User Needs Assessment: XSEDE-wide Group Features

April 6, 2017

XSEDE's Community Infrastructure team (XCI) conducted an enterprise-wide survey of how *groups* are currently handled within XSEDE tools and services. We attempted to answer the question, "Where are XSEDE personnel and users spending time defining, storing, accessing, and managing *lists of people?"* We looked at the following areas in particular: (1) XSEDE staff tools, (2) XSEDE resource/service allocations, (3) campuses and campus bridging, and (4) web applications and science gateways. This document summarizes our findings. XCI will use this information to identify and set priorities for new system features that will increase the value of XSEDE to the scientific research community.

Contents

Who we spoke to

XCI has limited personnel available for user needs analysis, so we weren't able to speak with every area within the XSEDE enterprise that might have needs for groups. We focused our effort on three kinds of group applications:

- Groups of XSEDE staff members (for administrative/operational purposes)
- Groups of XSEDE users (for XSEDE L1/L2 resource allocations)
- Groups of XSEDE users (defined by and for users themselves)

Based on these three priorities, we selected XSEDE areas to interview about their need for XSEDE-wide groups. Table 1 shows the areas that we did (and did not) speak with.

Areas we spoke with:	**Areas we *did not* speak with:**
Allocations and accounting Campus bridging Extended collaborative support service Operations Project management Science gateways User portal Cybersecurity User help desk	Campus champions Campus CIOs Community engagement and enrichment External CI service providers L1/L2 service providers Research PIs/project leaders Training, education, and outreach

Table 1. XSEDE areas consulted in this needs analysis

Group definitions

From listening to people in these areas, we learned that there are currently more than twenty tools and services used by XSEDE staff and community members for storing, managing, and using group definitions. There are two high-level types of group definitions: (1) groups of XSEDE staff members, and (2) groups of XSEDE community users. Table 2 summarizes the general ways in which each type of group is used.

How groups of staff members are used:	How groups of community members are used:
• Control who can administer resources and services • Control access to staff tools and data • Disseminate information (via email) • Assign tasks (tickets, project activities)	• Control access to XSEDE, L1, and L2 resources and services • Allow users to define and use their own groups for access control

Table 2. Uses of groups within XSEDE

Groups of staff members

Most groups of XSEDE staff members are created and managed in one of the following three ways. The list is ordered from most frequent to least frequent.

- The group is managed by the same XSEDE team from which members are drawn. (E.g., the ECSS team manages a group containing ECSS team members.)
- The group is managed by XSEDE project management staff.
- The group is managed by XSEDE operations staff.

Staff groups are used to define which staff members have administrative privileges with XSEDE services and to define who has access to XSEDE staff tools.

Groups of user community members

Groups of XSEDE user community members (individual researchers and research teams, science gateway operators, campus service providers) fall into two broad categories: groups that are defined by XSEDE staff and groups that are defined by the users themselves. The former category of groups is used to control access to resources and services provided by XSEDE itself or Level 1 or Level 2 service providers. The latter category of groups is used for whatever purposes the users envision, but it typically involves either: (1) further refining access control within that already established by the XSEDE staff, or (2) controlling access to resources and services provided by XSEDE Level 3, campus, or other service providers.

Applications in which groups are defined

XSEDE staff and community members define groups in more than twenty applications. Table 3 lists the applications in three categories, based on who defines (or manages) groups. Note that two applications appear in all three categories: the XSEDE central database and Google. Groups in most of the other applications are defined and managed by only one of the three categories.

Tools in which groups are defined by XSEDE project management	Tools in which groups are defined by other XSEDE staff	Tools in which groups are defined by XSEDE community members
Confluence/JIRA Google Qualtrics (survey system) RT (ticket system) Sharepoint Skype for business XSEDE central database XSEDE financial portal XSEDE metrics dashboard Zoom (teleconferencing)	Email lists Google L1/L2 resources Moodle (training system) RT (ticket system) Single sign-on hub Subversion (SVN) XSEDE central database XSEDE user portal	COmanage (Internet2) Globus Google Grouper (Internet2) HUBzero L3/campus resources Science gateways VOMS (Open Science Grid) XSEDE central database

Table 3. Applications in which XSEDE groups are defined

Access control using groups

Many of the groups defined in XSEDE systems and services are used to define which individuals should have administrative privileges and/or basic user access to resources, applications, or data. Of the more than twenty tools and services that have group management features, sixteen are known to be used for access or administrative control. Table 4 details these uses. (This is not an exhaustive inventory, because we did not talk with all XSEDE areas, nor did we ask for every known service.)

