Enabling efficient electronic collaboration between LIGO and other astronomy communities using federated identity and COmanage

Heather Flanagan[a], Marie Huynh[b], Ken Klingenstein[a], Scott Koranda[b], and Benjamin Oshrin[a]

[a]Internet2, 1000 Oakbrook Drive Suite 300, Ann Arbor, MI 48104, United States;
[b]University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, United States

ABSTRACT

Identity federations throughout the world including InCommon in the United States, SURFnet in the Netherlands, DFN-AAI in Germany, GakuNin in Japan, and the UK Access Management Federation for Education and Research have made federated identities available for a large number of astronomers, astrophysicists, and other researchers. The LIGO project has recently joined the InCommon federation and is beginning the process to both consume federated identities from outside of LIGO and to make the LIGO identities issued to collaboration members available for consumption by other research communities.

Consuming federated identity, however, is only the beginning. Realizing the promise of multi-messenger astronomy requires efficient collaboration among individuals from multiple communities. Efficient collaboration begins with federated identity but also requires robust collaboration management platforms providing consistent, scalable identity and access control information to collaboration applications including wikis, calendars, mailing lists and science portals. LIGO, together with collaborators from Internet2, is building the COmanage suite of tools for Collaborative Organization Management. Using COmanage and leveraging federated identities we plan to streamline electronic collaboration between LIGO and other astronomy projects so that scientists spend less time managing accounts and access control and more time doing science.

Keywords: identity, federation, collaboration management

1. LIGO AND THE LSC

The California Institute of Technology (Caltech) and the Massachusetts Institute of Technology (MIT) constructed and operate the Laser Interferometer Gravitational Wave Observatory (LIGO)[1] under a cooperative agreement with the United States National Science Foundation (NSF). LIGO operates two interferometers in the United States, one in Hanford, Washington and the second in Livingston, Louisiana. The combination of the Caltech and MIT LIGO staff with the LHO and LLO staff comprise the LIGO Laboratory with more than 200 current members.

Research groups at universities and other institutions sign memorandum of understanding (MOU) with the LIGO Laboratory to join the LIGO Scientific Collaboration (LSC).[2] The LSC carries out the science mission of LIGO and is organized around three general areas of research: analysis of LIGO and other interferometer data searching for gravitational waves from astrophysical sources, detector operations and characterization, and development of future large scale gravitational wave detectors. The LSC was founded in 1997 and today includes more than 800 scientists from dozens of institutions and 13 countries worldwide.[3] Note that many, but not all, members of the LIGO Laboratory are members of the LSC.

The GEO project[4] is a German-British collaboration that built and operates the GEO600 interferometer in Hannover, Germany. Through an MOU signed with the LIGO Laboratory and the LSC GEO itself is a part of the LSC and all GEO members are members of the LSC. Likewise the Australian Consortium for Interferometric Gravitational Astronomy (ACIGA)[5] is wholly part of the LSC. More recently the Korean Gravitational-Wave Group (KGWG)[6] and the Indian Initiative in Gravitational-wave Observations (IndIGO)[7] consortiums have joined the LSC.

Corresponding author: Scott Koranda, scott.koranda@ligo.org

2. LIGO IDENTITY MANAGEMENT PROJECT

The LIGO Laboratory and the LSC (hereafter "LIGO") have recently grown in size to more than 1000 members from across four continents. Building and operating LIGO and analyzing interferometer data in such a collaboration requires a large number of working groups and committees, each with its own needs for electronic collaborative tools. The larger working groups or committees often require email lists, wikis or electronic notebooks, software version control repositories, and tools for audio and video conferencing. LIGO also operates collaboration-wide services used across many different groups such as the Document Control Center (DCC)[8] and the LIGO Data Grid (LDG)[9] used for computationally intensive data analysis. Other services include data streaming, detector characterization portals, and various metadata services necessary for data analysis.

Starting in 2007 the LIGO Identity Management project began to design, develop, deploy, operate, and support an IdM infrastructure capable of supporting the needs of the LIGO community and easing the burdens of working collaboratively across four continents. The primary goal of the LIGO IdM project was at that time to enable each LIGO member to use a single LIGO electronic identity to efficiently access and consume all electronic resources, services, and tools necessary to carry out the LIGO science mission. Later as the IdM project matured and the need to support efficient collaboration with external partners grew the project began to plan for a future where LIGO resources efficiently consume federated identities both for internal and external collaborators.

2.1 Enrollment

After formally joining LIGO either by signing an MOU with the LSC or becoming a member of the LIGO Laboratory each person enrolls electronically using the MyLIGO web portal.[10]

The current MyLIGO tool is a custom set of PHP code developed by and for LIGO. New members enroll using different workflows depending on how each is joining the collaboration. Enrollment flows for joining the LIGO Laboratory are more complicated than those for the LSC since the laboratory needs more closely resemble those of a classic structured enterprise rather than a distributed or virtual organization (VO) like the LSC.

Often much of the information and attributes about LIGO members collected during the enrollment flows is already known by the user's "home" institution or organization. Since the MyLIGO application in production today is not able to consume federated identity and asserted attributes, however, the current enrollment flows must gather all the necessary information manually.

After a successful enrollment the MyLIGO web portal stores information and collected attributes such as given name, family name, address, and telephone number in a MySQL relational database. Using given, family, and sometimes middle name(s) the portal creates for each user a unique identifier to serve as that person's single LIGO electronic identity. The IdM project has branded that LIGO electronic identity as the "albert.einstein" or "@LIGO.ORG" identity. The branding has played a critical role in the adoption and uptake of the IdM infrastructure by collaboration members.

2.1.1 Example LIGO Laboratory enrollment flow

The LIGO Laboratory employs scientific staff, engineers, administrative staff, graduate students, and undergraduate students. The electronic enrollment or onboarding process flows differently for each employee type and requires specific information input, approvals, notifications, and provisioning.

To highlight requirements that need to be considered when designing any system to manage the complex enrollment process for a large scientific VO like LIGO, consider a particular enrollment scenario–the manager of a technical computing group at LIGO-Caltech hiring a new senior staff person. After receiving approval from LIGO Laboratory management and Caltech to hire, the manager logs into the MyLIGO portal. The portal recognizes the manager's role as a hiring manager and presents the manager with a choice of available tasks including adding a new LIGO Laboratory person. After selecting that task a web form prompts for the following fields:

Given, middle, family name and suffix: These are standard name attributes for VOs and since LIGO is an international project the fields must support unicode characters. All of the challenges familiar to web form designers for entering and managing name identifiers in support of multiple languages and cultures are expected for LIGO.[11]

The LIGO electronic identity is derived from a person's name identifiers so the form should dynamically check inputs, compute possible LIGO identities, and offer the ability to choose from a set of possible matches to existing identifiers to support the use case of a previous employee returning to LIGO.

