DataONE

Identity Management System Review

October 8, 2013

For Public Distribution

Jim Basney, Patrick Duda, Von Welch, Craig Jackson

About CTSC

The mission of the Center for Trustworthy Scientific Cyberinfrastructure (CTSC, trustedci.org) is to improve the cybersecurity of NSF science and engineering projects, while allowing those projects to focus on their science endeavors. This mission is accomplished through one-on-one engagements with projects to address their specific challenges; education, outreach, and training to raise the state of security practice across the scientific enterprise; and leadership on bringing the best and most relevant cybersecurity research to bear on the NSF cyberinfrastructure research community.

Acknowledgments

CTSC's engagements are inherently collaborative; the authors wish to thank the DataONE team, and specifically Ben Leinfelder, Bruce Wilson, and Dave Vieglais for the collaborative effort that made this document possible.

This document is a product of the Center for Trustworthy Scientific Cyberinfrastructure (CTSC). CTSC is supported by the National Science Foundation under Grant Number OCI-1234408. For more information about the Center for Trustworthy Scientific Cyberinfrastructure please visit: http://trustedci.org/. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Using & Citing this Work

Cite this work using the following information:
J. Basney, P. Duda, V.S. Welch, C. Jackson, "DataONE Identity Management System Review," Center for Trustworthy Scientific Cyberinfrastructure, trustedci.org, Oct 2013. Available: http://hdl.handle.net/2022/16926.

This work is available on the web at the following URL:
http://trustedci.org/dataone/.

Table of Contents

1 Executive Summary

DataONE (Data Observation Network for Earth, http://www.dataone.org/) is a federated data network built to improve access to Earth science data, and to support science by: (1) engaging the relevant science, data, and policy communities; (2) facilitating easy, secure, and persistent storage of data; and (3) disseminating integrated and user-friendly tools for data discovery, analysis, visualization, and decision-making.

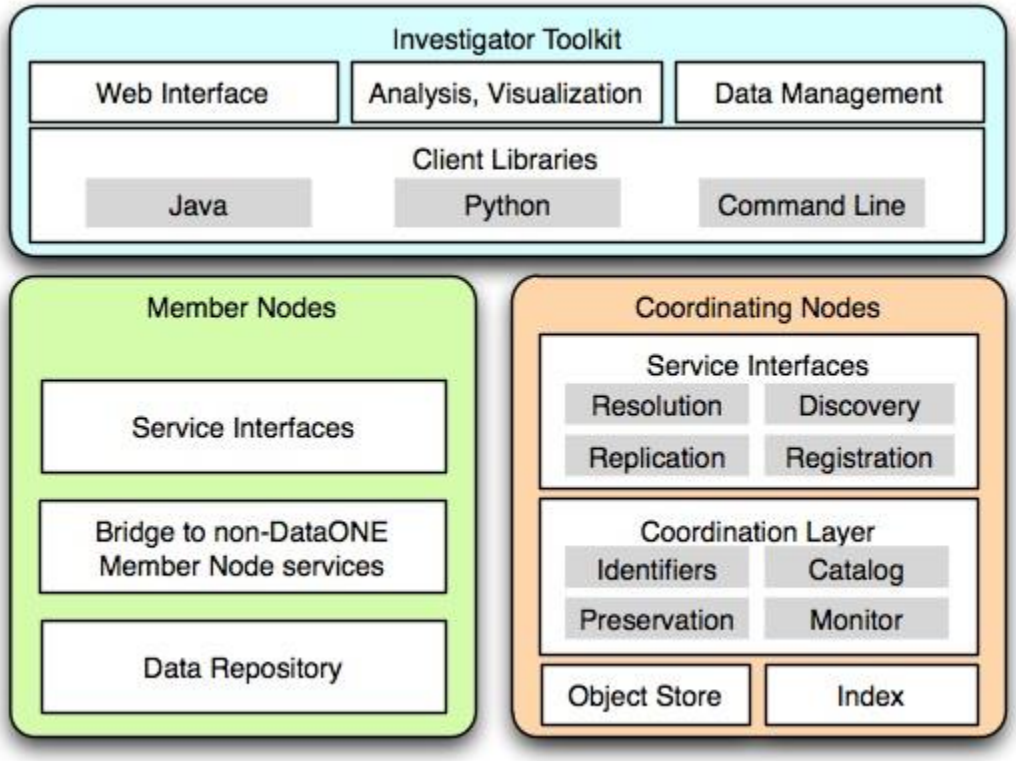
DataONE has invested significant effort into the development of an identity management (IdM) system that supports federated identities from a wide variety of identity providers and includes mechanisms and procedures to support access management.