Tool or system	Purpose of group	Who manages group
Confluence/JIRA	Control access to activity tracking & documentation	Project management staff
Google	Control access to XSEDE project documents/reports	Project management and XCI staff
L1/L2 resources	Administer resources	L1/L2 staff
	Control access to allocated resources (local)	L1/L2 staff
L3/campus resources	Administer resources	L3/campus staff
	Control access to L3/campus resources (local)	L3/campus staff
Moodle	Administer use of training sites	Training, education, outreach
	Control access to training sites	Training, education, outreach
Qualtrics	Administer use of survey tools	Project management staff
	Control access/invitations to specific surveys	Evaluation team members
RT	Administer ticket system	Operations staff
Sharepoint	Control access to XSEDE project documents/reports	Project management staff
Skype for business	Control access to XSEDE's telecon services	Project management staff UIUC
SSO Hub	Administer SSO Hub	Operations staff
SVN	Administer SVN services	XCI staff
	Control (write) access to SVN areas	XCI staff
XCDB	Administer and use accounting data	Accounting staff
	Administer the allocations system	Allocations managers
	Control access to allocated resources	Allocations managers
	Control access to allocated resources (discretionary)	L1/L2 staff
	Control access to allocated resources (within projects)	Researchers/scientists
	Control access to research proposals	Researchers/scientists
	Control access to XDMOD	XDMOD staff
	Control access to XUP apps	XUP staff
XSEDE financial portal	Control access to XSEDE financial data	Project management staff
XSEDE metrics dashboard	Control access to XSEDE metrics data	Project management staff
XUP	Administer the XUP	XUP staff
Zoom	Control access to XSEDE's telecon services	ECSS staff

Table 4. Groups known to be used for administrative privileges and access control

User-defined access control

In addition to the XSEDE and SP services shown in Table 4, it is also common for community members to use their own mechanisms and methods for controlling access to things. The “things” in question may be

data or code (especially common among researchers and their teams) or web applications, software licenses, and computing resources (especially common among science gateway providers). The mechanisms used for controlling access are usually closely tied to the tools being used. Rather than selecting a group management system and adapting their tools to use it, these users typically take whatever system is already used by their tools and create their groups and access control rules in it. This means that group definitions typically cannot be shared by research tools and must be re-created in each tool.

The following list offers some examples of community-defined groups and their uses.

1. A science gateway developer uses a Postgres database on the gateway server to store user identities (including login credentials and user profiles).
2. The science gateway developer uses the same database (see 1 above) to record which users have access to a personal license for a specific software tool, and only those users are allowed to access that tool within the gateway.
3. A web application developer uses an Open Science Grid VOMS service operated by a physics research collaboration to determine which gateway users should and should not have access to the team's data in the application.
4. An economist at Harvard uses a COmanage group to authorize research partners at other InCommon institutions to a collaboration server provided by his department's computing group.
5. An ecologist working at SUNY Buffalo uses a Globus group to authorize a set of high school biology teachers to access her field observation data on a workstation in her campus lab.
6. A research computing administrator at the University of Michigan uses the XSEDE SSO hub (single sign-on hub) to allow selected XSEDE users at other colleges to login to a departmental cluster.
7. A field biologist working in Yellowstone National Park uses a Google group to allow her colleagues at other sites to access her observation notes in Google Drive.

Methods of group management

The ways in which groups are managed (created and maintained) vary throughout the XSEDE community. The differences are mainly in *who* does *what* to the group, and the *interfaces* used to do those things. Though there is consistency within specific XSEDE areas (e.g., XSEDE project staff, science gateway developers), there are variations even within the same area.

