iPlant mtg notes

Meeting with iPlant - February 2011

iPlant Goals & Objectives

What is the community?

- resource providers (where resources are domain resources & tools)
    - ease of user management (SSO, ease of enrollment/registration)
    - ease of group management (inherit across providers; must align with user mgmt)
    - **data access across providers (iRODS)**
    - manage /quota/access at various levels (meter and throttle)
    - resource and services "federation" (sharability of workflow between services)
    - integration of 3rd party tools

- end-users / consumers (domain scientists)
    - SSO, linked to home institution
    - ability to form ad-hoc groups
    - ability to control access with keys/tokens that can be honored across services and providers, and yet limited based on time or something
    - can use web/API and command line apps w/ VO-based credentials (command-line apps where different libraries can be pulled in, can do more than just ssh)
    - keep data in one location, integration with many analysis platforms
    - activity dashboard, messaging and alerts, consolidated from providers

- *misc notes*
    - using google for calendar, document creation (wiki for final doc, but for the actual drafting, they go to google); use doodle for meeting scheduling
    - iPlant does not have grouper installed

- developers
    - user mgmt out of the box
    - access to federated resources
    - ability to meter and throttle (because developers easily create DOS situations; need handles in place so they know who is doing what)
    - unified sharing, reporting, dashboards
    - ease of integration to resources that need significant permissions (i.e. running compute data intensive tasks)
    - become part of a "market place" model

- *misc notes*
    - privacy of data is not a huge concern at this point from a contractual point of view, tho' individual researchers may be anxious about research not being exposed until the publication happens
    - internalization of VO may start to impact the importance of the privacy/restriction of data; possible assertion that if the data is under contractual restrictions, then maybe it is not a fit for iPlant

- educators (classroom, workshops)
    - make all iPlant resources easy to use in class settings
    - easy to work with ad-hoc user groups (workshops, tutorial with ease of provisioning/deprovisioning) especially for institutions with poor IdM
    - integration with LMS
    - integration with dashboard for management of self-paced tutorials/learning material

---

Outputs of iPlant

- toolkit for providers integrating with iPlant (iPlant does the heavy lifting with InCommon and others)
- toolkits for developers (domestication info; different toolkits for different needs and varying levels of sophistication)
- SSO features/capabilities for end consumers
- best practices for community (most resource providers have no idea about federation, domestication, VO IdM)
- a few domesticated apps
- infrastructure that promotes better collaboration
    - environments can be provided to researchers to give them a place to build their tools (they are designing the instructions and components for doing this; it will be platform independent)

Key applications needing domestication

- iRODS
- VNC
- others?
- iPlant current offerings

- *misc notes*
  - is there a flow diagram of all the ways researchers can access iPlant services? No, but this may be something the Bedrock grant-funded person may be assigned as an "intro" project; this basically will help understand scope
  - by end of meeting, that grant-funded position will be better scoped and hiring process can start
  - some time estimates for some of the work and priorities for iPlant would also be helpful, so we can determine where we need to wait to see how a space

---

Interactive presentation by Scott Koranda wrt my.ligo.org

---

iRODS

supports local database store for username/passwd, supports kerb, supports GSI; download as source code and compile and if you're using one of the listed forms of authentication, you are good to go from there; users can access via command line, webclient (webdav); API will be available through iPlant that will make iRODS accessible to the community; you can decide with whom you want to share a file, so with COmanage, knowing the membership of how you are sharing, what you are sharing, would be an important part of the service offering; ability to sync data is a really big thing (iDrop) because of how data will move from one repository to another transparent to the apps and the users; in terms of sharing and permissions, iRODS supports UNIX-like permissions; user and community data both is stored in iRODS, making it both a data store and a content delivery system

- check out DOI as a concept that links data, service, identity for publication
- not sure that COmanage itself can really help with the iRODS work; perhaps in making sure that information is expressed in a way the iRODS API can attach/consume?
- is iRODS a candidate for Moonshot?

