PROCEEDINGS OF SCIENCE

A Model for Identity Management in Future Scientific Collaboratories

Robert Cowles[1]
Indiana University
USA
E-mail: bob.cowles@gmail.com

Craig Jackson
Indiana University
USA
E-mail: scjackso@indiana.edu

Von Welch
Indiana University
USA
E-mail: vwelch@iu.edu

Shreyas Cholia
Lawrence Berkeley National Laboratory
USA
E-mail: scholia@lbl.gov

Over the past two decades, the virtual organization (VO) has allowed for increasingly large and complex scientific projects spanning multiple organizations and countries. The eXtreme Scale Identity Management (XSIM) project has surveyed a number of these VOs and the resource providers (RPs) that serve them, and built a model expressing the identity management (IdM) implementations supporting these large scientific VOs. The initial model was presented at eScience 2013. This work refines that initial VO-IdM model with XSIM efforts since the original eScience 2013 paper, capturing results from additional interviews and initial applications of the model, and begins to extend the model to include federated IdM environments, portal-based VOs and cloud and exascale RPs.

International Symposium on Grids and Clouds (ISGC) 2014
Academia Sinica, Taipei, Taiwan
23-28 March, 2014

[1] Speaker

1. Introduction: Background and Our Prior Work

Identity management (IdM) is the practice of creating and maintaining digital identities (composed of an identifier and attributes) regarding entities and conveying those identities in a trustworthy manner, such that other relying entities have some assurance about with whom (or what) they are communicating. These processes allow relying entities to make informed, confident decisions regarding, for example, how to service requests, log activities, and respond to security incidents.

In the early days of scientific computing, resource providers (RPs) had an unmediated relationship with their user communities, and therefore handled all aspects of identity management, what we refer to as the *classic model* of IdM. As scientific collaborations increased in both number of people and magnitude of computing requirements, they needed to obtain resources from multiple RPs. The concept of a virtual organization (VO) [1] emerged to coordinate the scientific collaboration and its relationship to the multiple RPs serving it.

The distributed and heterogeneous nature of the computing resources and unique position of the VO in negotiating and managing community relationships resulted in new opportunities and challenges for IdM. After roughly two decades of experience implementing VOs, a number of IdM approaches have been used. An initial objective of the *eXtreme Scale Identity Management for Scientific Collaborations* (XSIM) project is to develop a descriptive IdM model which includes the VO, encompasses the current solutions, and provides insight into the factors favoring one solution over another. Iterating on this model, XSIM is working towards its goal of providing practical advice to VOs and RPs on designing and optimizing IdM implementations fit for their particular needs.

Foundational to our model is the idea that VOs may be delegated IdM responsibilities classically carried out by the RP. VOs play an important role in brokering relationships between scientific communities and RPs, and in the context of IdM can (and often do) play some role in setting up mediated trust relationships between RPs and users. (For more on mediated trust, please see [2] at Appendix G). Our initial publication [3], presented at the 9th IEEE International Conference on eScience in 2013, established a simple VO identity model, described in the Section 3, which expressed the VO-RP relationship in terms of the amount of delegation of responsibility for IdM from the RP to the VO.

In subsequent work, presented at the 20th International Conference on Computing in High Energy and Nuclear Physics (CHEP2013) [4], we began exploring the motivations that VOs and RPs have for these delegations. It identified the following factors (on which we elaborate later in this paper): the need to provide isolation among users; persistence of user data at the RP; complexity of VO roles; cultural and historical inertia; scaling in terms of the size of the VO and number of RPs; and the RP's incentive to support the VO.

In this paper, we begin with a discussion of related work in Section 2, then describe our additional interviews in Section 3, refinements to our VO IdM in Section 4, and influential factors in Section 5. We conclude with a NERSC use case illustrating and applying our refined model in Section 6, followed by Conclusions, Future Work and Acknowledgements.

