

Extending Support for Large Distributed Projects Through Interoperability

Séverin Gaudet and the project team
Canadian Astronomy Data Centre



Extending Support Through Interoperability

- Why?
- What?
- An experiment
- How?
- Next steps



Interoperability for distributed resources

- International team members can bring regional resources
- Big data: moving code to data
- Resources are not simple

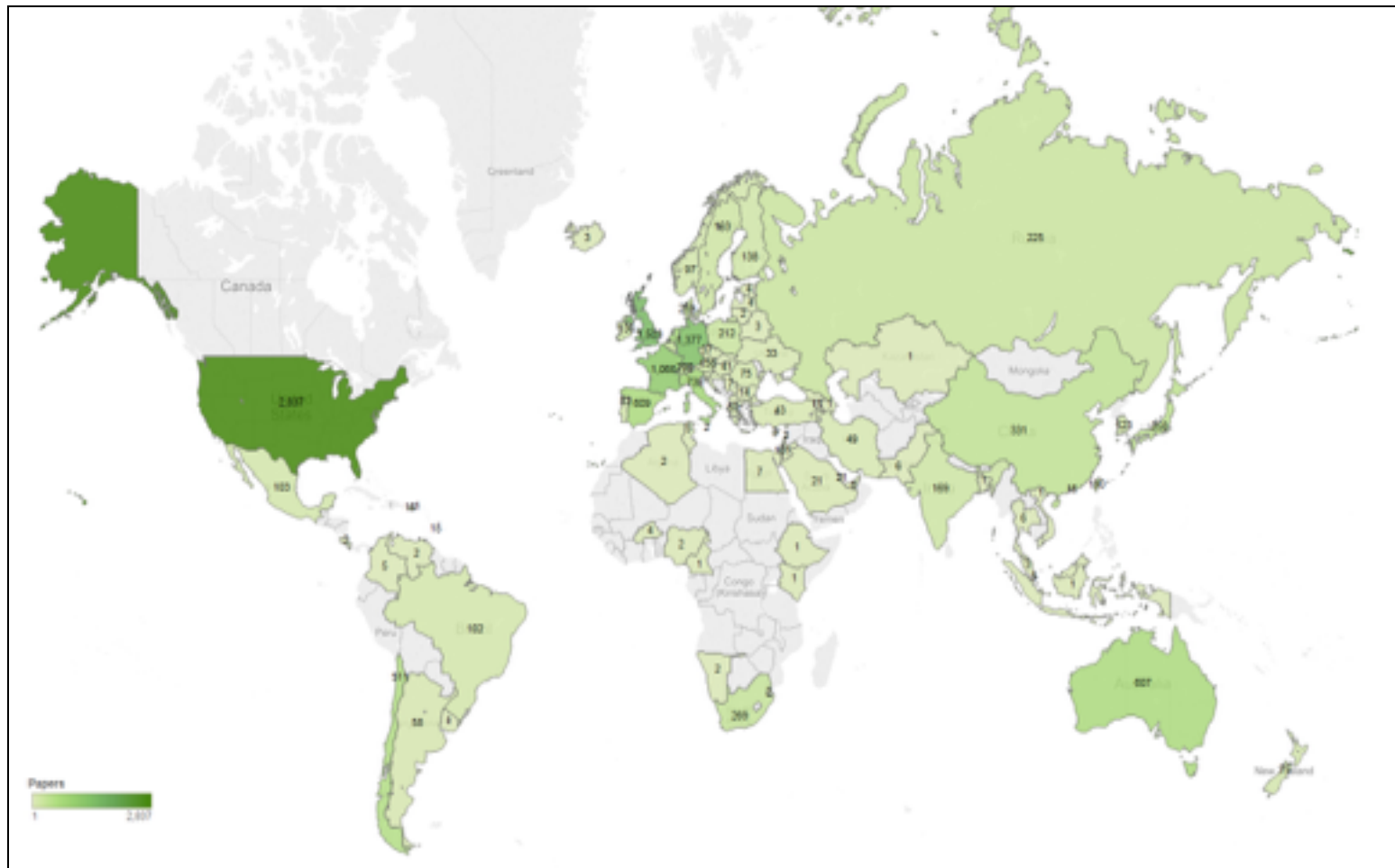


Interoperability for science teams

- Science teams are international virtual organisations
 - Forming around a given multi-year project
 - Handling large datasets
 - Faced with acquiring and building project infrastructure
- Require infrastructure
 - Larger datasets
 - Data management, data distribution, data processing
 - Challenging a team's ability to produce and maintain infrastructure
- May have access to national and regional infrastructure