The Center for Trustworthy Scientific Cyberinfrastructure (CTSC) conducted a design-level review of the DataONE IdM system implementation. This document represents the findings and recommendations from the review, which are in summary as follows:

- Documentation Issues
  - Missing and Out-of-Date Documentation
  - Missing Policies and Expectations
  - CILogon Extension Not Documented
  - Definition of "Verified User" Lacking
- Architectural Issues
  - Bindings of DataONE User Identities
  - Lack of Qualifier for Local User Identities
  - Supporting International Users
  - Managing Groups
  - Browser Authentication for Member Nodes
- Operational issues
  - Incident Response Plan
  - Backup of IdM Data
  - Logging
  - Intrusion Detection
  - Contacting Users
  - Dependency on CILogon
- Recommendations
  - Short-Term Recommended Tasks
  - Medium-Term Recommended Tasks
  - Longer-Term Recommended Tasks
  - Issues to be Tracked

2 Overview of DataONE

DataONE (Data Observation Network for Earth, http://www.dataone.org/) is a federated data network built to improve access to Earth science data, and to support science by: (1) engaging the relevant science, data, and policy communities; (2) facilitating easy, secure, and persistent storage of data; and (3) disseminating integrated and user-friendly tools for data discovery, analysis, visualization, and decision-making.



There are three major components in the DataONE infrastructure: Member Nodes which represent data repositories, Coordinating Nodes which serve data management and discovery services, and the Investigator Toolkit which contains a variety of end user tools for interacting with the infrastructure. For the purpose of the CTSC review, focus was put only on the Member and Coordinating Nodes since these items participate in the Identity Management System (IdM).

Participation in the DataONE infrastructure as a Member Node (i.e., implementing or utilizing DataONE service interfaces) provides several fundamental services upon which additional infrastructure, services, applications and communities may be built. These core community building services include:

- promotion of data preservation through automated replication of data and metadata

- support for arbitrary globally unique identifiers with guaranteed resolution and dereferencing
- extensible search and discovery services
- federated management of user identities and access control

Member Nodes primarily are existing data repositories (e.g., Dryad[1], the Knowledge Network for Biodiversity[2], ORNL DAAC[3]) that already fill an important role in their respective communities supporting data management, curation, discovery, and access functions. Existing or new repositories can participate in the DataONE infrastructure by implementing a simple set of APIs (Application Programming Interfaces) which represent a convergence of functionality expressed in a variety of existing systems. These APIs include basic operations such as listing and retrieving objects, support for creation of content, and the ability to generate low level system metadata describing the various objects (data, metadata) exposed by the service. Member Nodes may implement a subset of the full suite of *Member Node APIs*[4], and in this way participate in the network with minimal effort (e.g., as a "read only" data source). Member Nodes that implement the full suite of APIs will be able to accept data from other Member Nodes which assists with data preservation by ensuring multiple copies of all content are available, thus reducing the risk that content will be lost or inaccessible if a Member Node should go offline.

Member Nodes may eventually number in the thousands as progressively smaller repositories come online, perhaps even to the level of individual labs deploying their own Member Node to take advantage of the broad infrastructure enabled by DataONE.