2. Related Work

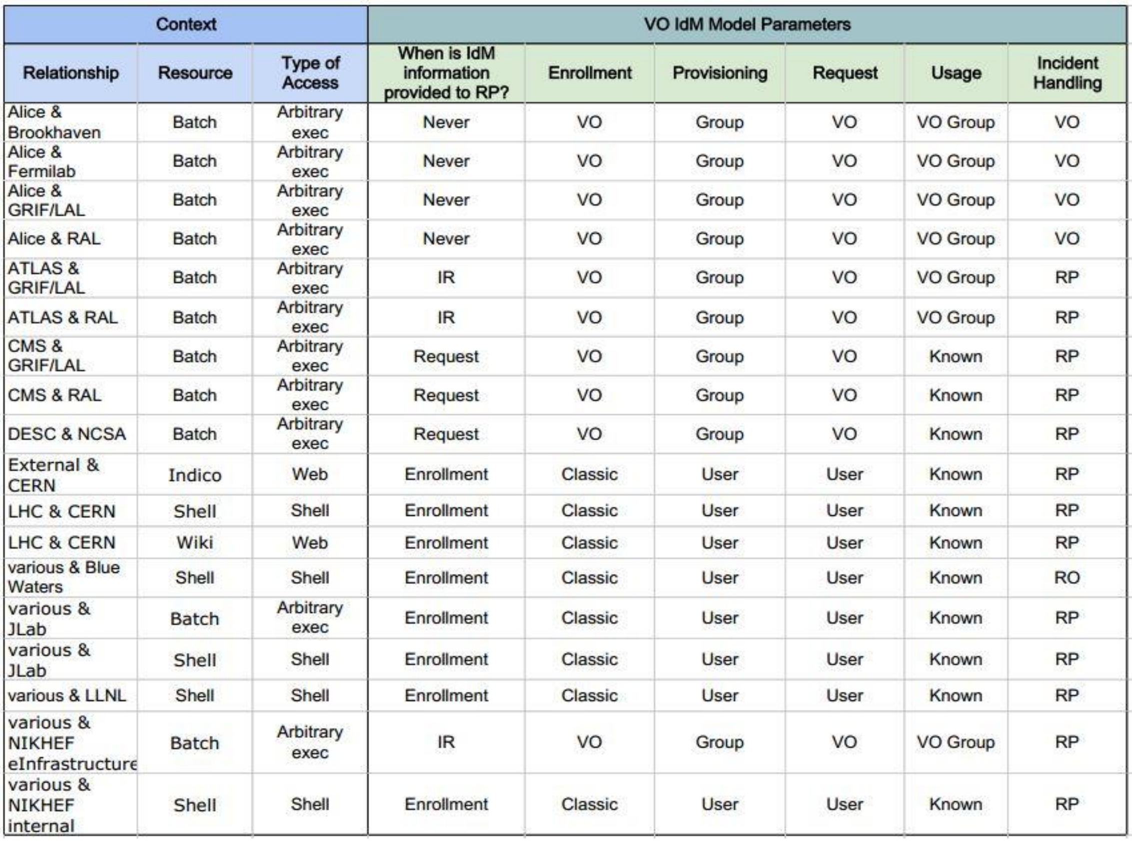
Our prior work was shaped by Landau and Moore [5], Broeder, et. al. [6], and Altunay [7], whose work shaped our theoretical underpinnings and assumptions. Work by Lin, Vullings, and Dalziel [8] explored factors related to trust in making access control decisions in the context of VOs, but was focused on and directly applicable to access control decisions rather than the whole identity management system.

Work in this paper was informed by others who had considered different ways of decomposing identity management systems and the flow of identity management information: work by Internet2 (e.g., see slides 21 and 22 of [9]), JISC's Identity Management Toolkit [10], the Identity Ecosystem Steering Group (IDESG) Functional Model Working Group [11], and Microsoft's Vision for an Identity Metasystem [12].

3. Results from Additional Interviews

In this section we capture the results of interviews we have conducted since our paper for eScience [3]. Table 1 shows these interviews captured as they would have been with our model at the time of the presentation to the eScience 2013 Symposium. At that time, a high-level, key parameter of our VO IdM Model was at which of the lifecycle stages, if ever, the RP becomes aware of the identity of the user. This stage is identified in one column of the table. For the lifecycle stages, we also identified the finer-grained options for how the identity information was managed between the VO and RP.

- **Enrollment**: An initial, typically one-time process by which the user is admitted into the VO. The table values are: VO makes decision to enroll unilaterally (VO); or RP makes decision (Classic).
- **Provisioning**: Following an enrollment, the one-time creation of any state associated with the user across the VO or RPs. Table values are: Shared group account or dynamically assigned Pool account (Group); or user account created (User).
- **Request**: The process by which the VO makes a request for resources from an RP to provide service to its users. A request can be in direct response to a user's action or can be an a priori reservation (e.g., a pilot job). Table entries are: RP authorizes based on per-user information (User); or RP authorizes based on VO membership or role (VO).
- **Usage**: Consumption of a RP's resource by a VO to provide service to a user. This can directly follow a request or may come sometime later. Table entries are: Known user of an isolated account (dynamic or persistent) or a shared group account. (Known); or VO member use of an isolated account (dynamic or persistent) or a shared Group account (VO Group).
- **Incident management/response (IR)**: An event that typically requires manual interaction with the user to resolve. This includes computer security incident response, a misbehaving user process, or a user support process. Table values are: RP handles incidents (RP); or VO handles incidents (VO).



Context			VO IdM Model Parameters					
Relationship	Resource	Type of Access	When is IdM information provided to RP?	Enrollment	Provisioning	Request	Usage	Incident Handling
Alice & Brookhaven	Batch	Arbitrary exec	Never	VO	Group	VO	VO Group	VO
Alice & Fermilab	Batch	Arbitrary exec	Never	VO	Group	VO	VO Group	VO
Alice & GRIF/LAL	Batch	Arbitrary exec	Never	VO	Group	VO	VO Group	VO
Alice & RAL	Batch	Arbitrary exec	Never	VO	Group	VO	VO Group	VO
ATLAS & GRIF/LAL	Batch	Arbitrary exec	IR	VO	Group	VO	VO Group	RP
ATLAS & RAL	Batch	Arbitrary exec	IR	VO	Group	VO	VO Group	RP
CMS & GRIF/LAL	Batch	Arbitrary exec	Request	VO	Group	VO	Known	RP
CMS & RAL	Batch	Arbitrary exec	Request	VO	Group	VO	Known	RP
DESC & NCSA	Batch	Arbitrary exec	Request	VO	Group	VO	Known	RP
External & CERN	Indico	Web	Enrollment	Classic	User	User	Known	RP
LHC & CERN	Shell	Shell	Enrollment	Classic	User	User	Known	RP
LHC & CERN	Wiki	Web	Enrollment	Classic	User	User	Known	RP
various & Blue Waters	Shell	Shell	Enrollment	Classic	User	User	Known	RO
various & JLab	Batch	Arbitrary exec	Enrollment	Classic	User	User	Known	RP
various & JLab	Shell	Shell	Enrollment	Classic	User	User	Known	RP
various & LLNL	Shell	Shell	Enrollment	Classic	User	User	Known	RP
various & NIKHEF eInfrastructure	Batch	Arbitrary exec	IR	VO	Group	VO	VO Group	RP
various & NIKHEF internal	Shell	Shell	Enrollment	Classic	User	User	Known	RP