Manual group management

Everyone that we spoke with acknowledged using manual, or human-managed, group functions. What this means is that a human being interacts with a user interface (almost always via a Web browser) and uses the interface to create groups, edit group properties (name, visibility, description, administrative roles), and add individual users to (or remove them from) groups. The following seem to be the most common requirements for manual group management across XSEDE areas.

- Someone creates a group and adds specific members to it via lookup or removes specific members from a member list.
- Someone creates a group and invites specific members to it via lookup and/or email address. (Invitees don't become members until they accept invitation.)
- Someone creates a group and lets candidate members know about it so they can request membership if they want to. (Requesters don't become members until their requests are submitted and subsequently approved by a group manager.)
- Someone discovers a group via search and requests membership. (Requesters don't become members until their requests are approved by a group manager.)
- A group owner authorizes someone else to manage the group's membership.

The ability to create and manage groups is needed by everyone involved in the XSEDE community. By itself, a group does nothing and affects nothing else. But the ability to create groups, manage them, and have the opportunity to reference a group in an access control list elsewhere in the system seems to be universal. In particular, the ability to create and manage groups should not be restricted to people with active XSEDE allocations. XSEDE groups--like XSEDE identities--should be available to anyone who can interact with the system in any way, and should be reusable indefinitely.

Systematic group management

Systematic--or automated--group management seems to be important for **XSEDE staff tools** and for **XSEDE allocations**. "Systematic management" means that the creation and management of a group is tied to other processes or changes in the system, rather than being performed manually by humans. Changes to specific groups are driven by business processes, such as billing & invoicing actions, project WBS (work breakdown structure) changes, allocation status changes, identity verification ("vetting") actions, etc. The following list provides several examples in which systematic group management is or would be helpful.

1. It would be helpful to have a group that contains all members of an XSEDE project WBS area. The group should be updated automatically whenever a team member is added/removed from the WBS area.
2. It would be helpful to have a group that contains individuals whose identities have been validated via a systematic XSEDE process. (Note: It isn't clear that XSEDE currently has an identity validation

process that would satisfy this need.) Membership in the group should be managed automatically as part of the identity validation process.

3. It would be helpful to have a group that contains all individuals who are currently members of an active XSEDE allocation project team. The group should be updated automatically whenever new projects become active or expire and whenever a project leader adds/removes a member from his/her team.

For systematic group management to be useful, the following requirements must be satisfied.

- Ownership / administrator roles must be clearly defined. How a group is managed (who has authority to make changes and which processes trigger changes) is as important as the membership of the group.
- Time-based changes need to be possible (e.g., expiration dates, timeouts). It would be useful to be able to specify that a group auto-expires or becomes inactive (unusable in access controls) on a certain date or after a certain period of time.

External group management

Group management that takes place in external systems is particularly important for **science gateways** and **campuses**. "External management" means that a group is created and managed in a another system (e.g., Google, OSG/VOMS, Grouper, Globus, HUBzero) but is usable within XSEDE (by XSEDE users and/or resource providers) for access control.

The following are some specific examples where the ability to recognize and reference groups defined and managed in other systems is important in the XSEDE context.

1. A research team based at a particular campus manages a Google group for their own local use, and that group is then used within a science gateway to control access to data belonging to the research team.
2. A multi-institution research team uses the Open Science Grid's VOMS service to define its group members, and an XSEDE L3 service provider at Indiana University uses that group to authorize jobs submitted by the team.

One requirement that seems important to the people we heard from is that in cases where a group is separately defined in more than one system (multiple copies are stored), the copies must be kept synchronized. If the synchronization of a group across systems can't be trusted, the users may not be willing to trust the group with their access control.

Unsatisfied user needs

The following sections record specific things people in the XSEDE staff and community do (or want to do) for which XSEDE's current system doesn't appear to provide a full solution. Note that these issues are not all strictly related to groups or group management. We intentionally recorded and are reporting on all of the unmet needs that we heard, rather than limiting the scope to what appeared to us to be group-related issues.