Coordinating Nodes implement critical services through the *APIs*[5] that enable identifier resolution, data preservation, data discovery, and supplement the federated identity system. Coordinating Nodes replicate all content between themselves, and in doing so create a small set (3-6 Nodes) of geographically and institutionally distributed systems that ensure ongoing operation of the infrastructure should any particular node be inaccessible. Coordinating Nodes maintain complete copies of all science metadata (detailed descriptions of science data objects and collections) and system metadata (low level metadata describing the type, size, ownership, and locations of data) and index this information to enable data discovery services.

[1] http://datadryad.org/
[2] http://knb.ecoinformatics.org/index.jsp
[3] http://daac.ornl.gov/
[4] http://mule1.dataone.org/ArchitectureDocs-current/apis/MN_APIs.html
[5] http://mule1.dataone.org/ArchitectureDocs-current/apis/CN_APIs.html

3 DataONE Identity Management System

The goals of the DataONE identity management (IdM) system are:

- Stay out of the credentialing business and not weaken authentication provided by identity providers.
- Maintain the autonomy of the Member Nodes (MNs), including those with legacy systems.
- Allow the Coordinating Nodes (CNs) to service searches in a way that predicts access control implemented by MNs.
- Give users control of equivalency of their identities.
- Allow MNs to control equivalency of their locally-managed identity and other user identities.
- Control access to restricted (non-public/moderate sensitivity) data.
- Support both web and command-line clients.
- Be sustainable for a century.

Identity Management for DataONE addresses the need to identify users that request the use of services and/or data/metadata resources within the DataONE virtual organization. DataONE recognizes that not all services/resources require user identification; thus, support for anonymous access to certain services/resources is possible using a Public identity. DataONE provides services for users to register their identity with DataONE in a user account so as to create a unique DataONE identifier, along with other attributes about that user. This account information may be used for authorization and logging of DataONE transactions.

Users may have multiple identities as a result of distributed research endeavors at different participating organizations and/or changes in organizational affiliation. Because of this, the DataONE Identity Management service will support user identity mappings, which allows users to authenticate using any one of their multiple identities, but still be recognized as the same DataONE identity. When a DataONE authenticated session begins, information pertaining to the user's identity is available for authorization purposes, which includes a listing of all mapped identities associated with that user. In general, these mapped identities serve equally well for authorization decisions—that is, within DataONE access control policies, reference to any mapped identity is the same as using any other of the user's identities. However, it is possible that some MNs will authorize some operations based only on the user's primary/active identity.

In addition, the Identity Management service provides a system for users and administrators to create, store, and modify groups of users that can be used in access control directives. Only the user creating a group will be allowed to delete the group or change the group's membership.

4 CTSC Findings

The results of the CTSC review are broken up into two parts: Strengths and Issues.

4.1 Strengths

The overall assessment of the DataONE (IdM) system was positive. There are many strengths in the current DataONE IdM design, including support for authentication using federated identities, equivalence mapping of multiple identities for the same person, and a well-specified access policy language. The IdM system design is a good match for the system goals (specified above).

Of particular note were the system's operational strengths. There are many good safeguards and security practices in place:

- All administrator access uses a separate `LDAP` service.
- Strong `sudo` practices are being used.
- Security updates are being applied.
- There is multi-master mirroring of Coordinating Nodes and `LDAP`.
- All web access is done via `HTTPS`.
- Good change control practices are in place.
- Strong testing is done before rollout of updates.
- Procedures are clearly defined for how the system will update.
- Use of `syslog`.
- Good controls over DataONE CA root keys.
- Physical access to servers is limited and controlled.
- Disaster recovery plan in place and being followed.
- Server certificates issued by offline DataONE CAs are delivered securely (by SCP).

4.2 Issues

The IdM system issues have been organized into three broad categories: Documentation, Architectural, and Operational.

The Documentation section delineates issues found in the current documentation, policies and procedures. The Architectural section addresses issues found in the design and overall architecture of the DataONE IdM system. Finally, the Operational section addresses perceived weaknesses in the operational security of the DataONE IdM system.