Table 1*: Results of our interviews since our eScience 2013 paper captured using the model we used in that paper.*

In analyzing these interviews, we noted that relationships involving supercomputer facilities (e.g., Blue Waters, LLNL) tended to retain the classic model for IdM since those facilities normally granted shell access and often required multi-factor authentication. The interview involving CERN, which provides services to the LHC experiments in the classic model, also expressed a desire to reduce the IdM costs via a move to a federated IdM environment.

4. Refined VO-IdM Model

We now present our refined VO IdM model, a significant advancement from our prior work. The model remains “data-centric” as was our earlier eScience model; that is, it is based on the flow of identity information. However we have refined it in two ways by (a) decomposing the data into three information types common to VOs, and (b) introducing the notion of *producers* and *consumers* of the information. In this section we describe each of these refinements. Section 5 discusses the influential factors we introduced in [4], cast into this refined model. Section 6 illustrates the application of these concepts and how they combine to model IdM in the context of VOs.

4.1 VO Identity Information Types: Data for Supporting Scientific Workflows

Our initial model considered a single flow of user-centric identity information. In analyzing the results from our interviews, we noted there actually exist three different types of user information that are commonly produced and consumed in the context of VO-IdM:

- ***Digital Identifier:*** That is, an identifier of the scientist/VO member issued by an IdP. Examples of this information type include an X.509 distinguished name [13], an eduPerson Principal Name (ePPN) in a SAML assertion [14] and a username.
- ***VO Membership & Role:*** Minimally, each VO produces information about who is a member of the VO. For example, attribute data as captured in Virtual Organization Management Registration Service (VOMRS)[2] system. Some VOs have richer expressions of membership that include a scientist's role(s) and privileges in the VO.
- ***Contact Information:*** Often, but not always, contact information (e.g., email address, phone number, postal address) is collected from a scientist.

The latter two types of information can be, and often are, referred to as "user attributes" [15], and certain attributes are included in the VOMS Attribute Certificate associated with some grid requests [16]. However, we distinguish them because, as we describe subsequently, they are often generated by different parties and utilized for different purposes.

A reader familiar with IdM also will notice we do not include other types of attributes; e.g. a scientist's institution and their role and department at that institution. Our interviews have not revealed evidence of these attributes being in common use in the VO context. There are, at least, two possible reasons for this: (1) there is a lack of demand for this information, that is, it's not useful to VOs or RPs; and/or (2) sufficient information is available by using clues from the email address or the authentication domain associated with the Identity Provider (IdP).

4.2 Identity Production and Consumption: Functions Enabled by IdM

Earlier iterations of our model focused on transmissions of identity information at stages of a VO user's lifecycle, and did not account for the multiple purpose-driven flows of specific types of identity information in VO-RP relationships. What that earlier iteration gained from simplicity, it lost in utility as we began working with VOs to address the mechanics of designing or evolving their IdM implementations. We found it necessary to evolve our model to account for this complexity. For example, consider the multi-user pilot job factory example in Section 4.3: VO membership information is used to authorize a request, the user's identity information is recorded for audit purposes, and contact information is collected and retained for incident response (e.g., user support or security investigations).

Our refined model reflects our observation that identity data is, similar to any data, produced, stored, transformed, transmitted and consumed. As such we turn to the concept of Data Flow Diagrams (DFD) (originally presented in [17] and described concisely in [18][19]) to model the flow of the three types of identity data in the context of VOs. While DFD offers a rich framework, we borrow its simple concepts of entities being *sources* and *terminations* of

[2] http://computing.fnal.gov/docs/products/vomrs/

identity data, though we use the terms *producer* and *consumer* as we believe they more clearly convey the process in the IdM context.

Identity information is produced by administrative action by an entity. In the VO context, producers may include the VO, the RP, or (introducing a new, but well recognized party to our model) an IdP. Each of the three types of identity information can be produced by any one of these three parties. Production may entail generation of previously non-existent information (creating a username) or conversion/translation of existing information into digital form (e.g., recording contact information). For example, some common patterns in the VO context are:

- Identity is produced by an identity provider when a credential is generated for the user.
- VO membership information is produced by a VO when the user successfully applies for membership.
- Contact information is collected by the VO when VO membership is granted to the user.
- In the classic model, the RP would serve as the producer of all three types of information.

Flowing from a producer, identity data may arrive at one or more consumers who use the data for the purpose of providing some service. We have identified seven common functions supported by identity information in the VO context:

- ***Authentication***. Consumes externally provided identity information and produces an internally trusted identity/attribute “bundle” for use by other functions.
- ***Authorization***. Consumes identity information (identity, VO membership/role) to implement access controls on resources.
- ***Allocation / Scheduling of resources.*** Consumes identity information (identity, VO membership/role) to make decisions regarding how to allocate or schedule resources to service a request.
- ***Accounting.*** Consumes identity information to account for resource consumption.
- ***Auditing.*** Consumes and records identity information to allow for the proper decision making regarding a request and to provide information in case user support or incident response is necessary.
- ***User Support.*** A typically manual process that consumes identity information in order to communicate directly with the user initiating a computing workflow in order to resolve some apparent malfunction.
- ***Incident Response.*** A typically manual process that consumes identity information in order to communicate directly with the user initiating a computing workflow in order to resolve a possible security violation.