Previous LIGO identity: As noted above the form should dynamically compute and present as choices possible matches to known LIGO identifiers based on name identifier inputs. This field is included in the form so that when a previous LIGO identity is known for a returning LIGO employee but the system does not offer it as a dynamic choice, perhaps because the algorithm used to compute a LIGO identifier from a set of name identifiers has changed, the hiring manager can still input the known LIGO identity.

Email: The infrastructure provisions an email address of the form `albert.einstein@ligo.org` for each LIGO member but emails sent to that address must be forwarded to another address. This field collects that forwarding address.

Title and affiliation: The new employee's title and one of faculty, student, staff, member, affiliate, or employee, used to populate the `eduPersonAffiliation` entry in the LIGO LDAP directory.

Organizational unit: One of Caltech, MIT, LHO, or LLO to indicate into which LIGO unit the person is being hired. The default value is indicated by the role, identity, and attributes for the hiring manager but exceptions occur regularly and it is necessary that the hiring manager be able to override the default.

Sponsor: Each new member of LIGO must be sponsored. Eligible sponsors are determined by LIGO management and should be presented as a list from which the hiring manager may choose.

Valid from and through: Effective dates for LIGO membership must be recorded and dates both in the past and in the present need to be supported by the infrastructure.

LSC member: Some but not all LIGO Laboratory members are members of the LSC. This field allows the hiring manager to indicate his or her understanding of whether the new laboratory member should be a LSC member, but the final determination and approval rests with the office of the LIGO Director.

Percent FTE, research, LIGO research: If the new employee is to be a LSC member then the percent full-time employee (FTE), research, and LIGO research must be recorded. See below the discussion of an LSC enrollment for a more detailed explanation of the LSC membership requirements.

LSC valid from and through: Effective dates for LSC membership must be recorded and it cannot be assumed the effective membership dates for the LIGO Laboratory and LSC membership are the same. Both dates in the past and in the future must be supported.

After the form is submitted and input validated any remaining reconciliation needed between existing organizational identities and the new name identifiers must be diagnosed and managed appropriately. Some reconciliation steps may only be resolvable by direct administrator intervention.

With reconciliation complete the new LIGO identity is created and the enrollment petition marked as pending. Notifications are sent next to various actors and observers. The details of whom should receive notifications, mostly by email, vary across the spectrum of onboarding flows as does the specific details of how people are to be notified–actors should receive an email with their address in the `To:` field while observers prefer to only be `Cc`'d. For this specific example an email notification is sent to the hiring manager submitting the form and the Deputy Director of the LIGO Laboratory who is responsible for directly approving most hires at the LIGO Laboratory at Caltech.

After being notified the Deputy Director logs into the portal and is given the opportunity to edit certain parts of the enrollment petition, most notably whether the new hire should be a LSC member and the corresponding FTE, research, and LIGO research time commitments. After making any necessary edits the Deputy Director is expected to approve or deny the enrollment petition leading to another round of enrollment-specific notifications. Approved petitions cause an email notification event to the enrollee that includes a URL to be used to complete the enrollment flow.

Various provisioning flows must be initiated next to prepare the infrastructure to effectively support the new LIGO identity. As detailed below LIGO electronic identifiers are also Kerberos principals and a primary provisioning task is the creation of the new principal in the Kerberos key distribution center (KDC). Other provisioning tasks are both enrollment flow and organizational specific with different requirements for the four different LIGO Laboratory sites.

After being notified of an approved petition the enrollee browses to the prescribed URL to set an initial password for his or her LIGO identity and complete a web form to submit the following information:

Address and phone: Work address and phone numbers.

Email: With an approved enrollment petition the enrollee may choose at this time to change his or her email forwarding address.

Additional email: Other email addresses from which the user may wish to send mail and have it accepted by the LIGO email list servers.

Author name: Some enrollees, especially those who are LSC members, are expected to be authors on collaboration papers and will need to specify a name identifier to be used on published papers.

Preferred name: Other name identifiers the enrollee may prefer.

Demographic self report: The NSF requires LIGO to collect member demographic information. The enrollee should be prompted at this time to enter demographic information or actively elect not to submit demographic information (opt-out).

After the enrollee completes and submits the form and it is validated the infrastructure marks the identity as active.

This is just one representative enrollment flow for the LIGO Laboratory. Any platform or infrastructure intended to manage onboarding for the LIGO Laboratory must be flexible and extensible enough to accommodate many different flows, each with its own requirements and details.

The flow above highlights the combination of standard onboarding requirements common to many organizations such as collecting name identifiers, addresses, and telephone with LIGO specific details including LSC membership details and specific notification and provisioning flows. We expect a number of onboarding flow details to be shared across various astronomy and other scientific VOs but also recognize each VO may have unique enrollment flow requirements.

2.1.2 Example LSC enrollment flow

After a university research group or other organization signs an MOU with the LIGO Laboratory to join the LSC its members are eligible, with the principal investigator's (PI) approval to join the LSC and obtain a LIGO electronic identity. Today the PI's enrollment must be bootstrapped by technical staff. We expect in the future that the infrastructure will enable the LSC Spokesperson to bootstrap the enrollment of the PI immediately after the MOU is completed.

Today new LSC members from a university research group or other organization initiate the enrollment flow themselves rather than having the PI begin the flow. Unauthenticated users browse to the MyLIGO portal and select the LSC enrollment flow and then fill out and submit a web form with the following information (this description of the enrollment flow includes both current functionality and planned or required and missing functionality):

Given, middle, family name and suffix: The same as for the LIGO Laboratory flow detailed above.

Previous LIGO identity: The same as for the LIGO Laboratory flow detailed above. As students and postdoctoral researches move from institution to institution it is particularly common for LSC members to have an existing LIGO identity. Any platform for managing LSC onboarding and offboarding flows must efficiently support these use cases when members move from one institution to another within the collaboration, sometimes with extended absence periods during which few LIGO electronic privileges should be available.

Email: The same as for the LIGO Laboratory flow detailed above.

Address and phone: Work address and phone numbers.

Title and affiliation: The same as for the LIGO Laboratory flow detailed above.

Organizational unit: The LSC member institution or organization to which the enrollee belongs and is supporting his or her membership in the LSC.