4.2.1 Documentation Issues

4.2.1.1 Missing and Out-of-Date Documentation

Overall, DataONE's IdM documentation is well done and in good order. However, during the review, it was noted that there were some issues with the documentation. There are several documentation areas marked with "TODO". Also, some of the documentation no longer reflects the current functioning or design of the system.

4.2.1.2 Missing Policies and Expectations

The relationship between the Coordinating Nodes, Member Nodes and the DataONE user community involves trust and expectations that is presently implied. The following policies stand out as absent:

- An acceptable use policy (AUP) that DataONE users agree to abide by. Some example AUPs that DataONE may want to consider in drafting their AUP include:
  - Open Science Grid: http://osg-docdb.opensciencegrid.org/cgi-bin/ShowDocument?docid=86
  - XSEDE: https://www.xsede.org/usage-policies
  - iPlant: http://www.iplantcollaborative.org/node/1934
- A privacy policy which states how DataONE will and will not collect, use and share data regarding its users.
- A set of expectations stating how CNs and MNs behave with regard to IdM.

4.2.1.3 CILogon Extension Not Documented

One of the key interfaces in the system is the XML that is added to the CILogon certificate. This XML expresses the identities and privileges of the users and is intended to be consumed by CNs and MNs. The reviewers did not find a schema or definition for this structure in the current documentation. Currently, DataONE has written all the software that parses this extension, but if an MN needs to write software that does so, they will need documentation on its syntax and semantics.

4.2.1.4 Definition of Verified User Lacking

The term "verified user" is found in the documentation but there is no definition of what this term really means. There does not seem to be an explanation of how one really becomes verified and who performs the verification.

4.2.2 Architectural Issues

4.2.2.1 Bindings on DataONE User Identities

DataONE supports binding of user identities (i.e., marking two user identities as equivalent) to support users with multiple identities. Use cases for identity binding include: 1) users wanting to use different identity providers (e.g., Google, InCommon) at different times, 2) users migrating from one identity provider to another (i.e., changing university affiliation), 3) user name changes, 4) users losing access to a previously used identity (e.g., lost password, identity provider out-of-business).

The CTSC reviewers note two concerns with the current identity binding design. First, DataONE does not bind external user identities (e.g., CILogon distinguished names) to DataONE-internal identities, but rather uses external user identities directly. This can potentially increase the disruption caused by changes in external identity providers.

However, the CTSC reviewers acknowledge that using external user identities directly simplifies the internal DataONE system design and implementation. Second, bindings between user IDs (i.e., one user with multiple IDs) are pairwise and not transitive in DataONE. To use a new identity in DataONE without losing existing authorizations and group memberships, the user must bind that new identity explicitly to each identity the user has previously registered with DataONE. As user identities change over time, this binding process will become increasingly cumbersome, particularly when users no longer have access to old identity providers to establish bindings to old identities. However, making identity bindings transitive increases risks, as the compromise of any identity would spread transitively to all bound identities for that user. Different user identities may have different privileges, and DataONE MNs can decide locally whether to trust identity bindings. In conclusion, the CTSC reviewers acknowledge the design trade-offs in the DataONE identity binding design and recommend that DataONE re-evaluate these trade-offs as more operational experience is gained.

4.2.2.2 Lack of Qualifier for Local User Identities

The DataONE IdM system does not qualify local MN identities with an indication of their origin (i.e., to make the identities globally unique). This could lead to inappropriate authorization for replicated data if two users were to have the same identity at two different MN sites, i.e., it is unclear if *jsmith* at MN1 is the same person as *jsmith* at MN2. This issue becomes critical as data is replicated across MNs with local MN identities in authorization policies.