All consumption may take place at the RP, or, for a function that has been delegated, at the VO. The location of the production and consumption is an indicator of whether responsibilities have been delegated to the VO; information flows between the producers and consumers serves to show what identity information is used for a particular function.

We note that the Data Flow Diagramming methodology can scale to much more complex system descriptions than we set out here, and the method includes a number of concepts that may prove useful in to setting out the fine details of an IdM system design. For example, it defines *stores* (roughly equivalent to databases and credential stores), and processes which can transform data (which seem equivalent to security token services such as CILogon [20]). This assures us that the Data Flow Diagram can be used to reflect highly-complex identity flows if needed (i.e., by using level 2 Data Flow Diagrams), but we resist incorporating them into our model until proven necessary in order to keep it simpler.

4.3 Example Applications of the Model

In this section we apply our model to two case examples. In Figure 1, we show the simplest possible example: a classic implementation with the RP handling all identity management. In this case, two types of identity information are produced and consumed by the RP (and, there is no VO membership information since there is no VO).

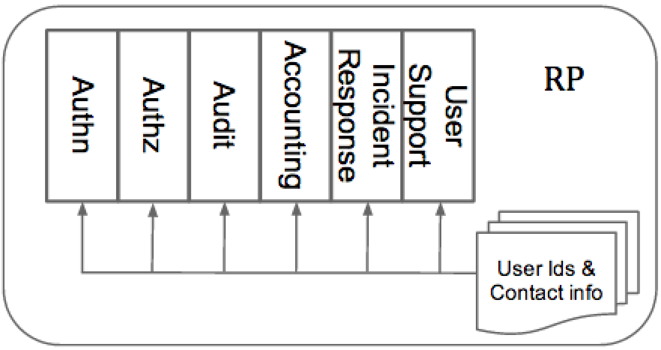


Figure 1*: The classic model with RP handling all identity management.*

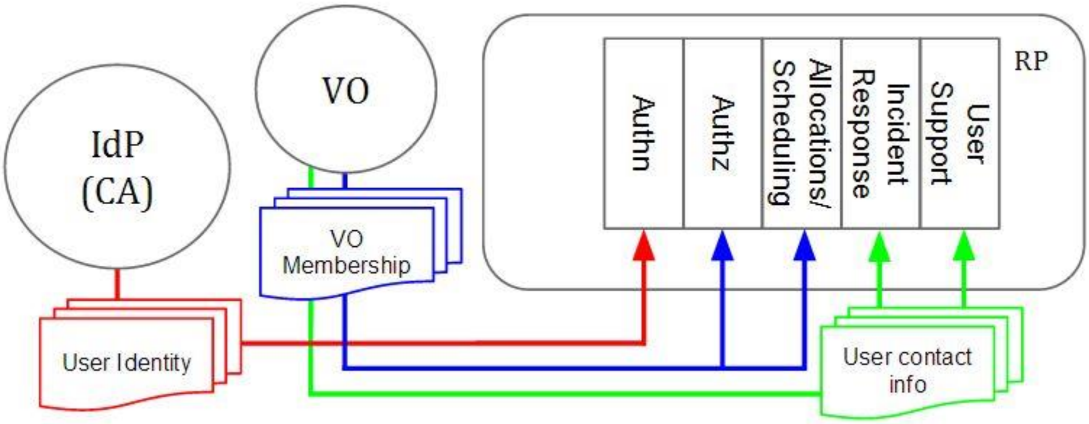


Figure 2*: An example of a multi-user a job factory expressed in our VO IdM Model. User identities come from an identity provider (a certificate authority) are used to authenticate the user's compute job, VO membership information is used by the RP to authorized and allocate resources for the job, and the user's contact information from the VO is used by the RP in the event there is an incident or user support needs to be undertaken.*

Figure 2 shows a more complex model where the RP has retained responsibility for identity management consumption and associated functionality, but has delegated production of user identity to a certificate authority, and determination of VO membership and collection of user contact information to the VO. We believe our model supports the clear expression a complex implementation. It not only conveys the flow of identity information, but also allows for ready inference of the trust relationships and delegations involved.

5. Factors Impacting RP-to-VO Delegation

We now turn to the factors that commonly influence whether and what identity management responsibilities are delegated from RPs to VOs and IdPs. The factors here represent an enhancement of our work in [CHEP] in the following ways: (a) In addition to factors that motivate or disincentivize delegation, we have identified a set of factors that *enable* delegation, but do not in themselves motivate or disincentivize that delegation; (b) we have dropped the “Incentive for Collaboration” factor, which was tenuous and has failed to stand up to validation; (c) we have merged “Isolation of User Data” and “Persistence of User Data” into a more general factor around technology limitations. Other factors remain generally the same with some refinement to reflect further thinking.

5.1 Motivations for Delegation

Scaling and Dynamicity of the VO. Scale can affect the VO/RP relationship in two main ways: The number of RPs involved and the number of users (both total and in terms of turnover) involved in the VO may motivate the parties to delegate production of IdM identity to the VO or an IdP, where it will be centralized instead of replicated at multiple RPs.