After the form is submitted and input validated any remaining reconciliation needed between existing organizational identities and the new name identifiers again must be diagnosed and managed appropriately.

With reconciliation complete the new LIGO identity is created and the enrollment petition marked as pending. The PI and any other delegates are notified that a petition is pending and requires action.

Next the PI or delegate logs into the portal and is given the opportunity to edit certain parts of the enrollment petition. Most importantly the PI must set the three attributes detailing the research and time commitments to the LSC for the enrollee:

1. FTE%: The fraction of time a member spends as part of the group. This is less than 100% for multi-homed members, and 100% for members who belong to the LSC only through that group.
2. Research%: The fraction of time each group member has available for research (e.g., faculty who teach will have less than 100% of their time available for research).
3. LSC%: The percentage fraction of *available research time* devoted to LSC-related research and service.

The LSC bylaws use these values as input, along with effective membership dates, into a formula to determine author eligibility for collaboration papers.

After making any necessary edits and setting the three attributes above the PI approves or denies the enrollment petition. Approved petitions cause an email notification event to the enrollee that includes a URL to be used to complete the enrollment flow.

As with the LIGO Laboratory flow various provisioning flows must be initiated next to prepare the infrastructure to effectively support the new LIGO identity. Enrollment flows for LSC members outside the LIGO Laboratory generally require less provisioning.

After being notified of an approved petition the enrollee browses to the prescribed URL to set an initial password for his or her LIGO identity. Demographic information may be requested if the LSC group or institution is within the United State. A notification of all approved petitions for LSC membership is sent to the LSC Spokesperson.

That some LSC member organizations are themselves geographically distributed federations, such as GEO and ACIGA, may lead to more complicated onboarding and offboarding flows in the future. So far both GEO and ACIGA leadership have elected to manage enrollments with ad-hoc delegation arrangements. As the LSC grows in size and complexity those types of arrangements may no longer scale. This is especially true with the recent addition of KGWG and IndIGO to the LSC. Other LSC flows currently include enrollment for administrative or support staff needing LIGO identities to carry out delegated tasks but for whom authorship on collaboration papers is not a consideration. We have observed that as the LSC grows the onboarding and offboarding needs evolve and remain fluid with use cases and requirements often not prescribed until after new LSC members urgently need access to LIGO resources.

2.2 Identity

LIGO electronic identities take the form `given.family@LIGO.ORG` and uniquely identify a single LIGO member. Identifiers are not reused. At the time the MyLIGO portal creates the LIGO identity or identifier it provisions that identity into the LIGO master Kerberos Key Distribution Center (KDC). Each LIGO identity is simultaneously a Kerberos principal in the `LIGO.ORG` realm. Associated with each identifier is a password or pass phrase. Users may reset or change their password using the MyLIGO portal. The portal uses the administrative interface to the master KDC to set the password for the Kerberos principal. The KDC is the only password store used for all LIGO identities and at no time are passwords stored in plain text or encrypted and stored in any other way besides the KDC.

2.3 LDAP

The MyLIGO portal provisions an entry for each LIGO member into the LIGO master LDAP server. LIGO uses the OpenLDAP 2.4.x slapd server. Each LDAP record is an instance of the standard organizationalPerson, inetOrgPerson, and eduPerson object classes and each includes standard attributes such as cn, eduPersonAffiliation, givenName, locality, mail, mailAlternateAddress, mailForwardingAddress, postalAddress, sn, and telephoneNumber. Users may manage certain attributes such as mailForwardingAddress and postalAddress using the MyLIGO portal with changes provisioned into the MySQL relational database and LDAP server as necessary.

2.4 Group management

Principal Investigators (PIs) for LSC MOU groups and LIGO Laboratory managers also use the MyLIGO portal for simple group management tasks such as removing (de-enrolling) members from the group and for managing representation on the LSC Council. The MyLIGO PHP code uses web services calls to drive LIGO's Grouper deployment, the data store and foundation for the majority of LIGO's group management.

LIGO leverages Grouper[12] from Internet2 for the majority of its group management needs. Only IdM project members directly use the Grouper administrative interface for group management. Simple LIGO-specific interfaces, including the MyLIGO group management front end, have been built to enable users to manage certain group memberships. For example many of the LIGO email lists are "opt-in" and managed by joining a particular group. The majority of groups and group memberships within LIGO managed by Grouper are provisioned or reflected into the LIGO LDAP master server using the Grouper `ldappc-ng` (renamed the Provisioning Service Provider or `PSP`) tool.

2.5 Authentication, authorization, and services

All authentication for services and tools supported by the LIGO IdM project use or will use Kerberos and the LIGO electronic identities (Kerberos principals) for authentication. No other credential store is used for authentication. To enable a robust authentication infrastructure available to a widely distributed set of services and resources the LIGO master KDC is replicated to a number of slave KDCs throughout the world.

To enable single sign-on for web services and tools LIGO has deployed the Shibboleth[13] Identity Provider (IdP) and a SAML2 based infrastructure. The production IdP currently delegates authentication to the Apache `mod_auth_kerb` module so that users logins and passwords are tested against the LIGO KDC. A future enhancement will use a dedicated Kerberos JAAS-based login handler more tightly integrated with the IdP to provide a more customized user experience and support extra functionality such as forced re-authentication for high risk services. As part of the SAML2 identity assertion the IdP queries the LIGO LDAP server network for user attributes including group memberships and asserts those attributes for consumption by services.

LIGO has deployed over 50 instances of the Shibboleth Service Provider (SP) to consume identity and attribute assertions from the IdP and manage access to web content like electronic notebooks, wikis, and data analysis results. Working together with the IdP the SPs support a single sign-on experience across the majority of LIGO web services. Using attribute assertions from the IdP the SPs manage authorization to services. Most authorization decisions are based on group memberships as asserted by the IdP, which retrieved them from LDAP where they had been provisioned by the Grouper suite of tools.

LIGO Data Grid users currently authenticate to LDG resources including Linux clusters used for data analysis using RFC 3820 proxy certificates (derived from X.509 certificates) and tools enhanced to support them using the Globus Grid Security Infrastructure (GSI).[14] Today LIGO users in the United States request and receive X.509 certificates signed by the DOEGrids[15] certificate authority (CA) while users from other countries rely on regional or national CAs that are members of the International Grid Trust Federation (IGTF).[16] Each user is responsible for managing her own certificate and private key and the encryption of the private key is neither managed by the LIGO IdM infrastructure nor related in any way to her LIGO identity (Kerberos principal).