4.2.2.3 Supporting International Users

As the DataONE project moves forward, it plans to expand both its user base and Member Nodes to include international partners. This will bring challenges of international identity federation and require adjusting to international privacy expectations/laws (e.g., in the European Union). CILogon supports international users using OpenIDs (i.e., Google IDs) but does not otherwise support identity providers outside InCommon. DataONE needs a plan for supporting identity providers outside the US.

For international support of SAML web single sign-on, CTSC's recent work with LIGO to explore options for international identity federation (https://dcc.ligo.org/LIGO-G1300768/public) can be a useful reference for DataONE. In summary, CTSC and LIGO found that international interfederation among research and education identity federations (for example, InCommon in the US) is a work in progress. InCommon is currently considering joining eduGAIN (http://edugain.org/), which interconnects identity federations around the world. Until InCommon joins eduGAIN, it is necessary for research projects to join national identity federations one-by-one for international interoperability or even worse, to coordinate directly with international identity providers one-by-one.

For international support of X.509 certificate authentication, the International Grid Trust Federation (http://igtf.net/) provides a community of certification authorities serving academic researchers. CILogon participates in the IGTF, and the CILogon certification authorities are included in the IGTF trust anchor distribution.

Additionally, identity providers around the world are adopting the standard OAuth interface supported by CILogon, primarily for integration with Globus Nexus (https://www.globusonline.org/). Examples include University of Exeter (UK), European Grid Infrastructure, and University of Canterbury (New Zealand). DataONE could integrate directly with these additional OAuth providers (besides CILogon) and/or integrate with Globus Nexus for support for these additional providers as they come online.

In summary, providing support for international users beyond the existing Google ID support poses significant technical challenges. Work in progress by InCommon, Globus, IGTF, and others has the potential to address these challenges in the future. DataONE could actively engage with these leading-edge efforts (potentially in collaboration with CTSC), or DataONE could monitor this ongoing work and wait until more complete solutions are available (for example, when InCommon joins eduGAIN).

4.2.2.4 Managing Groups

The DataONE IdM system provides a very basic user group capability, with an unmanaged namespace and no method for transferring group ownership. To support user groups appearing in long-lived data access control policies, the CTSC reviewers recommend globally unique group names and improved group management capabilities (including transfer of ownership and multiple owners of a group).

4.2.2.5 Browser Authentication for Member Nodes

DataONE's design for user browser-based authentication to member nodes is unclear. The current DataONE design supports certificate-based authentication, but certificates in browsers are notoriously difficult to manage. The possibility of a web single sign-on mechanism (OpenID, SAML, etc.) between the CNs and MNs was discussed during the review and requires further design work.

4.2.3 Operational Issues

Operational security is an important part of total system security. As noted in the Strengths section, DataONE has many good operational mechanisms in place; however, a few operational issues were found while reviewing the overall system.

4.2.3.1 Incident Response Plan

DataONE currently lacks an incident response plan that describes when and how Coordinating and Member Nodes communicate during a security incident.

4.2.3.2 Backup of IdM Data

DataONE's user and group data is not being backed up at the CNs, though it is replicated across the CNs. As this data potentially includes personally identifying information, any future plans for backing up this information should include appropriate safeguards (e.g., encryption of backups). The DataONE disaster recovery plan should address IdM data including recovery of identity bindings and group memberships that are relied upon in long-lived data access control policies.

4.2.3.3 Logging

DataONE currently uses decentralized system logging that may not include complete logging of identity mappings and group memberships (such as when mappings/memberships were added/removed and by whom). CTSC recommends centralized log collection to assist with incident response and disaster recovery.

4.2.3.4 Intrusion Detection

At the time of the assessment there is a lack of intrusion detection done at the Coordinating Nodes (CNs), but that is being addressed. The user and group information managed by the CNs is critical for appropriate access control and should be appropriately monitored for intrusions using network- and host-based intrusion detection.

4.2.3.5 Contacting Users

Currently DataONE users are not required to provide contact information upon registration, and the user email addresses provided by CILogon are not recorded by DataONE. While it is laudable for DataONE to avoid unnecessary collection of personal information, having user contact information is valuable for incident response and support purposes. The CTSC reviewers recommend that DataONE consider requiring collection of user email addresses for these purposes.