Aggregate identity management effort is roughly O(#RPs x #Users) if all the IdM for a service is done at the RP. The more control is centralized to the VO, the more the number of RPs drops out of this equation bringing the effort down to O(#Users). The amount of effort based on #RPs is initially very steep, but once it exceeds a handful of RPs, the mechanisms are typically in place to support a much larger number.

Secondary factors include the number of institutions at which the VO’s members are distributed. This will serve to increase the complexity as more effort is needed to coordinate the users, and the dynamicity of the users in the VOs (and, in theory, the RPs serving the VO), more frequent turnover of users serves to also multiply the distributed effort.

We note that with the inclusion of identity providers in our model, this factor is the main factor influencing the use of a third party identity provider.

Complex VO Member Roles and Privileges. The more heterogeneous the privileges of different VO users, the more complex the access control policies will be and, if RPs are responsible for enforcing those policies, the more complex the communication between the VO and RP will need to be to communicate the policy and necessary information to enforce it. Hence greater complexity of VO roles tends to push authorization functionality to the VO.

VO-wide Collaboration Services. Many VOs have the need to have provide services that support collaboration to their communities: e.g., forums for communication, source code repositories for development, means for sharing and collaboratively analyzing data. Since operating these services requires both effort and identity information (to authenticate and authorize users), this encourages RPs to delegate identity information consumption to the VO so that it can take on this effort.

Alignment with RP's Mission. RPs have their own missions, often heavily influenced by the missions of their funding agencies. In the context of scientific VOs, typically RPs are generally motivated to help VOs achieve science results, though they may be more strongly motivated by VOs tightly aligning with their missions or when specifically funded to help a particular VO. Commercial RPs (e.g., cloud providers) are primarily motivated by payment.

5.2 Enablers of Delegation

This set of factors serve to reduce the barriers to the delegation (i.e., reduce the amount of motivation needed from the first set of factors), but do not themselves motivate delegation.

Established Trust Relationships. When the RP has an established trusting relationship with the VO, this reduces the barriers to the delegation. Examples include a history of prior collaboration, the VO being closely associated with the RP organizationally, and a reputational history of trustworthy VO behavior with other RPs.

Available VO IT/IdM Effort and Expertise. A VO's available IT staff time and expertise in running services (IdM services in particular) is a straightforward, but critical enabler of delegation. A VO that is highly capable, or at least on par with the RP, makes delegation easier. VOs without members with IT expertise, or interest in operating IT services, naturally dissuade delegations of IdM to them.

Availability of Traceability Mechanisms. Increasingly, traceability [21] -- i.e., the ability to trace events back their initiator on an as-needed basis to facilitate user support and security incident response -- is a viable and in-demand mitigation against the reduced RP real-time visibility into user identity that comes along with increased IdM delegation.

5.3 Barriers to Delegation

Historical Inertia and Introduction of Risk. For RPs with a history of doing their own identity management, the delegation of identity management will often require some time for acclimatization. This may also be true for RPs' funding agencies or other stakeholders who set policy for them. These entities frequently have formal policies, informal cultures, and respected reputations around information security and risk which have evolved over time. Delegating even a portion of the information security domain means a change in risk profile, as "decisions to establish trust relationships are expressions of acceptable risk." [2] Our recent interviews with supercomputing centers have reinforced the validity and importance of this factor. We observe these RPs taking more conservative steps, and beginning to delegate IdM to VOs once implementations have proven viable and benign in other settings.

Compliance and Assurance Requirements. IdM-related compliance and/or assurance may present barriers to delegation. Strength of authentication, traceability, auditing, and accounting may be critical responsibilities, and usually lie with the RP by default. Note that external stakeholders of the RP and VO must often be considered here. Stakeholders of RPs, in particular, tend to introduce higher requirements for IdM. There has been some recent relaxation of requirements in recognition, in some cases, that the identity requirements were for persistence and/or valid contact information rather than traceability to a legal identity.

***Technology Limitations*.** The technologies (e.g., software stacks) to be used in the VO/RP context must be considered. Many contemporary tools require identity to function, but allow only for authenticated individual access (e.g., an individual logging in with a username and password or certificate), access by an undifferentiated group to an individual user account, or anonymous access (e.g., a public website, a read-only data server). Some technologies have been extended to allow access by a group to an individual user account while carrying information about the individual user to the RP. For example, this is what VOMS does by embedding a VO credential in a batch job request.

The less sophisticated the technologies in terms of their IdM support, the more effort is required to distribute IdM functionality between two parties and hence encourages IdM to be concentrated at one party or the other (typically the one that is more resourced).

6. Case Study: NERSC Collaboration Accounts

The National Energy Research Scientific Computing Center (NERSC) is the primary scientific computing facility for the Office of Science in the U.S. Department of Energy. As one of the largest facilities in the world devoted to providing computational resources and expertise for basic scientific research, NERSC is a world leader in accelerating scientific discovery through computation. NERSC is a division of the Lawrence Berkeley National Laboratory. More than 5,000 scientists use NERSC to perform basic scientific research across a wide range of disciplines, including climate modeling, research into new materials, simulations of the early universe, analysis of data from high energy physics experiments, investigations of protein structure, and a host of other scientific endeavors.

