would be a good time. Just jump in for a moment, we're about full. Wes, do you have anything else you'd like to add?
- Wes:** Yeah, there's one of the things that we recently changed, is the way entries get into our database; the people branches are effectively fed from our Kerberos System. I thought that might be of some interest.
- Tom B:** So you're coalescing the administration of a three people office into one system, down from two or more. Is that correct?
- Wes:** The way people used to get in is either someone with the manage level of access would add them. Or, when we got our institutional feed, they'd be added. Under the new system, pretty much you have to have...well, we've reorganized the database so that the DN is formed with the UID.

So you have to have been assigned a UID, and the UID space is owned effectively by the Kerberos System.

Michael: What is the frequency of update on the Kerberos System? And then what is the frequency of update on between Kerberos and the directory?

Wes: Those are all done interactively. The Kerberos System has a backside. Our Kerberos System, anyway, has a backside side that when an entry is created, or destroyed, or renamed, that information propagates on the order of seconds into the LDAP database.

Michael: Ok, just to clarify, is the Kerberos System institutional fed, or is that done effectively by human interface?

Wes: That's done by administrators.

Michael: Ok.

Wes: In fact that's done by...used to be done by administrators effectively in every department. We've put another level of access there, such that all of the people who used to be able to create Kerberos identities, continues to create Kerberos identities, but only a handful of authoritative accounts offices – these are offices which are able to check whether or not the person is actually affiliated with the University. Only people in those offices are able to set the flag that would cause people to appear in the directory.

Tom B: John Balem of Brown, any other last minute points?

John: A question for Wes, a quick question. Wes, you said that when you flattened out your model, you are now keeping person groups in the user object. How are you supporting referential integrity like that, or are you doing that right now?

Wes: We're not storing groups in the person objects – we're actually storing affiliation in the person objects. So, for instance, if you were to search the University of Michigan directory, you'd see that my affiliation is something about information technology, bla, bla, like that; that's just text. Those fields are stock fields that we get, and that's part of our institutional feed. The actual group membership in terms of, for instance, mailing lists

and whatnots are stored in the groups themselves as DNs to the people. Referential integrity is maintained in part by users editing their own groups and also by an administrative process that we run through to periodically prune dangling references.

Tom B: Steve Carmody?

Steve: Nope, nothing too important.

Tom B: Has Tom Dopirak had to leave our conference at this point?

Tom D: No, I learned that if you don't use the mute button correctly, you're talking to yourself.

Keith : And you thought we were ignoring you. We thought you were gone.

Tom D: I was sitting here thinking while you were discussing role-base authorization that one of the problems that really consumes a lot of time is not official University – departments and classes and things like that – but it's all the ad hoc organizations. They seem to consume an enormous amount of support time, and cause us to create an awful lot of special casing and software that we write.

Keith: Do you mean the Chess Club and things like that?

Tom D: Club and clubs that extend beyond the University boundaries, and ones that include part-time students and adjunct faculty, and people that fall in between the cracks. And I know that we are guilty of not spending enough time thinking about what the changing face of a member of the University is. And I think that is one of the things that we need to talk about as a group, and how does privacy enter and the people that are protected by FERPA, and Employee Relationships and things like that.

Tom B: Thanks. That's a good note, actually, to end on because also one of the things I worry most about these days is how to provide for new kinds of clientele, prospective employees, prospective students, and things like that; that's just inevitable, I think, and I expect them anyway.

Folks, thank you very much. I should ask, Michael or Keith, do you have any last-minute comments or questions? Any wrap up from you?

Michael: No.

- Keith:** No, I think the role-base stuff will get really interesting over time. I also agree with the questions that Tom raised at the end that are rich and we'll get into at the end of our 3rd and 4th hour of this call.
- Steven:** I think it would be useful, if there is another go round to talk about a set of people; the next ring out which would be applicants, parents, alums. But Tom also mentioned – once you get into student clubs, it gets really fuzzy. And you can use groups in the core ring. And in the second ring -- I just mentioned -- you can use the groups for bulk mail, you can use them for authorized access to web and other kinds of applications. When you get into that third ring -- someone who's a member of the Outing Club because they're friendly people – what are you giving them? And so it becomes much more than just a directory question; it's a policy question. What kind of provisioning are you really doing when you put this person in a group definition, or in the directory some how?
- Tom B:** Very good. Well, I like ending on a rhetorical note. This is Tom. So, Steven you're the perfect one. Folks thank you all very much.
- Keith:** Thank you Tom and Ellen.
- Ellen:** And everyone, did a very good job of saying who they were.
- Michael:** All right. Make sure everyone hangs up. Save the money. Bye