4.2.3.6 Dependency on CILogon

CTSC notes that the DataONE IdM system is highly dependent on CILogon. A major natural disaster impacting the University of Illinois at Urbana-Champaign or a major funding or personnel change impacting the CILogon project could have major negative impact on DataONE. In collaboration with DataONE and XSEDE, the CILogon project is pursuing geographic redundancy for the CILogon service for fail-over and disaster recovery, with the National Institute for Computational Sciences (NICS) at the University of Tennessee as a potential backup hosting site. The NSF XSEDE project and DOE grant DE-SC0008597 support CILogon operations at the University of Illinois through 2015. CTSC reviewers recommend that DataONE consider establishing a service level agreement with CILogon (possibly through XSEDE) for long-term support. CILogon uses open source software, so the possibility exists that DataONE could operate a CILogon service of its own in case the CILogon project at Illinois is unable to meet DataONE's needs. Alternatives to CILogon such as Globus Nexus could also be considered.

5 Recommendations

The CTSC reviewers make the following recommendations to DataONE regarding its IdM system. The recommendations are grouped into short (less than a couple weeks), medium (less than a couple months) and longer-term tasks. Within each group, tasks are prioritized into high, medium and low priority. We also include a set of recommendations on issues to monitor for potential problems.

5.1 Short-Term Recommended Tasks

- (R1)(HIGH) Clarify "Verified User" as described in 4.2.1.4. The term "verified user" is not clearly defined in the documentation. This definition should also include use cases showing how a user becomes verified. Architectural design and API information should be included.

5.2 Medium-Term Recommended Tasks

- (R2)(HIGH) Complete system design: Section 4.2.2.2 and 4.2.2.5 describe gaps in the current IdM system design that should be rectified.

- (R3)(HIGH) Complete documentation regarding policies and expectations between Coordinating Nodes, Member Nodes and DataONE users as described in 4.2.1.2. A more complete description of the roles and trust agreements between the Coordinating and Member Nodes and the users should be included. This description should detail which Node does what with regard to access, authorization and the management of the user's identity, along with how identity mappings are maintained, created and used. Policies and procedures that DataONE needs to develop include:

    - Member Node Memorandum of Understanding (MOU), i.e., DataONE's expectations for MN security practices
    - Privacy Policy, i.e., how DataONE manages user data
    - Acceptable Use Policy, i.e., expectations of user behavior

- (R4)(MEDIUM) Author incident response policies and procedures to address 4.2.3.1. As noted in the "Documentation" section, an incident response plan should be developed. This plan should cover incidents such as inappropriate content being uploaded and stored on systems. Member Nodes need to know their responsibilities in monitoring and reporting malicious activities. The mechanism for reporting these activities should also be covered in the plan. If an incident occurs on one of the Coordinating Nodes, this needs to be communicated to the Member Nodes and users. The lack of a clearly defined response plan is a risk to the current system.

  See Appendix D for additional guidance on developing incident response plans.

- (R5)(MEDIUM) Logging of identity mappings as described in 4.2.3.3. Review of the audit logs should be performed. Verify that good information is being logged when it comes to identity mappings and group membership. Verify that information on when and by whom these mappings were created is captured.
- (R6)(LOW) Complete and update documentation as described in 4.2.1.1 and 4.2.1.4. Updates should be made to documentation to remove all sections that are marked "TODO". These "TODO" sections should be filled-in with current information or moved to their own section in the design addressing future work. The documents should also be reviewed by the DataONE team, and updated and corrected with current information and designs. In the event a Member Node wants to implement its own identity-handling or access control software, it will be useful to have a full description of the XML extension included by CILogon. This description should include a schema specification and detailed information on each section and element of the XML document.