NERSC supports a very diverse range of science and has over 700 unique projects. Given the large and diverse number of scientific projects supported, there was a strong need to enable collaboration within a virtual organization, in a scalable manner. This case study describes how NERSC is enabling collaboration and sharing of data and jobs within projects running at NERSC. This case study represents a relatively minor change in terms of our model, though a significant step in terms of delegation and trust by NERSC. For an example of a more complex delegation using our model, we refer the reader back to the multi-user job factory example in Section 4.3. We present the case study first described by our co-author (Cholia) from NERSC and then analyzed in our VO IdM Model.

6.1 Use Case

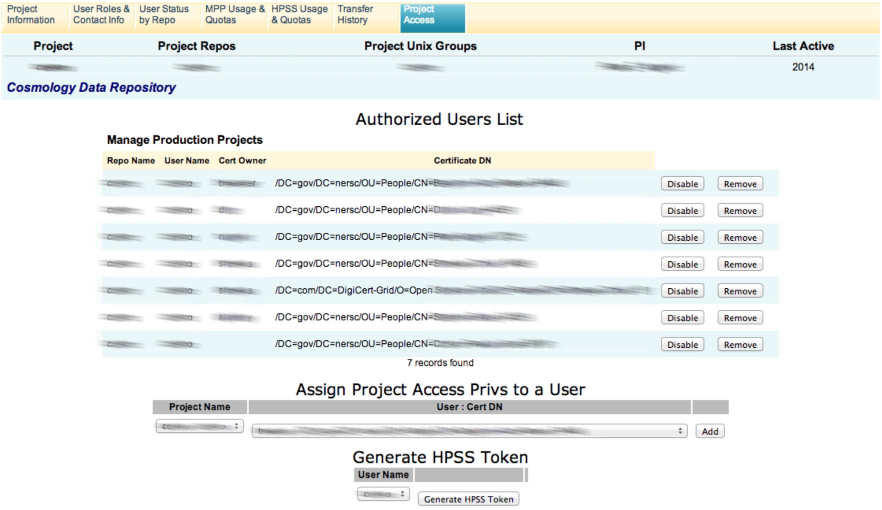
NERSC is implementing a special type of login account called a “collaboration account” to facilitate collaboration and sharing of data. The purpose of the collaboration account is to allow collections of users in a VO to access and manipulate files and jobs run by other members of their VO. Given that a bulk of scientific computing is now performed by teams of scientists, it becomes very important to be able to share access to compute and data resources in a secure, scalable manner.

UNIX and POSIX ACLS have inherent limitations with respect to providing shared access. They do not provide shared control over the batch system for running jobs (e.g., user X starts a long-running job for the collaboration, but user Y needs to stop the job because he finds a problem in the parameters). Incorrectly applying UNIX permissions on files also has the effect of locking out other users in the collaboration.

6.2 User Management

Collaboration accounts allow users to access a shared account to control job and file ownership while still maintaining traceability with respect to the user performing any actions within the system.

VOs requesting collaboration accounts can manage access to this account through the NERSC Information Management (NIM) web portal[3]. Users within the VO can be added as authorized users with access to the collaboration account by the VO's principal investigator or their proxy. This mapping (between user->collaboration account) is propagated to all the systems at NERSC, and users can switch their roles to this collaboration user using a special set of tools.



Project	Project Repos	Project Unix Groups	PI	Last Active
[redacted]	[redacted]	[redacted]	[redacted]	2014

Cosmology Data Repository

Authorized Users List

Manage Production Projects

Repo Name	User Name	Cert Owner	Certificate DN		
[redacted]	[redacted]	[redacted]	/DC=gov/DC=nersc/OU=People/CN=[redacted]	Disable	Remove
[redacted]	[redacted]	[redacted]	/DC=gov/DC=nersc/OU=People/CN=[redacted]	Disable	Remove
[redacted]	[redacted]	[redacted]	/DC=gov/DC=nersc/OU=People/CN=[redacted]	Disable	Remove
[redacted]	[redacted]	[redacted]	/DC=gov/DC=nersc/OU=People/CN=[redacted]	Disable	Remove
[redacted]	[redacted]	[redacted]	/DC=com/DC=DigiCert-Grid/O=Open [redacted]	Disable	Remove
[redacted]	[redacted]	[redacted]	/DC=gov/DC=nersc/OU=People/CN=[redacted]	Disable	Remove
[redacted]	[redacted]		/DC=gov/DC=nersc/OU=People/CN=[redacted]	Disable	Remove

7 records found

Assign Project Access Privs to a User

Project Name	User : Cert DN	
[redacted]	[redacted]	Add

Generate HPSS Token

User Name	
[redacted]	Generate HPSS Token

Figure 3*: NERSC Information Management web portal interface for VO principal investigator to authorize users.(User and Project info are anonymized.)*

6.3 Account Access

It is a requirement for both to NERSC and its VOs to be able to determine which specific individual is running a given process under the common collaboration account. The key is to make sure that users initially authenticate to NERSC as themselves (using their own credentials), and to enable access to the collaboration account through a controlled privilege switching mechanism, once the user has logged in.