Interactive presentation by Benn Oshrin wrt COmanage mockups

- Reporting
  - need some more requirements - are these reports purely IdM? Are they querying the apps? Are they usage reports for the platform?
  - iPlant has social scientists on board who want to have reports on who/what is working with whom, probably more
- History
  - keeping information is important, but keeping it private is also important; iPlant needs to consider what they'll need in workgroup situations
- Details/Questions
  - Is everything being stored in LDAP or some other way? Current implementation is built on top of a database (currently postgres, but that's adjustable); the registry piece should, in the long-term, be considered a black box - the VO shouldn't have to worry about this; the applications themselves will be looking at an LDAP or getting exposed to SPML
  - How much of this is exposed to API-style calls? Yes, most of the code is RESTful; everything that is documented on the wiki is implemented

---

Details

iPlant's enrollment process (today) http://www.iplantcollaborative.org

- enrollment is based on users wanting access to apps; they learn about the discovery environment through word of mouth;
- discovery environment and My.iPlant - they request access via website form, based on form that TeraGrid uses; notice then goes to a person who will run the scripts to create an account; backend is mostly a manual process at this point; they have a suite of python scripts that create accounts in LDAP and in the apps that need local accounts (i.e. JIRA); scripts also send out notifications to users letting them know account has been created and what their password is
- DNASubway has a different process for enrollment, they do not have an LDAP, different forms; but the account creation is automatic
- it would be desirable to be able to delegate the invite process to appropriate managers, and not core staff
- the information that is requested in part feeds the reports in to NSF, so we may need to define what items are actually required

iPlant's enrollment process (proposed to rollout in the next few months)

- will be done in Python/Jango
- this will be a centralized place to mange app services, api, and personal information; the registration process itself is not likely to change (much)
- apps will be both iPlant apps and Community apps
- allocations will be viewable (how much storage and where, how much resources (CPU, Memory, Block storage) in "Atmosphere", how much time in HPC; can also request larger allocations; they have the capabilities for some of this already
- still to do: systems access, including Atmosphere and virtual hosting, communications sections (users managing their own lists through this portal), group management (creating groups, verifying membership), HPC-related integration

Flow diagrams

- first, review of LIGO's enrollment and Expiration diagrams

- iPlant has not touched the whole question of expiration policy; it may be simpler because their access is not necessarily tied to the involvement from any particular institution
- some of the components in the system might be relevant for faculty to teach students who don't want to go to all groups themselves but want to be able to approve membership, registration, and possibly pre-populate student data

Roles

Role	Examples	Create accounts	Delete accounts	Run partner tools (TeraGrid)	Run local tools	Access Community Data	Add/Delete Community Data	Add /Delete Groups	Use Collaboration Tools (chat)	Add /Delete User data	Allocate Resources	Request Resources
Constrained user	psuedo-anonymous, temporary, guest, possibly student, conference /tutorial specific accounts	✗	✗	✗	✓	✓	✗	✗	✗	✓	✗	✗
iPlant user	identifiable user	✗	✗	✓	✓	✓	✗	✓	✓	✓	✗	✓
Steward	community data steward, local tool owner	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓
Administrator	Faculty, TA, Admin Asst.	✓	✓	✓	✓	✓	✗	✓	✓	✓	⚠	✓
Developer	tool developers /tool users (in Atmosphere), image creators	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓
Organization	creating a CO, allocating resources to the CO	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Legend	
✗	not allowed
✓	allowed
!)	allowed but with limited scope

- Constrained User - psuedo-anonymous, temporary, guest, possibly student, conference/tutorial specific accounts (there may be user registration specifically for that tutorial; this may be a special set up depending on the situation; these may be pre-created not individual registration; these may be very restricted in capability/access; if there are a large number of temporary accounts like this, will that trigger audit issues with granting agencies?); some sort of persistent, recurring access while not tracking personal information
    - they can run many (but not all, not TeraGrid, for example) of iPlant managed resources, they can view and subscribe all the community data in the system, they can add data temporarily, they can sign themselves up for mailing lists; they cannot create groups, they cannot invite other users, nothing that falls under collaboration (collaboration and communication requires more identification)
    - not an iPlant user, as they can be restricted by the faculty more than what a general iPlant user will be; if the individual behind the student account wants more access, they may request a full iPlant user account that will be a wholly separate account than their student account

- iPlant user - fully registered user
    - depends on the characteristics of the user; there are a set of services that have policies that may impact across the board and require higher levels of identity assurance; in addition to what the pseudo-anonymous user can do, they can link their allocations that they might be getting from other sources, they can request compute resources and most of the exceptions go away for them, they can add and manage their own user data (which has a different life-cycle than the p-a ones), they can participate in collaboration tech, they can create groups, they can create/edit their own analyses, they can invite other users (but they cannot approve those invitation), they can create requests (i.e. submit patches in an open source context)
    - they cannot edit community data or analyses; they can request but not assign allocations; they cannot create other users