5.3 Longer-Term Recommended Tasks

- (R7)(MEDIUM) Deploy an intrusion detection system for the Coordinating Nodes per 4.2.3.4. Currently, there is no intrusion detection system in place. Intrusion detection should also be part of the overall incident response plan, including the steps to be taken in response to a detected intrusion. See Appendix B for additional guidance on intrusion detection.
- (R8)(MEDIUM) Develop a contingency plan for CILogon per 4.2.3.6. Possible options include establishing a replica of CILogon for both redundancy and to bring expertise in-house or exploring the use of Globus Online/Nexus.
- (R9)(MEDIUM) Backup IdM data as described in 4.2.3.2. Perform encrypted backups of user and group data from the identity management system for disaster recovery.
- (R10)(LOW) Centralized logging as described in 4.2.3.3. Centralized auditing might be considered to help with monitoring the system. See Appendix C for guidance on centralized logging.
- (R11)(LOW) Track efforts in international identity federation and privacy issues as described in 4.2.2.3.

5.4 Issues to be Tracked

CTSC noted the following attributes of the DataONE IdM system which do not obviously need remediation, but should be monitored closely by the DataONE team.

- (R12) Lack of DataONE-internal user identifiers (4.2.2.1).
- (R13) Lack of user contact information (4.2.3.5).

Appendix A: DataONE/CTSC Engagement Process

At the request of DataONE (http://www.dataone.org/) CTSC was asked to perform an architectural review of DataONE's current Identity Management System (IdM). CTSC agreed upon the following goals of the review with DataONE:

- Identify the specific software components to be reviewed.
- Identify (or assist in creating) documentation of those components sufficient to review them.
- Perform and document an assessment of those components that is of use to DataONE.
- Assess potential vulnerabilities, scalability, interoperability and supportability

The focus of the review was on identity management for DataONE's data infrastructure. This includes Member Nodes, Coordinating Nodes, CILogon, certificate management, and identity mapping. The DataONE Identity API and Authorization API are the focus. Other off-the-shelf components such as LDAP are less of a concern and therefore not the focus of the review. Furthermore, the DataONE project internal IDM system for project collaboration is out of scope.

The following plan was developed for the review:

1. Discuss with members of the DataONE team to understand both the DataONE IdM system and its role in the overall DataONE system.
2. Identify the specific scope of the DataONE IdM system to be assessed.
3. Work to identify documents related to the DataONE IdM system.
4. Work to document any portions of the DataONE IdM system that are missing, incomplete, or lack sufficient detail for assessment. The onus for creating such documentation lies with DataONE, though CTSC can help.
5. CTSC will then undertake an assessment of the system, with DataONE staff being available to answer questions and provide additional details if needed.
6. CTSC will document their assessment and deliver an initial draft to DataONE.

The review process ended up taking the form of a 2 day series of face-to-face meetings between the CTSC and DataONE teams. These meetings were held in an informal ad-hoc style. Prior to the face-to-face meetings, the CTSC team reviewed documents identified by DataONE as being valuable to understanding the IdM system. After reviewing the documents a basic system characterization was developed. The review of these documents has been included in this report.

During the face-to-face meetings, DataONE engineers presented their design and use cases for the IdM system. Through this presentation CTSC personnel were able to ask questions and gather information needed to understand the system as it currently exists and plans for enhancements. Prior to the end of the first day of meetings, a summary of findings and issues was presented by CTSC. Based on those findings and questions from DataONE, the

schedule for the next day's meeting was determined. The second day meeting consisted of CTSC staff presenting suggested architectural changes to the current IdM and this prompted more information being revealed by the DataONE staff. During this second day several important issues were identified.

The results of the document review and 2 days of meetings are presented in this report.