Soon, however, LIGO users will begin to retrieve short-lived X.509 certificates from the CILogon[17] service after authenticating via the LIGO IdP using their LIGO identities (Kerberos principals). The short-lived X.509 certificates and RFC 3820 proxy certificates based on them can be used with the current set of GSI-enabled tools. This change will reduce the burden of users having to manage their own certificate and private key and directly tie LDG authentication to the single LIGO identity.

Authorization for access to LDG resources is currently managed via static access control lists or grid-mapfiles. Each static grid-mapfile lists the X.509 certificate subjects authorized to access the resource, and when necessary include a mapping to a local account needed by the service. After transitioning to using short-lived certificates issued by the CILogon service where the certificate subject name is directly tied to the LIGO identity or Kerberos principal authorization will be managed using grid-mapfiles derived programmatically from LDAP. The group of users that should be authorized to access a particular LDG service will be managed using Grouper with the group membership being provisioned into LDAP. The grid-mapfile generation tool will simply query LDAP with the name of a group(s) that should be authorized to obtain memberships and receive a list of X.509 subject names derived from the corresponding LIGO identities.

Shell and terminal access to general computing resources at the LHO site, as well as authenticated access to email services like IMAP, POP, and SMTP is managed using Kerberos for authentication and authorization against LDAP groups. A particularly elegant design choice made by the IdM project architect at LHO is to use a local Kerberos realm just for the LHO site and enable cross realm trust so that LHO members may seamlessly use their @LIGO.ORG identities while preserving flexibility for the local LHO infrastructure. At this time the LDAP groups used for authorization are not managed using Grouper but that enhancement is planned for the near future. The other LIGO Laboratory sites are at various stages of transitioning their infrastructure to leverage a similar design.

LIGO makes heavy use of electronic mailing lists. Until recently the mailman tool[18] managed most lists but soon after the initial deployment of the LIGO LDAP network the collaboration began switching most email lists to be managed using Sympa.[19] In addition to a number of other useful features Sympa supports list membership queries through LDAP and integrates easily with the Shibboleth SSO infrastructure. With sophisticated group management including opt-in/out and composition provided by Grouper and reflected into LDAP, Sympa easily consumes email list information about subscribers, owners, and moderators that is more easily and naturally managed with Grouper.

2.6 Example inter-LIGO collaboration management scenarios

The LIGO IdM project provides a solid infrastructure foundation that has already substantially eased electronic collaboration for LIGO members. Two inter-LIGO collaboration management scenarios, however, demonstrate how the existing infrastructure still lacks the necessary management and provisioning tools to fully enable collaboration among LIGO scientists.

2.6.1 GstLAL

Recently a group of LIGO postdoctoral researchers and staff decided to explore a new approach to creating and managing LIGO data analysis workflows. The group chose as their development and analysis framework the open source multimedia framework GStreamer.[20] The project aims to build a LIGO data analysis workflow environment built using GStreamer and the LIGO Algorithm Library (LAL).[21] The project is named GstLAL.

The GstLAL team self organized and immediately needed a set of collaborative tools:

email list: Anxious to make progress the group turned to a local administrator and directly asked for an email list managed by the older mailman tool. More recently the group has requested the LIGO IdM project to provision a new email list using the Sympa, LDAP, and Grouper infrastructure. The provisioning requires multiple groups to be set up in Grouper for subscribers, moderators, and owners of the email list as well as the associated composite groups for managing delegation and other privileges. The structure and names of the groups, however, follow a known pattern and the process could be automated. At this time only an IdM project administrator can perform this work.

After the groups are created the population of the memberships can be easily delegated to the GstLAL team. The required Sympa configuration, however, cannot be delegated and must also be performed by an IdM project administrator at this time. Again the configuration follows a known pattern and the process could be automated.

Since the structure of the required Grouper groups and the Sympa configuration both follow known patterns that only depend on the name of the project and other simple metadata the infrastructure should have allowed the team to create and manage its own email list without needing to involve IdM project administrators.

code repository: The GstLAL team prefers to use Git for software version control. The team requires a centralized repository with most members having read access and only a few having commit privileges. IdM project administrators created the central repository and configured the necessary privileges. The structure and details of the repository again follow known patterns that depend on only a few details about the GstLAL project. The infrastructure should have allowed the team leaders to provision the repository and configure the necessary privileges without direct intervention from an administrator.

The team inherently understands this and demonstrated it when they asked "Why can't this just work like GitHub?"

wiki: So far the GstLAL team has leveraged wiki spaces from other LIGO projects rather than requesting its own wiki space in part because they understand that at this time the work must be done by already burdened IdM project administrators. The team would, however, prefer to have its own wiki space that supports anonymous read access for some pages but that includes areas requiring authorization. All write access should require authorization. Since the team is making contributions to the GStreamer upstream codebase it is conceivable that the infrastructure needs to support in the future federated authorized access from non-LIGO users. These deployment and provisioning patterns for federated wiki access are well known and understood and the infrastructure should support automated processes to enable the team to deploy the spaces without administrator intervention.

bug tracking: The GstLAL team would like to leverage the same bug tracking tool (Redmine[22]) that a number of other LIGO projects use. Today, however, provisioning of a new project within Redmine and the configuration of privileges requires administrator intervention. The infrastructure should support provisioning for new Redmine projects following the well established project patterns known to LIGO.

video conferencing: Like other LIGO groups the team relies on EVO[23] for videoconferencing. Reservations, accounts, authorization, and other details again follow well established patterns known to LIGO. The IdM infrastructure should support teams being able to provision their own meeting spaces on demand.

The GstLAL team is an energetic self-organized group of scientists within LIGO fully capable of provisioning and consuming collaborative tools without administrator intervention if the infrastructure supported it.

2.6.2 LigoDV-web

For a number of years LIGO instrument scientists have used the LIGO Data Viewer (LigoDV),[24] a thick client desktop tool, for detector characterization work. Recently a small team of software engineers with input from the instrument scientists has begun building a web browser version of the thick client. The project is known as LigoDV-web.