- steward - a group manager; this is a role that would be granted, not something available on a form; ex. "clade owner"; anyone who is a data steward is also a general user, not sure exactly what privs this role would carry, or if it informs other roles (can they delegate to other specific individuals/roles? would that make that other person still a data steward or another role? we'll say they would all be data stewards for now)
    - responsible for community resources; they have everything that an iPlant user has plus read/write access to community data
    - not responsible for people data

- administrator/faculty users - want to use this for classroom projects, possibly independent research assigned out to students; they hear about DNASubway via workshop attendance, then go to registration form to request access/participation; they then go back to their classes and tell students to register as well; they will have situations of student teams which could be a shared single account; all accounts are equal
    - eventually may want to create these from course rosters but will have business processes to anonymize user names (concern about FERPA)
    - desirable: would like to have at least 4 roles in DNASubway, and those roles may apply more broadly, a "teacher" role (can see everything) and a "student" role (will be assigned to class groups), an "assistant" role (only see a subset of students, or have limited

approval for receiving delegated authority), generic/default user (not part of any class but can view publicly shared info); people can have more than one role
    - administrator can: create their own account, create accounts for students, group those accounts in to whatever organization the faculty member chooses; be able to view what all the students information; they can allocate resources (but there is a question regarding how constrained they are in what resources they can allocate to their students; this depends on what kind of resources (teragrid? atmosphere?) are to be allocated or re-allocated)
    - allowed to create and assign privs, possibly scoped down to specific projects or areas, create accounts

- developer role
    - tool developers/tool users (in Atmosphere) - functionality will be similar, but how they would use it would be different; a tool developer would create their own images and may have different allocations
    - bundle an image and make it public (but others can do that); they have access to development sandboxes and debug logs and possibly other metrics; ability to participate in some development processes; they will have a different set of permissions than what other users might have; it may be in the iPlant instance with their peer CO that they all should be just one CO in the system

- organization role
    - creating a CO, allocating resources to the CO

- for these roles, the steward, administrator, developer would all need to be iPlant users as well; they could request specific roles, but approval would be by the user committee(s); would it be possible for someone to automatically become a steward for the data they expose as community-accessible?
    - more will need to be explored regarding data access, exposure, creative commons as a useful construct?

Lifecycle of a partner CO

- from the COmanage perspective, iPlant needs the ability to administer COs that are separate but equal and possibly overlapping - why, really? Is this just a federation problem and an access management problem?
- in creating the relationship initially, iPlant must know about the CO, the users in the CO, to be able to map allocations appropriately (back to access management, possibly done via groups and attribute exchange)
    - how does the CO collect that identity data? there is no standard or comprehensive way this is done; they may or may not be in structured data at all; there are currently under 100 collaborations working with iPlant now
    - right now, COmanage is designed to have a super-user that handles the platform and can add/remove CO, and then each CO has its own admin or set of admins that are constrained to work within their own CO and not across other CO on the platform
- once the MOU's are signed, committees have blessed a collaboration, what happens next? an administrator will need to create the existence of that organization (users given accounts, groups of users based on their organizational affiliation; allocations will be assigned based on whatever was agreed to in the MoU); ideally there will be collaborators who could be administrators over components of the system (see role above), in fact there are no roles that they wouldn't be able to be assigned, tho' granted those roles may potentially be scoped/restricted

---

What would iPlant find most useful next?

- group management across all applications (grouper + domestication)
- more sophisticated relationships and ability to manage those relationships between CO (an inheritance relationship between CO); some format of how to exchange information on how iPlant CO was configured that they could import
- automate or build a work flow of the provisioning/de-provisioning process, tools that the would allow the committees that do the approvals could click a button (rather than send an email) to have approvals happen, and have that process be flexible based on application or CO or which committee needs to approve

---

Misc Notes

- iPlant is such a conglomeration of different organizations that some of the flows and business processes will be decided by the organization not by iPlant; there is likely to be separate enrollment flows, with separate roles; need to be able to have the flexibility for new roles to be created on the fly
- *Stanford AuthZ Model* - Business view includes roles, functions, tasks, and Internal/system view in turn includes entitlements and systems; we will need to decompose the iPlant roles to get to this
    - will want to talk about triggers for changing roles
- *provenance management* is something to think about, tho' it may be handled more via policy than via technology
- when the COs go international, when we have international collaboration, will need to have a broader conversation about what attributes international institutions will be willing/able to release (note that it is all about consent in Europe (plus "unconsent" in Switzerland), so a consent form may be sufficient for many things)

---

what is the vetting process and timeframe for what we've discussed today?