[3] http://nim.nersc.gov

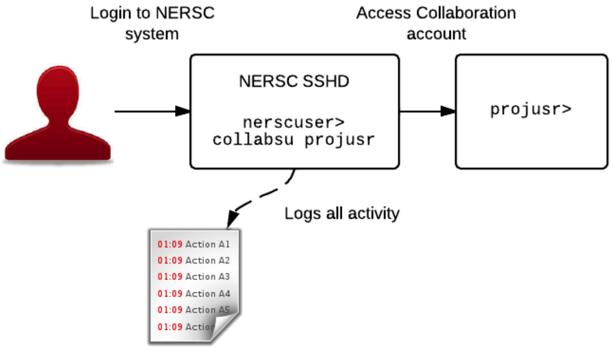
In particular, interactive access to NERSC systems is controlled through a special instrumented SSH daemon where all user commands are logged. NERSC uses a modified version of SSH[4] on all systems that allows NERSC to record and analyze the content of interactive SSH sessions. The data collected with this version of SSH is sent to one of NERSC's security systems where it is analyzed by an intrusion detection system called Bro[5]. The logs from this version of SSH can help determine precisely what a user did during their session.

Crucially, the logs also record the transition between the user account and the collaboration account, so that all future operations can be traced back to the end user.

Access to the collaboration account is handled in one of two ways. When the PI adds a VO member to the collaboration account, the user is (1) added to a control group that manages access to the collaboration account and (2) added to a gridmap file that maps the users X.509 certificate to the project account.

The user can then access the collaboration account, either with a special version of the UNIX "su" (switch user) command or using the gsissh command. The user is granted access to the collaboration account by virtue of being in the control group for that account (entry in sudoers file generated from NIM), or by having a grid mapping between the user's certificate and the collaboration account (generated from NIM). In both cases the user authenticates with the user's own credential (either a password or an X.509 certificate) and then proceeds to access the collaboration account. All actions are now performed under this shared account.

The entrypoint (the NERSC SSHD service) logs all keystrokes and enables us to track actions back to specific users forensically in the event of unusual activity or a security incident.



```
nerscuser>
collabsu projusr
```

```
projusr>
```

```
01:09 Action A1
01:09 Action A2
01:09 Action A3
01:09 Action A4
01:09 Action A5
01:09 Actio
```

***Figure 4**: NERSC's instrumented SSHD enables auditing of a user's actions after they have been authorized based on their VO membership.*

The collaboration account provides a powerful mechanism for VOs to enable users to share access to compute and data resources, while maintaining individual traceability and accountability with respect to specific actions performed under that account.

[4] http://www.nersc.gov/users/accounts/user-accounts/computer-security/instrumented-ssh/
[5] https://www.bro.org/

6.4 Analysis with XSIM Model

6.4.1 Functional Data Flow

Prior to this change, NERSC fit the classic model shown in Figure 1. With the change, the VO now acts as a producer of VO membership information consumed by NERSC to grant access to the collaboration account. Figure 5 shows this simple change in a clear manner.

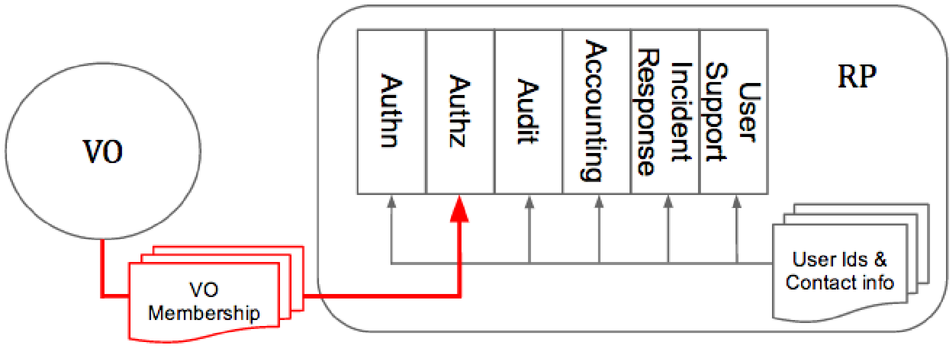


Figure 5*: New model representation for NERSC*
showing VO control of membership information used for authorization at NERSC.

6.4.2 Influential Factors

Motivations. A strong motivator for NERSC's move is the desire to support collaboration at the project/VO level. Closely aligned with that desire is the ability to support project scaling and turnover by putting management of VO membership in the project PI's hands. Complexity of VO member roles and privileges appear to play little or no role at this time.

Enablers. Two enablers clearly show themselves in as key to minimizing the barriers to delegation: First, there are strong, established trust relationships between NERSC and the VOs served. Second, the availability of forensic tracing mechanisms mitigates risks to NERSC associated with decreased *ex ante* visibility into the projects'/VOs' member identities. Available IT and IdM effort at the VO and expertise appears not to be a significant enabler in this case since NERSC supplies the needed mechanisms through NIM.

Barriers. In general, we note that barriers exist to more thorough or complex IdM delegation in the NERSC context, particularly the need to integrate with existing tools and mechanisms at NERSC and concerns regarding risk.

7. Conclusion and Future Work

In this paper we presented a refinement to our previous work presented in [3] and [4], extending our previous one-dimensional model based on a VO user lifecycle to a more expressive model based on three types of identity information commonly found in VOs (user identifier, VO membership/role, and user contact information) and expressing that, by leveraging the concept of data flow diagrams [17][18][19], as information flows between data producers and data consumers. Data consumers use the information to provide functions related to identity management: authentication, authorization, allocations/scheduling, auditing, incident response and user support. The graphical representation, as shown by figures in this paper, allows a ready communication of what aspects of IdM have been delegated.