Many of the collaboration tools needed by the GstLAL project are also needed by LigoDV-web: email lists, code repository, wiki, and bug tracking. In addition the LigoDV-web team requires the following collaborative tools:

request tracker: Since LigoDV-web is used by many people across LIGO the project team needs to manage help requests from users. The team is leveraging the Request Tracker (RT)[25] deployment used by other projects within LIGO. The RT queue structure and email integration for the LigoDV-web project follow known patterns and the infrastructure should have supported the team to provision the queue without administrator intervention.

privilege management: Different capabilities of the LigoDV-web service require users to have specific privileges. As with most other LIGO tools and services those privileges are managed using the Grouper deployment. The LIGO IdM project infrastructure should have automatically made a stem or namespace in Grouper available with appropriate delegation of privileges to the LigoDV-web team at the beginning of the project so that the team could configure and manage the necessary privileges for its own project without needing administrator intervention.

delegation management: Because the LigoDV-web service is primarily concerned with detector characterization it is natural for the service to consume information from other LIGO web services as well as be consumable. LigoDV-web plans to leverage the existing SAML2 delegation capabilities of the LIGO Shibboleth infrastructure along with other LIGO web services. The operators for these services should be able to appropriately configure the necessary delegation within the LIGO IdM project following established patterns without needing administrator intervention.

The LigoDV-web project needs represent a common and reoccurring theme within LIGO as more and more services move away from thick clients to the web browser. The LIGO IdM project infrastructure should enable the development and operations teams to quickly spin up these new services without needing extended dialogues between team members and administrators who will simply follow already established processes and procedures.

3. EXTERNAL COLLABORATORS

Scientists rely on analyzing data from a world wide network of gravitational wave detectors to confidently detect and locate astrophysical sources of gravitational waves. The French and Italian Virgo[26] collaboration is composed of more than 200 scientists mainly from the Centre National de la Recherche Scientifique (CNRS) and the Istituto Nazionale di Fisica Nucleare (INFN) laboratories and from the European Gravitational Observatory (EGO). The collaboration operates the Virgo gravitational wave antenna. The Kamioka Gravitational wave detector, a large-scale cryogenic gravitational wave telescope, is currently under construction in Japan by the KAGRA project.[27]

Realizing the full science potential of LIGO and other gravitational wave detectors requires collaboration with scientists from other astronomy and astrophysics projects ranging from the NASA Swift Gamma-Ray Burst Mission[28] to the IceCube South Pole Neutrino Observatory[29] and the Numerical INJection Analysis (NINJA)[30] Project bringing together numerical relativity and gravitational wave data analysis. To date the LSC has signed MOUs with more than a dozen different projects for collaborative work to fully explore and realize the science potential of the LIGO and GEO interferometer data.

4. FEDERATION WITH EXTERNAL COLLABORATORS

As detailed above, LIGO scientists need to efficiently collaborate with researchers from a number of other projects. In the past enabling external collaborators access to LIGO resources required provisioning a new LIGO identity for each user. As is often the case requiring “yet another login and password” places a heavy management burden on users, requires significant helpdesk resources from the project, slows collaboration, and discourages well managed and secure electronic identities.

Today federated identity management and tools that consume federated identities streamline collaboration and lower the burdens on both users and administrators. Users authenticate with an existing electronic identity, often provisioned by their "home" university, institution, or organization. After a successful authentication the home IdP asserts attributes, including privileges, to the various service providers (SPs) to which a user needs to be authorized. Users rely on the home institution IdP operations team to support the usual login and password management requirements, rather then having to navigate the requirements and processes for each individual service. The LIGO IdM project has chosen infrastructure technologies and tools that support federated identity management.

For web services LIGO has chosen to leverage its SAML2/Shibboleth infrastructure to enable federation. LIGO joined the InCommon[31] identity federation in the United States and is beginning to pursue federation with other SAML based identity federations in Europe, Japan, Australia, and Canada.

Unfortunately not all scientists with which LIGO researchers need to collaborate have access to secure and well managed federated identities. For this reason LIGO has deployed an "identity provider of last resort". The LIGO Guest infrastructure[32] provisions electronic identities of the form `given.family@GUEST.LIGO.ORG` for collaborators with no access to federated identities but who need to access LIGO web resources.

As the number of identity federations increases and interoperability between identity federations matures[33] it is expected that someday VOs like LIGO will no longer need to provision electronic identities since all the VO members bring with them trusted federated identities from their "home" campuses, institutions, or other organizations. A necessary requirement for VOs like LIGO to decommission their IdPs is that the federated identities be easily consumable not only in the web space but also by services and tools across the grid or cluster and shell spaces. Recent work[34,35] indicates that future may no longer be far off. Until that time LIGO will continue to operate both its `LIGO.ORG` and `GUEST.LIGO.ORG` IdPs.

Federation within the grid space and federated access to LDG resources occurs through LIGO's reliance on X.509 credentials issued by CAs that are members of the IGTF. Likewise LIGO scientists may use their current X.509 certificates issued by the DOEGrids CA or other regional CAs to access non-LIGO grid resources. Soon LIGO will transition away from the DOEGrids CA and rely on the CILogon[17] service.

Federated access for LIGO scientists to non-LIGO grid resources will continue after LIGO transitions to using the CILogon service since the CILogon CA initially targeted for use is recognized by many grids. Note, however, that the initial CILogon CA that LIGO will use is not IGTF accredited. Use of the CILogon CA that is accredited by IGTF will require LIGO to achieve the InCommon Silver accreditation for its IdP. LIGO plans to take that action and assert InCommon Silver in the future.

4.1 Example intra-LIGO collaboration management scenarios

The LIGO IdM project provides a solid foundation ready to support identity federation and remove the burden from users and administrators of managing multiple electronic identities when accessing LIGO services. Three intra-LIGO collaboration management scenarios, however, demonstrate how the existing infrastructure still lacks the necessary management and provisioning tools to fully enable collaboration between LIGO scientists and their external collaborators.

4.1.1 Joint LIGO and NASA Swift analysis

The LSC has signed an MOU with the Swift team to enable joint analysis of the data from the LIGO instruments and the Swift observatory. Swift is a "first-of-its-kind multi-wavelength observatory dedicated to the study of gama-ray burst (GRB) science.[28]" A small joint working group has been formed and continues to work on the project.

To facilitate collaboration the LIGO InCommon administrator has injected metadata for the Compact Binary Coalescence (CBC)[36] wiki into the InCommon metadata to enable federation of that SP. Access to the particular wiki pages necessary to support the working group was granted to a non-LIGO researcher from MIT who authenticated using his MIT identity and gained access to the resource. Later LIGO Guest identities were provisioned for three NASA scientists without access to federated identities and the wiki access control configured to enable access for those three scientists using the LIGO Guest credentials.

Even with the basic configuration needed for federation complete, however, the process to enable access for the four collaborators was tedious, time consuming, and prone to errors. Because there is no source of group information regarding the working group members that could easily be managed by the scientists the lead LIGO team member needed to contact the IdM project administrators. Currently the LIGO Grouper deployment only resolves LIGO identities and cannot consume federated identities or generate federated group memberships. For this reason the administrators chose to authorize access to the wiki pages using specific user identities rather than authorizing a group or using a privilege.