XSEDE staff: WBS-driven groups

XSEDE's operations staff and project management staff currently manage the access control rules that authorize staff access to a large number of tools and services. There is no central database of the "official" staff roster that can be used as a source for this information, however. The project staff changes frequently and staffing changes are not communicated systematically throughout the project, so it is quite difficult to keep the access control settings synchronized with staffing changes. Staff members are often frustrated to find that they or their team members have been left off of important email lists or the access control lists for staff tools, or that former staff members are left on access control lists long after they leave a team.

The best attempt to resolve this issue that we've heard of so far has been by the project management staff who maintain groups in XSEDE's Confluence/JIRA service from Atlassian. These groups are only partially populated, and they are all administered uniformly by the JIRA/Confluence administrators (as opposed to being maintained by the leaders of each area, which might be more manageable).

A similar group structure is maintained in the RT (trouble ticket) system, but it is driven largely by user support and issue resolution tasks, so it also isn't a comprehensive reflection of the entire project staff. The RT groups and the Confluence/JIRA groups are not automatically synchronized or linked.

XSEDE has a formal staff "onboarding" process that can be leveraged to help maintain a central, WBS-driven staff database, but it hasn't yet been used to populate a central staff database that serves as a source for other tools and services.

Table 3, specifically the first and second column, contains a list of staff tools that ought to be managed centrally. The tools below are those arguably most in need of WBS-driven groups.

- RT (ticket system)
- Confluence
- JIRA
- SSO Hub
- Email lists
- The (currently non-existent) staff database

As a specific example, the manager of the ECSS area noted that there are currently ~80 individuals assigned to areas within the ECSS WBS. Most have partial (as opposed to full-time) assignments. Staffing changes are frequent, and keeping track of them is time-consuming. The tools currently used to track ECSS staff and staffing assignments may be sufficient to record changes, but they don't offer connectivity with the other tools used by ECSS, so they aren't being used for access control in the tools on an ongoing

basis. It would be helpful to have a centralized mechanism for recording ECSS staff assignments that could be used throughout all of the tools used by the ECSS area.

XSEDE staff: Vital XSEDE community groups

In addition to XSEDE personnel, XSEDE's project management personnel are also responsible for a nontrivial number of groups of people closely related to, but not employed by, the XSEDE project. Membership of these groups needs to be tracked on an ongoing basis. These groups should ideally have the ability to use some of XSEDE's tools with access control restrictions (e.g., Confluence, maybe JIRA, RT, and/or XUP apps).

At the moment, each group has an XSEDE staff person responsible for tracking the group and managing the group's access to XSEDE services and tools, but the methods used are *ad hoc*. The following is a non-exhaustive list of these kinds of groups.

- Campus champions
- Domain champions
- Regional champions
- XSEDE Advisory Board (XAB) members
- User Advisory Committee (UAC) members
- Service Provider (SP) Forum representatives
- Key NSF personnel (e.g., program officers)

Science gateways, L3 and campus service providers: Identity validation services

XSEDE currently performs a limited degree of identity validation for community members who apply for and receive research allocations on XSEDE-allocated resources. This service is not currently extended beyond the users of XSEDE-allocated resources. Identity validation is important when authorizing access to services that have significant costs (e.g., HPC systems) or that have legal restrictions (e.g., sensitive data and technologies, licensed software). Both campus service providers and science gateway operators would like to be able to offer services to validated users, but neither currently has a good mechanism for validating the identities of individuals beyond their immediate (formally employed or enrolled) communities. For campuses, validation is limited to members of their own campus communities (excluding people from other campuses), and for science gateways validation is limited to their own development and operation teams (excluding most gateway users).

Campus service providers, Level 3 service providers, and science gateway operators would use validated identities to provide basic access to verified members of other research institutions, to offer higher qualities of service (e.g., access to expensive services or services that employ sensitive technologies), to offer use of services that use licensed software if it could be known that the user has such a license, and to provide access to datasets that may include sensitive data.