This work represents progress towards XSIM's ultimate goal of providing VOs and RPs with the ability to more clearly articulate their relationships in terms of IdM and providing guidance in how to shape those relationships based on commonly influential factors. The next steps are to continue our work with VOs and RPs to refine and validate this work, continue developing our understanding of the specific relationships between the motivating factors and the delegation of individual identity management information production and consumption, and explore the ramifications of that when failures occur in trust relationships.

8. Acknowledgments

In addition to our continued appreciation to interviewees in our prior work, we thank John Allen, Ashley Barker, Tom Barron, David Groep, Jens Jensen, Tony Johnson, Ross Miller, Susan Parete-Koon, and Adam Slagell. We thank those who attended XSIM talks at LBNL, NERSC and ISGC, and provided feedback on our efforts. We thank the US Department of Energy Next-Generation Networks for Science (NGNS) program (Grant No. DE-FG02-12ER26111) for funding this effort. Also, this research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. All opinions are those of the authors and do not necessarily reflect those of the Department of Energy or any other organization or person (including the interviewees).

References

[1] I. Foster, C. Kesselman, S. Tuecke, *The Anatomy of the Grid: Enabling Scalable Virtual Organizations (PDF), International Journal of Supercomputer Applications* (May 10, 2001). Retrieved October 15, 2011.

[2] NIST Special Publication 800-39, *Managing Information Security Risk,* March 2011.

[3] R. Cowles, C. Jackson, V. Welch. *Identity Management for Virtual Organizations: A Survey of Implementations and Model, 9th IEEE International Conference on eScience*, 2013, http://www.computer.org/csdl/proceedings/escience/2013/5083/00/5083a278-abs.html.

[4] R. Cowles, C. Jackson, V. Welch. *Identity management factors for HEP virtual organizations. 20th International Conference on Computing in High Energy and Nuclear Physics (CHEP2013)*, 2013, http://www.vonwelch.com/pubs/CHEP2013.

[5] S. Landau , T. Moore, *Economic tussles in federated identity management, First Monday*, vol. 17, no. 10, October 2012, http://journals.uic.edu/ojs/index.php/fm/article/view/4254/3340.

[6] D. Broeder, et al., *Federated identity management for research collaborations, CERN-OPEN-2012-006*, Apr 23, 2012, http://cds.cern.ch/record/1442597?ln=en.

[7] M. Altunay, *How OSG Resource Providers Consume Identity Information, unpublished presentation to the MAGIC committee*, Dec. 4, 2012.

[8] A. Lin, E. Vullings, J. Dalziel, *A trust-based access control model for virtual organizations," in IEEE Proc. Fifth Int. Conf. Grid and Cooperative Computing Workshops (GCCW'06).*

[9] K. Klingenstein, K. Hazelton, *An Introduction to Identity and Access Management, TERENA EuroCAMP Identity Management and Federated Access*, 7 - 9 November 2005, Porto, Portugal. http://www.terena.org/activities/eurocamp/nov05/slides/day1/introtoidm.ppt.

[10] *JISC Identity Management Toolkit, undated and unversioned document*, visited April 6, 2014, http://www.jisc.ac.uk/media/documents/programmes/aim/IdMToolkit.pdf.

[11] The Identity Ecosystem Steering Group (IDESG) Functional Model Working Group, *Functional Elements, undated and unversioned document*, visited April 6th, 2014. https://www.idecosystem.org/wiki/File:Functional_Elements_(Package_for_Committee_Review).pdf.

[12] Microsoft Corporation, *Microsoft's Vision for an Identity Metasystem*, May 2005, http://msdn.microsoft.com/en-us/library/ms996422.aspx.

[13] D. Cooper, et al., *RFC 5280: Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile, IETF*, May 2008.

[14] Internet2 Middleware Architecture Committee for Education, Directory Working Group (MACE-Dir), *Internet2-mace-dir-eduperson-201203*, April 23, 2012, http://middleware.internet2.edu/eduperson/docs/internet2-mace-dir-eduperson-201203.html

[15] *InCommon Federation Attribute Summary, undated document*, http://www.incommonfederation.org/federation/attributesummary.html.

[16] V. Ciaschini, *A VOMS Attribute Certificate Profile for Authorization*, April 3, 2007, http://egee-jra1-data.web.cern.ch/egee-jra1-data/glite-stable/stage/share/doc/voms/AC-RFC.pdf.

[17] W. Stevens, G. Myers, L. Constantine, *Structured Design*, *IBM Systems Journal*, 13 (2), 115-139, 1974.

[18] *Data Flow Diagram*. http://en.wikipedia.org/wiki/Data_flow_diagram.

[19] D. S. Le Vie, Jr, *Understanding Data Flow Diagrams, undated document*, visited April 6, 2014, http://ratandon.mysite.syr.edu/cis453/notes/DFD_over_Flowcharts.pdf.

[20] J. Basney, T. Fleury, J. Gaynor, *CILogon: A Federated X.509 Certification Authority for Cyberinfrastructure Logon, Concurrency and Computation: Practice and Experience*, 2014, http://dx.doi.org/10.1002/cpe.3265.

[21] G. Garzoglio, A. Padmanabhan, M. Altunay, K. Hill, *Running Jobs without End User Certficates: Assessing Traceability of User Jobs In the Grid, Presentation at International Symposium on Grids and Clouds 2014*, 23-28 March 2014, Academia Sinica, Taipei, Taiwan, http://indico3.twgrid.org/indico/contributionDisplay.py?sessionId=29&contribId=133&confId=513