The wiki software configuration requires a valid wiki name in the access control list (ACL). The wiki name must be generated from attributes asserted by the IdP to the SP hosting the wiki. This particular wiki software is configured to generate the wiki name from the eduPersonPrincipalName (ePPN) attribute value for the user asserted by the IdP (one of MIT, LIGO, or LIGO Guest in this scenario), since ePPN is a common attribute asserted by many IdPs, especially those in the InCommon identity federation. Only the wiki and IdM project administrators know these details and were able to ascertain what ePPN value the IdPs would assert for each user and then edit the ACL appropriately. Scientists should not be burdened with managing these details and the IdM project infrastructure should have made this process flow without any intervention from the administrators.

4.1.2 LIGO KAGRA

Work is underway to directly federate the SP supporting the LIGO Document Control Center (DCC)[8] with the University of Tokyo IdP to enable federated access to the DCC for a number of KAGRA scientists in support of a LIGO-KAGRA project. We expect that federation to be completed this summer.

Even with the configuration necessary to support federation complete, however, the processes for managing access will be cumbersome. As noted above the LIGO Grouper deployment currently does not support federated groups and is not able to consume federated identities. The most likely short-term solution will be a group entry in the LIGO LDAP server with membership managed by hand by the IdM project administrators. Membership changes to the LIGO-KAGRA project will require emails from both the LIGO and KAGRA managers to the administrators and management will be tedious and error prone. The details of the electronic identities asserted by the University of Tokyo IdP will need to be directly communicated over email by the IdP operator to the LIGO IdM project staff.

4.1.3 NINJA

The LSC has signed a MOU with the Numerical INJection Analysis (NINJA) project.[30] The goal of the NINJA project “is to bring the numerical relativity and data analysis communities together to pursue projects of common interest in the areas of gravitational-wave detection, astrophysics and astronomy.” Many members of the NINJA project are also members of LIGO and have LIGO identities but a number of other NINJA members are not part of LIGO. At this time the NINJA project is not large enough and does not have enough staff to join an identity federation such as InCommon, but it is large enough (approximately 100 members) to experience many of the pain points around electronic identity and collaboration. To begin to address some of the issues NINJA project staff at Syracuse University have deployed a Shibboleth IdP and have issued “NINJA credentials” to users (this deployment pre-dates the deployment of the LIGO Guest infrastructure). Users from both the NINJA project and LIGO need to access material hosted by the project wiki.

Because there is no federated source of group membership or privilege assertions, however, management of access controls within the wiki is cumbersome and error prone. It is common for information to be posted in the wiki that only LIGO identities can see or that only NINJA identities can see when anybody from the joint NINJA-LIGO project should be able to see the information. Frustrated users and NINJA administrators have taken to giving LIGO users NINJA identities, thereby undoing the benefits of federated identity. Sometimes it is just easier to manage a new identity than to leverage federated identity when the infrastructure does not support the collaboration workflows needed.

5. COLLABORATION MANAGEMENT WITH COMANAGE

Managing access to resources for federated identities brings with it its own challenges, slows collaboration, and adds a burden to LIGO IdM project administrators. To ease that burden and further enhance collaboration LIGO is helping to develop COmanage.[37]

Collaboration management platforms (CMPs) and services complete the basic vision of federated identity by adding components to do effective and scalable access controls and permission management. Together attributes are able to be created, managed and transported to relying parties that can then directly make decisions about users accessing resources.

CMPs are intended to provide consistent, scalable identity and access control information to both collaboration applications (including wikis, mailing list services, bug tracking, code repositories) and domain applications (including data grids, shell-based services, science gateways and portals, etc.).

Such platforms are built by repackaging enterprise middleware, such as Shibboleth, Grouper, and LDAP for use in virtual organizations. At this point, CMPs are typically built by assembling a coherent service from a variety of separate middleware and application servers, but deployable virtual machine appliances and the resulting cloud-based services are expected once the field is better understood.

The Internet2 middleware activity has received a NSF OCI grant, beginning September 1, 2010, called "Building from Bedrock: Infrastructure Improvements for Collaboration and Science." The intent of "Bedrock" is to enhance and package enterprise tools for collaborative organization (CO) use and work intensively with several major collaborations, including LIGO, for deployment. Staff from LIGO and the iPlant Collaborative, a project to develop cyberinfrastructure and computational tools to solve grand challenges in plant science,[38] are co-investigators on the grant.

Bedrock has an active development effort underway, focusing on the identity management needs as well as the domestication and integration of applications for large and small VOs. Working with a variety of VOs, the work being done is smoothing the way towards more efficient collaboration and better science within the VO.

Participating VOs in the initial stages of Bedrock development include:

- LIGO
- iPlant[38]
- The Internet Society (ISOC)[39]
- The Earth Science Women's Network (ESWN)[40]
- Project Bamboo–a multi-institutional, interdisciplinary, and inter-organizational effort that brings together researchers in arts and humanities, computer scientists, information scientists, librarians, and campus information technologists.[41]

5.1 COmanage architecture

Figure 1 illustrates a high-level view of the current COmanage suite reference architecture. Further details are available at the COmanage web site. COmanage Registry combines group management with configurable and flexible onboarding and offboarding workflows to support the quick and easy spin up of collaborative organizations (COs) focusing on a common task or goal. Registry supports COs requiring identity provisioning and management (i.e. creating new logins and passwords), COs only consuming federated identity, or a combination of both.

COmanage platform administrators can configure a Registry deployment so that certain details of a workflow enrollment, such as which name identifiers to collect, are required for all COs using the platform. CO administrators using the platform may further configure and customize the enrollment workflows used for adding new CO members. Registry supports CO administrators adding CO-specific attributes that need to be collected during an enrollment flow, such as the LIGO-specific FTE%, Research%, and LIGO Research% attributes detailed above.

