Astronomical Data Analysis Software and Systems XXVI
ASP Conference Series, Vol. 521
Marco Molinaro, Keith Shortridge, and Fabio Pasian, eds.

Extending Support for Large Distributed Projects Through Interoperability

Séverin Gaudet,[1] Giuliano Taffoni,[2] Sara Bertocco,[2] Brian Major,[1] Patrick Dowler,[1] Marco Molinaro,[2] David Schade,[1] and Fabio Pasian[2]

[1]*Canadian Astronomy Data Centre NRC, Victoria, BC, Canada;*
Severin.Gaudet@nrc-cnrc.gc.ca

[2]*INAF-Osservatorio astronomico di Trieste, Trieste, Italy;*
taffoni@oats.inaf.it

Abstract. Many astronomy projects today are executed by distributed science teams with access to different computation and storage resources. As we move into the era of petabyte and exabyte datasets, it is recognized that moving the code to the data becomes necessary as the alternative becomes infeasible. The question becomes how can resource infrastructures support these large projects such that a team has integrated access to the different distributed resources available to a project. Examples of resources that could be integrated are files and directories, storage allocations, processing allocations, containers and virtual machine images, databases and tables, etc. A first step in this direction is the interoperability of authorization services.

The International Virtual Observatory Alliance (IVOA) has developed many standards to support access and interoperability of infrastructure such as Single-Sign On (SSO), Credential Delegation Protocol (CDP) and VOSpace. Both Canadian Advanced Network for Astronomical Research (CANFAR) operated by the Canadian Astronomy Data Centre) and INAF-Osservatorio Astronomico di Trieste (INAF-OAT) use these standards for provision of user storage to support projects. In the VOSpace implementation, users assign read-only and read/write permissions to groups that are defined in their respective home institution Group Management Services. In 2015, the EGI-Engage project in Europe partially funded an exploration of interoperability of authorization services in a joint project between the CANFAR and INAF-OAT. This has also led to the inclusion of this work in the Advanced European Network of E-infrastructures for Astronomy with the SKA (Aeneas) proposal. The joint CANFAR/ INAF-OAT project has added support to interoperate its VOSpace services by adding the capability of granting authorization to access a resource to groups defined in an external Group Management Service and to allow for the dynamic creation of internal user IDs that are associated with an external identify provider.

1. The need for interoperability

1.1. Supporting science teams

The need for interoperability begins with supporting science teams. A science team is a virtual organisation consisting of several to many individuals and is often international in scope. The team forms around a given multi-year project and may be dealing with a large dataset. The team is then faced with acquiring and building project infrastructure to support team activities which can include data management, data distribution and

data processing. This is challenging for a distributed team where team members may have access to local or regional resources that they may able to bring to the project but these resources do not interoperate, different silos essentially operating independently. The challenge is to provide integrated resources to allow teams to self-organize around their project goals and devote their available resources to the science tasks rather than to building and maintaining infrastructure.

1.2. Big data: moving code to data

As the ability to generate data increases, the community has been exploring ways to reduce the need to move large datasets to a users resources. In the past few years, the idea of moving code to processing resources co-located with the data storage is seen as the primary strategy to mitigate this data movement. This was the one of the foundational ideas behind the CANFAR system (Gaudet et al. 2009), where user accessible processing and storage was co-located with telescope data. This is also the concept adopted for processing in the EUCLID Science Ground Segment data (Poncet et al. 2019) and is one of the drivers in creating SKA regional centres. To enable moving code to data, an interoperable authorization framework could be part of the solution.

1.3. Resources are not simple

Up until recently, the authorized resource in an astronomy data centre was limited to meta-data and/or data files. But in the era of providing support for science teams and moving code to data, the resources required go beyond files and databases. They now involve storage allocations, processing allocations and shared virtual machines or containers. Authorization also applies to services, job listing and control, user information and group information. To create an effective work environment that supports teams or allows moving code to data, these resources need to be integrated and to interoperate.

2. Interoperability through authorization

The following section describes authorization based on a group management service (GMS) (Damian et al. 2012) and how it can be used to support interoperability.