- need to describe the entirety of the problem needs to be part of the document, from there we can determine more priorities and requirements for what's next

---

Groups (not roles)

- Real-time plumbing in to the infrastructure, but not pushing enterprise groups in to iPlant (or vice versa)
- notifications and performing actions based on membership are the most attractive aspect of groups
    - approvals or change of approvals
    - you run an analysis (3-4 tools strung together) and this is of interest to the group you are in, so as the analysis completes, notification goes to the entire group
    - group profiles similar to individual profiles, so that groups could have a pool of interest and receive notifications as new data comes available that would be of interest to a group
    - Discovery Environment and My.Plant both have this need
- allocation of resources based on groups is something that is desired but not available to iPlant today
- need auditing across the board, including point in time auditing
    - metering, which can be seen as a more active form (real-time) of auditing; this probably has to be on the application side
- wouldn't it be nice to have a better connector between Grouper and the Grid space?
- may want to consider that you can "paint" attributes on groups; the whole question of group structure is one that iPlant will need to study
- can we say that everyone in a class/group is an iPlant user, or is there something else we want to capture; are roles defined by groups, or are groups defined by roles, or neither of the above?
- what is a clade? by shared access to data, a common profile of interest; there are discussion forums and pictures and home pages for each clade; there may be a specialized invite process, so this may be a little more than just a group - that is why iPlant has the idea of steward, to simplify this

---

Domestication

- note that in the future, iPlant is seeing an increase in the need to create a framework for developers to create and add in tools to the environment, and consider how other resource providers will connect/become available to the environment; the question of domestication metadata would be helpful
- in terms of existing tools "out there" that are desireable:
    - federated ssh, or better, a federated PAM module (using PAM for ssh, VNC) - Moonshot has great potential here
    - iRODS
    - databases (mySQL, PostGres hosting)
    - many of the tools that they expect will come through the iPlant API, and iPlant doesn't have any concrete API that does authorization or authentication yet
- the desire is there to domesticate all tools currently in use, including DNASubway (which drives so many people to iPlant)
- not providing wiki space as a service
- so, has a CMS, mailing lists, data repository, data analysis tools, XMPP, wants to offer help/service tickets (enabled for DE only right now), . No wiki, no web conferencing
- **we need to talk about what is the protocol(s) for domestication, and then what are the attributes for domestication**

Open conversation

- what are the outputs from this meeting?
    - roles document (Sonya, Heather)
    - data document = licensing structures on the data a la creative commons, fine grain access control issues; what are the policies around the data; need to scope what it is not going to look at (provenance, for instance; probably not metadata, ontologies, taxonomies) (Sonya, others @ iPlant)
    - roll up on how what we learned through iPlant can be described more generally for what VO might want to consider (Heather)
    - catch up with Nirav (Ken)
- metering and throttling
    - what resources would this apply to? what urgency? probably not heavily active in the next 3 months at least, but possibly in the 6 month time frame
    - trying to throttle access to the services and to the HPC resources
    - this is through the API
    - this is probably not a primary responsibility of the core COmanage piece; it may become sort of part of provisioning as quotas are pushed out to the apps; one could imagine a standalone service (like a licensing server) that can keep track of this; this could be a tool that COmanage would provide a simple interface to it, but the major work would be done in this standalone app
- people can update info in COmanage, can apps? yes
- where are are the individual's pictures stored? they could be stored in the person registry, it could be just another attribute about a person
- what is the iPlant schedule like?
    - fairly app specific, tho' in general we can expect that the iPlant team will be pushed to deploy/upgrade new tools over the summer
    - stuff is also rolled out in time for SuperComputing and for NSF site visits
    - in general, this year is pretty busy and booked with projects
- provisioning/de-provisioning
    - some apps may have a separate process to the organizational process
- what about a workflow/business process tool?
    - probably not necessary for the most basic question of "are you qualified to be a person?" but when we get in to access management, that may be a useful tool

---

what would be the early wins between COmanage and iPlant

- quick access to grouper, since iPlant doesn't have it - if it were an early part of COmanage, that would be great
    - would be nice to have an evolution of grouper doc, a la evolution of shib doc
- better management of mailing lists and integration with groups
    - currently using mailman, not opposed to using sympa; tho' perhaps this is not a quick, non-disruptive win
- automation of any of the notifications that happen in the people business process flows

---

Next round of conversations

- migrations and transitions (next 3-5 months)
- provisioning/de-provisioning details (at both org and app level)
- there will be ad hoc conversations, but a once-a-month roll up to inform the higher level individuals about progress (HF to schedule)