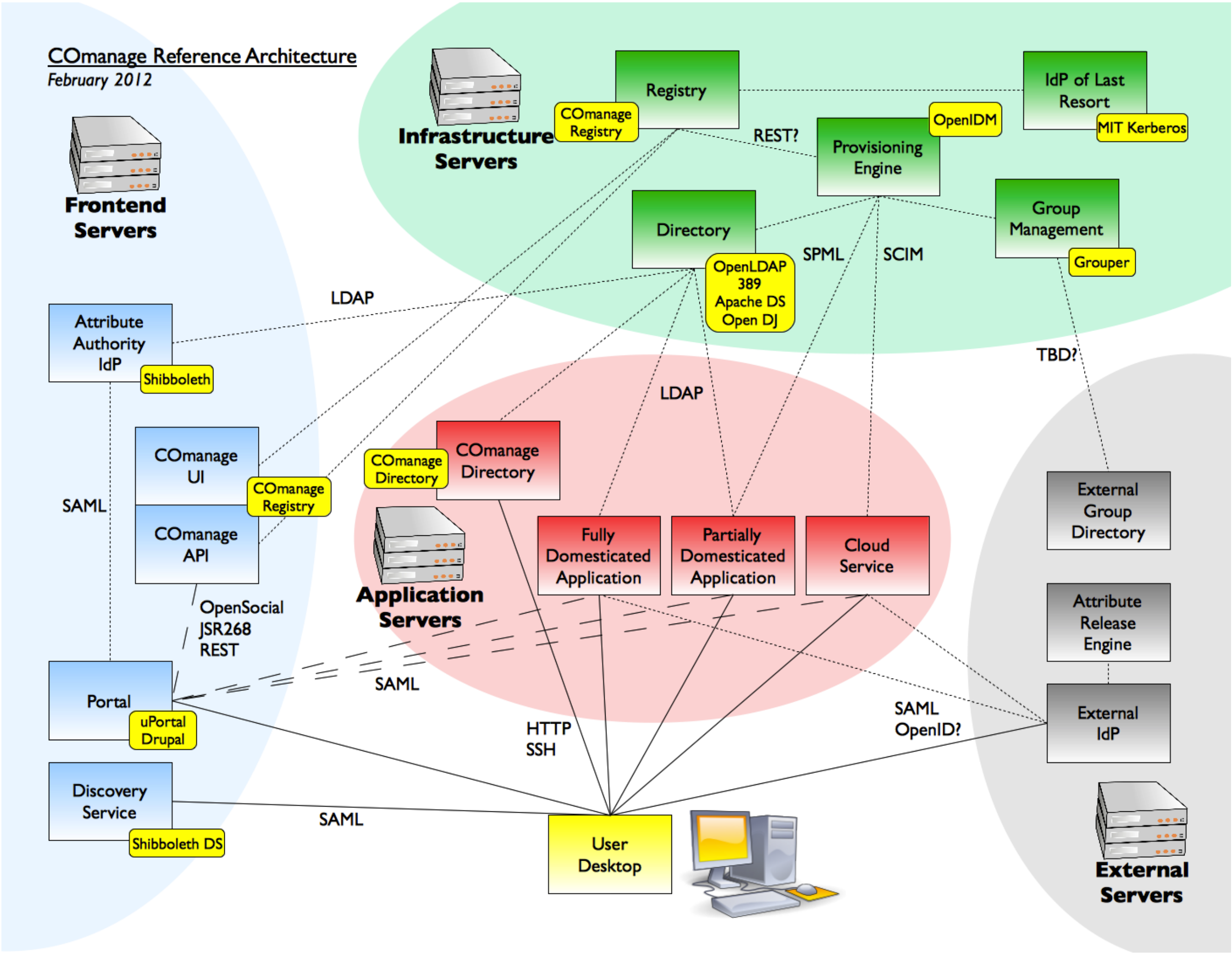

Figure 1. COmanage suite reference architecture.

CO membership is provisioned or reflected either directly into LDAP or via Grouper and then into LDAP, depending on how the platform is configured. The membership information is then available for consumption by IdPs for attribute assertions during authentication events or directly by applications.

Integration between COmanage Registry and a flexible and configurable provisioning engine enables CO and platform specific provisioning events to be orchestrated before, during, and after onboarding and offboarding processes. Registry itself supports a limited number of standard provisioning operations common across many COs, along with the ability to consume simple plugins that support provisioning related functionality, for example implementing specific password rules or mechanisms for COs that require identity provisioning and password management.

The COmanage Directory web application provides the standard directory or "white pages" functionality common to many organizations, both real and virtual or collaborative. Platform administrators configure which attributes and details are released about subjects for all COs and CO administrators may further configure CO-specific attribute release. When deployed with Registry the Directory supports "hot linking" of directory information with Registry data so that CO members may simply click to edit (when authorized).

Applications like wikis, code repositories, mail list servers, and calendars may consume CO membership, group, role, and privilege information directly from both Registry and Directory through REST interfaces, or indirectly through standard mechanisms like LDAP or SAML assertions. So-called "fully domesticated" applications are those considered ready to consume federated identity and group information from external sources and that do not require any substantial provisioning. Applications requiring specific provisioning before they can consume federated identity or group information are described as "partially domesticated".

5.2 CO and COU

COmanage Registry combines group management with configurable and flexible onboarding and offboarding workflows to support the quick and easy spin up of collaborative organizations (COs) focusing on a common task or goal. Put another way, a CO binds groups with onboarding and offboarding workflows, assurance levels, metadata management, and organizational processes.

Many COs, however, may require different enrollment flows for various working groups, departments, committees, or other sub-organizations. The organizational structure and different enrollment requirements for the LIGO Laboratory and the LSC clearly demonstrate this. COmanage therefore introduces the notion of a collaborative organizational unit or COU.

The COU is an optional construct to allow CO managers to define an organizational structure within a CO (e.g. a self-contained collection or department within a CO, or a collection of privileges within a CO). The workflow for enrolling people may have details specific to a COU.

If an organization has common goals and policies but yet within that understanding finds sub-groups with unique requirements and different paths to joining and participating in those groups, the organization is a CO that contains COUs. If the organization has one common set of policies that define how individuals are added or removed from the CO then the organization does not have COUs, even though there may be sub-groups for various other organizational reasons or for managing simple access control. The primary distinction is the extent to which more than one enrollment flow exists within the CO.

5.3 COmanage development status

Full details of the COmanage roadmap and release schedule are available at the COmanage web site.[37]

COmanage Registry and Directory are being developed in tandem by the COmanage development team. A catalogue of domesticated, partially domesticated, and not yet domesticated applications needed to support COs is being assembled. Active domestication of applications has not begun yet outside of specific work done by particular COs.

Work on Registry so far has focused on use case and requirements gathering, data model design and implementation, support for configurable enrollment workflows and CO-specific attributes, and user interface design and implementation. Integration with Grouper is underway and at this time work on the notification engine and provisioning will begin soon.

Work on Directory so far has focused primarily on simple “white pages” functionality and the link to Registry when deployed together.