2.1. Authorization based on Groups

- Groups
  - A group is a set of users
  - A user creates a group (owner)
  - The owner adds users and/or groups as members
  - The owner adds users and/or groups as administrators
  - A group can exist on its own. Its does not have authorization or privileges until a grant action associates it with a resource (see below)
- Granting action
  - A granting action associates a group to a resource
  - Any member of group X can query this row (operational in the CADC query service)

  - Any member of group Y can read this file (operational in the CANFAR user storage)
  - Any member of group Z can execute this virtual machine (operational in the CANFAR batch processing system)

- Granting read-only to public is also supported

3. An experiment

3.1. The project

This EGI-funded Data Commons project is to explore interoperability with other Authentication and Authorization infrastructures. The approach is to leverage an operational Group Management Service and IVOA services in CANFAR with a similar deployment at INAF-OAT such that at the IVOA services can interoperate. This is not a prototyping activity but rather deploying operational services at each site. This will enable:

- Identities and groups defined at INAF-OAT to be used in CANFAR
- Identities and groups defined in CANFAR to be used at INAF-OAT

3.2. Interoperating VOSpaces

The experiment is based on deploying two independent sets of services, one in CANFAR and one in INAF-OAT consisting of a VOSpace service (Graham et al. 2013), a group management service (GMS), a VO credential delegation service (CDP) (Plante et al. 2010) and supporting the VO single sign-on standard (Rixon et al. 2008). Details of the software stack and the experiment are described in (Bertocco et al. 2019). In summary, the following was enabled:

1. A user A with a VOSpace allocation in CANFAR logs in (authenticates) to CANFAR and then grants read permission on subdirectory S by setting a URI to a group X defined in the INAF-OAT GMS.
2. A member B of group X at INAF-OAT authenticates and retrieves a proxy certificate from the CDP service at INAF-OAT and issues an http get command with the proxy certificate to the CANFAR VOSpace service.
3. The CANFAR VOSpace service takes on one hand user B's identity and on the other hand the list of groups allowed to access the subdirectory and asks each of the GMSs defined by the URIs if user B is a member of the group. The INAF-OAT GMS responds positively so user B is authorized to read the file and the transfer is initiated.

4. Next steps

4.1. IVOA

The current IVOA mandate states:

- *Allow astronomers to interrogate multiple data centers in a seamless and transparent way*
- *Give data centers a standard framework for publishing and delivering services using their data.*

In the context of moving more services into data centres, the IVOA could consider extending their mandate by facilitating interoperability standards at the resource level:

- Give data centers a standard framework for integrating interoperable authorization
- Allow astronomers to interrogate or use multiple data centers in a seamless and transparent authorized way

4.2. Resource providers

For resource providers, be it telescopes, archives, data centres or processing centres, begin to interoperate authorizations:

- Integrate authorization across all resources being provided to users
- Allow external trusted group information providers
- Allow users to create and manage their own groups
- Allow users to grant authorization to their groups to their allocated resources

For the current project, the next step is to work with Compute Canada and EGI on interoperating at the resource provider level such that group management can be provided as part of regional research computing organisations and to interoperate as trusted group information providers. This work is also part of the AENEAS (Advanced European Network of E-infrastructures for Astronomy with the SKA), an exploratory project that includes an interoperability component.

References

Bertocco, S., et al. 2019, in ADASS XXVI, edited by M. Molinaro, K. Shortridge, & F. Pasian (San Francisco: ASP), vol. 521 of ASP Conf. Ser., 61

Damian, A., et al. 2012, in ADASS XXI, edited by P. Ballester, D. Egret, & N. P. F. Lorente, vol. 461 of ASP Conf. Ser., 311

Graham, M., et al. 2013, IVOA recommendation: VOSpace specification v2.0, IVOA Recommendation 29 March 2013. 1509.06049

Plante, R., et al. 2010, IVOA Credential Delegation Protocol Version 1.0, IVOA Recommendation 18 February 2010. 1110.0509

Poncet, M., Le Boulc'h, Q., & Holliman, M. 2019, in ADASS XXVI, edited by M. Molinaro, K. Shortridge, & F. Pasian (San Francisco: ASP), vol. 521 of ASP Conf. Ser., 588

Rixon, G., Graham, M., & Grid and Web Services Working Group 2008, IVOA Single-Sign-On Profile: Authentication Mechanisms Version 1.01, IVOA Recommendation 24 January 2008. 1110.0506