One possible way to give gateway operators, L3 and campus service providers access to vetted user identities would be to provide an "on-request" vetting service for anyone who is registered with XSEDE. By submitting a request via the XUP, any registered user could initiate staff vetting of their own identity, after which their account would be marked as "vetted" in the XSEDE central database. Science gateways could then require that users login to the gateway with a vetted XSEDE identity--or otherwise demonstrate that

they have a vetted XSEDE identity--in order to use a specific feature or access a specific resource within the gateway.

Another way to satisfy this need would be to provide a "self-service" validation feature. For example, a registered XSEDE user could access a feature in XUP or a related website that allows him/her to demonstrate that he/she can authenticate via an InCommon Bronze or Silver IDP, and, on success, automatically marks them as having a certain degree of vetting consistent with the InCommon criteria. (Note: There are currently no InCommon Silver IDPs.) CILogon has code to do this, and Globus Auth offers the possibility that XSEDE users can link their InCommon identities to their XSEDE identity, so this might require only a modest addition to the existing system. This solution would have the advantage of having no staff cost, since it would rely on existing certifications and user vetting that is already being performed by identity providers.

Science gateways: "Login with XSEDE"

The purpose of XSEDE science gateways is to provide useful scientific/research services--*making use of XSEDE's resources*--to a broader base of researchers than those who apply for and receive individual allocations. XSEDE requires science gateways to collect, manage, and report data on the individuals who use the gateways, but it currently does not provide any tools for authenticating individual users. Each science gateway currently creates its own mechanism for authenticating users, keeping track of user profiles, and establishing access controls based on user identities. Two undesirable effects of this are: (1) the effort and technical expertise required to develop and operate a science gateway is higher than it needs to be, and (2) there is more variation in the quality (and reliability) of authentication mechanisms than their ought to be, given the importance of the data to XSEDE and XSEDE's service providers.

Encouraging science gateway developers to allow users to authenticate using XSEDE's Globus Auth service would provide a number of benefits.

1. It would free gateway developers from having to develop their own user registration and authentication interfaces, while allowing them to continue having full control of authorization (access decisions).
2. It would help the XSEDE community de-duplicate identities in usage data (the same person using a gateway vs. using other XSEDE services).
3. It would free gateway users from having to create and remember userids/passwords for each gateway. (In addition to the XSEDE userid and password, Globus Auth also allows users to authenticate using their credentials from InCommon-participating campuses and organizations.)
4. It would reduce the number of unique authentication mechanisms used in science gateways and provide a more consistent level of quality.
5. It would enable the gateway to use XSEDE's (and Globus's) other features (XUP, training services, groups, data transfer and sharing) without requiring users to re-authenticate to XSEDE or Globus.

Science gateways: User-defined groups

Science gateway developers expressed an interest in an XSEDE-recommended, XSEDE-provided group management mechanism that would allow gateways to offer group creation, group management, and group-based access control features to their users. Features that would be especially desirable to gateway developers follow.

1. Group data (definitions, profiles, membership data) would be hosted by XSEDE services rather than by the gateway. The gateway would implement its user-facing group features using an XSEDE-provided API (presumably a RESTful web service).
2. Gateway users could create and manage groups on their own initiative without gateway operator or XSEDE staff assistance. The gateway could either provide an integrated/custom user interface for accessing the group functions or direct users to a generic, XSEDE-provided group management user interface.
3. Groups created in a gateway would be accessible in other gateways and other XSEDE services, and vice versa.
4. Groups created in a gateway would outlive the gateway. In other words, the gateway could cease operation but the groups could still be reused in other gateways or services. (Note: This implies that XSEDE's group features should not be tied to allocations.)
5. Mechanisms to access groups defined in other group management services (e.g., COmanage, Grouper, Globus, Google) would be especially helpful.

Example 1: A gateway creates a group for access to a specific project area within the gateway, and the same group can also be used in Globus to allow file upload/downloads to the project area in the gateway.

Example 2: A research group that uses one gateway creates a group for access to their "team area." They subsequently begin using another gateway. Rather than recreate the research group in the new gateway, they'd prefer to import or share the original group.