Appendix B: Intrusion Detection

Available open source intrusion detection systems include:

All Inclusive	Security-onion (http://code.google.com/p/security-onion/)
Host Based	OSSEC (http://www.ossec.net/) Samhain (http://la-samhna.de/samhain/) Tripwire (http://sourceforge.net/projects/tripwire/)
Network Based	Snort (http://snort.org/) Suricata (http://suricata-ids.org/) Bro (http://bro-ids.org/)

Security Onion

Security-onion (SO) is a very useful package of open source tools to help with network security monitoring. SO provides three core functions: full packet capture, network and host-based intrusion detection systems (NIDS and HIDS), and analysis tools. This provides visibility into network traffic and adds context around alerts and anomalous events. It's based on Ubuntu and contains Snort, Suricata, Sguil, Squert, Snorby, Bro, NetworkMiner, Xplico, and many other security tools.

Since SO works off of full-packet capture, disk space can become an issue. SO allows for some tuning of this capture. For example, you can ignore things like your backup server but for the most part all packets are being captured all of the time. SO does perform some disk management for you. When the data reaches 90% of disk space, SO begins to purge out the captured packets and log files that it has been saving and using. With this in mind, adequate sizing should be used when setting the system up.

Currently, SO is only offered on Ubuntu. SO is released as an ISO and can be easily installed as a VM. This is the recommended installation practice too. Analysts are encouraged to install SO as a VM running on their workstation.

While automation and correlation can enhance intelligence and assist in the process of sorting through false positives and malicious indicators, there is no replacement for human intelligence and awareness. SO only provides the tools; the administrators must actually monitor what is going on. You cannot install the system and walk away thinking things are safe.

Appendix C: Centralized Logging

Centralized logging aggregates in a central location multiple log files from multiple of servers, which is very helpful in comparing logs across systems when investigating security incidents. Logging to dedicated log collector servers also provides redundancy in the case local system logs may have been manipulated by an attacker.

Security Onion provides the ELSA centralized syslog framework (https://code.google.com/p/enterprise-log-search-and-archive/), which supports log queries, alerts and dashboards.

Examples for setting up centralized logging using syslog-ng can be found at the following:

- http://www.balabit.com/sites/default/files/documents/syslog-ng-ose-3.3-guides/en/syslog-ng-ose-v3.3-guide-admin-en/html/syslog-ng.conf.5.html
- http://www.deer-run.com/~hal/sysadmin/SSH-SyslogNG.html
- http://www.enterprisenetworkingplanet.com/netsysm/article.php/3596656/Build-a-Secure-Logging-Server-with-syslogng.htm
- http://www.enterprisenetworkingplanet.com/netsysm/article.php/3598146/Build-a-Secure-Logging-Server-with-syslogng-Part-2.htm

Examples for setting up centralized logging using rsyslog can be found at:

- http://www.rsyslog.com/receiving-messages-from-a-remote-system/
- http://www.rsyslog.com/sending-messages-to-a-remote-syslog-server/

For systems that are logging information via Log4j, SyslogAppender can be used. Good examples for configuring and running with this setup can be found at:

- http://kazed.blogspot.com/2009/12/using-syslog-appender-in-log4j.html
- http://blog.trifork.com/2010/01/14/logging-to-the-syslog-from-a-java-application/

Appendix D: Incident Response

The Open Science Grid wiki provides a good example of an incident response process tailored to distributed cyberinfrastructure, along with additional incident response references:

- https://twiki.grid.iu.edu/bin/view/Security/IncidentResponseProcess
- https://twiki.grid.iu.edu/bin/view/Security/IncidentResponseReferences

Another potentially useful reference when developing an incident response plan for a federated system is the CIC Federated Security Incident Response Policy, which has been adopted as an InCommon recommended practice:

- http://www.cic.net/Libraries/Technology/Federated_Security_Incident_Response.sflb.ashx
- https://spaces.internet2.edu/x/8o6KAQ

Lastly, the technical report "An Analysis of the Benefits and Risks to LIGO When Participating in Identity Federations" provides an example risk assessment that can be a useful input when preparing incident response plans:

- https://dcc.ligo.org/public/0070/G1100964/002/LIGOIdentityFederationRiskAnalysis.pdf