5.4 COmanage and addressing LIGO scenarios

When the COmanage suite of tools and services is sufficiently mature LIGO will replace the current MyLIGO web portal with a deployment of COmanage Registry and Directory and use it to manage the internal collaboration needs of both the LIGO Laboratory and the LSC. LIGO Laboratory will be a CO with COUs for LIGO at Caltech, MIT, and the LHO and LLO sites. Different enrollment workflows for each COU will include customized notifications and when necessary provisioning. LSC membership will be indicated by a CO-specific attribute managed using Registry.

Another CO will represent LIGO collaborators with each LSC group that has signed an MOU represented by a COU. For example the Syracuse University LIGO group will be a COU within the LIGO collaborators CO. GEO, ACIGA, IndIGO, and KGWG will themselves be COUs with further COU structure. If necessary to support future organizational structure requiring specific enrollment workflows this nested COU structure may be altered.

This COmanage deployment will include direct integration with the LIGO Grouper deployment so that all CO and COU membership is reflected by Grouper and ultimately LDAP so that it may be readily consumed by a number of existing infrastructure components. Registry will enable provisioning and management of LIGO electronic identities for each member of the LIGO Laboratory and LSC.

Integration of Registry with applications like wikis, the Sympa email list server, and code repositories will enable LIGO scientists to quickly spin up new COUs and groups and use the COmanage infrastructure to quickly provision the necessary collaboration tools and services without direct intervention from the platform administrators. This integration will require a substantial effort to integrate and evolve a sophisticated provisioning engine as well as to fully domesticate the necessary applications.

A second COmanage deployment at a neutral web location outside of `ligo.org` will enable intra-LIGO COs to be quickly created and managed without regard to the politics of which VO appears to own and operate the infrastructure. One candidate web location is `gw-astronomy.org` but other locations will be considered during any joint work.

With this second deployment each CO is expected to be an effort across organizations such as LIGO-Swift, LIGO-NINJA, and LIGO-KAGRA. We expect over time that COs not involving LIGO will use the deployment and participate in its management and operations. This COmanage deployment will only consume federated identity and will not be configured to provision electronic identities or manage passwords. With its own Grouper and LDAP sub-deployments, the platform will fully support a federated identity ecosystem capable of participating in attribute and group assertions with trusted service providers.

5.5 Other collaboration management platform projects

While many of the core ideas for CMPs came from the United States, other countries, most notably the Netherlands, Switzerland and Norway have advanced the practice considerably. Several countries are planning a national level collaboration service. Specific projects or groups include:

- SURFConext, a “next generation collaboration infrastructure that creates new opportunities to collaborate online based on a combination of applications from different providers”, from SURFnet in the Netherlands.[42]

- SWITCHaai, an authentication and authorization infrastructure operated by SWITCH in Switzerland is exploring the CMP space using the concept of a Virtual Home Organization (VHO) and related ideas.[43]

- SWAMI is the Swedish Alliance for Middleware Infrastructure and is the Swedish University Computer Network’s (SUNET) virtual project organization.

- The Globus Online (GO) project from the Computation Institute at the University of Chicago and Argonne National Lab in the United States is building out a hosted collaboration management platform targeted at scientific organizations.[44]

Many of the people working in the CMP space benefit from sharing use cases and requirements and it is not unexpected that VOs may choose to mix and compose CMP solutions to suit their specific needs.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of the members of the LIGO Identity Management Project: Stuart Anderson, Warren Anderson, Melody Araya, Juan Barayoga, Phil Ehrens, Jonathan Hanks, Shannon Roddy, Abe Singer, Larry Wallace, and Roy Williams.

We also gratefully acknowledge the COmanage Technical Advisory Board: Tom Barton, Steven Carmody, Emily Eisbruch, Michael Gettes, Keith Hazelton, and Steve Oshansky.

We continue to benefit enormously from our SURFnet colleagues Niels van Dijk and Remco Poortinga.

Heather Flanagan, Marie Huynh, and Benn Oshrin are supported in part by NSF award 1032468. Scott Koranda is supported in part by NSF awards 0600953 and 1104371.

REFERENCES

[1] B. Abbott, et al., "The Laser Interferometer Gravitational-Wave Observatory," *Rep. Prog. Phys* **72**, 076901 (2009).

[2] http://www.ligo.org.

[3] http://roster.ligo.org.

[4] H Grote (for the LIGO Scientific Collaboration), "The status of GEO 600," *Class. Quantum Grav.* **25**(11), 114043–114052 (2008).

[5] http://www.anu.edu.au/Physics/ACIGA.

[6] http://kgwg.nims.re.kr.

[7] http://www.gw-indigo.org.

[8] http://dcc.ligo.org.

[9] https://www.lsc-group.phys.uwm.edu/lscdatagrid/.

[10] http://my.ligo.org.

[11] http://www.w3.org/International/questions/qa-personal-names.

[12] http://www.internet2.edu/grouper/.

[13] http://shibboleth.net/.

[14] http://www.globus.org/security/overview.html.

[15] http://www.doegrids.org/.

[16] http://www.igtf.net/.

[17] http://www.cilogon.org/.

[18] http://www.gnu.org/software/mailman/index.html.

[19] http://www.sympa.org/.

[20] http://gstreamer.freedesktop.org/.

[21] https://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html.

[22] http://www.redmine.org/.

[23] http://evo.caltech.edu.

[24] https://www.lsc-group.phys.uwm.edu/daswg/projects/ligodv.html.

[25] http://bestpractical.com/rt/.

[26] https://wwwcascina.virgo.infn.it.

[27] http://gwcenter.icrr.u-tokyo.ac.jp/en.

[28] http://www.nasa.gov/mission_pages/swift/main/index.html.

[29] http://icecube.wisc.edu/.

[30] https://www.ninja-project.org/doku.php.

[31] http://www.incommon.org/.

[32] https://guest.ligo.org/.

[33] https://refeds.org/.

[34] http://www.project-moonshot.org.

[35] http://tools.ietf.org/html/draft-cantor-ietf-kitten-saml-ec-01.

[36] https://www.lsc-group.phys.uwm.edu/ligovirgo/cbc/.

[37] https://spaces.internet2.edu/display/COmanage/Home.

[38] http://www.iplantcollaborative.org/.

[39] http://www.internetsociety.org/.

[40] http://www.eswnonline.org/.

[41] http://www.projectbamboo.org/.

[42] http://www.surfnet.nl/en/Thema/coin/Pages/default.aspx.

[43] http://www.switch.ch/aai/index.html.

[44] https://www.globusonline.org/.