Interoperability for science teams

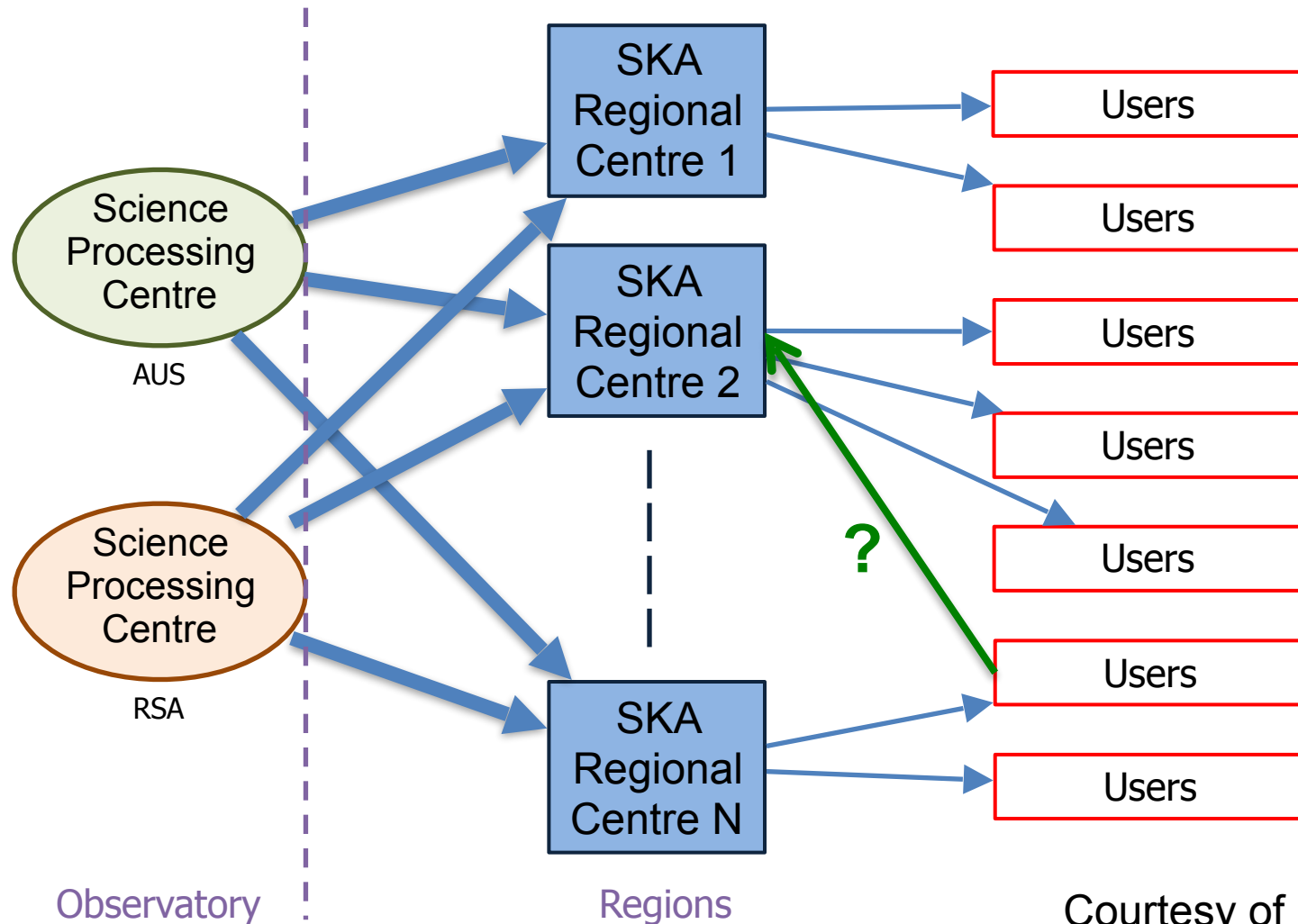


Co-authors by country for Canadian first author papers in 2014 – Crabtree

Moving code to data



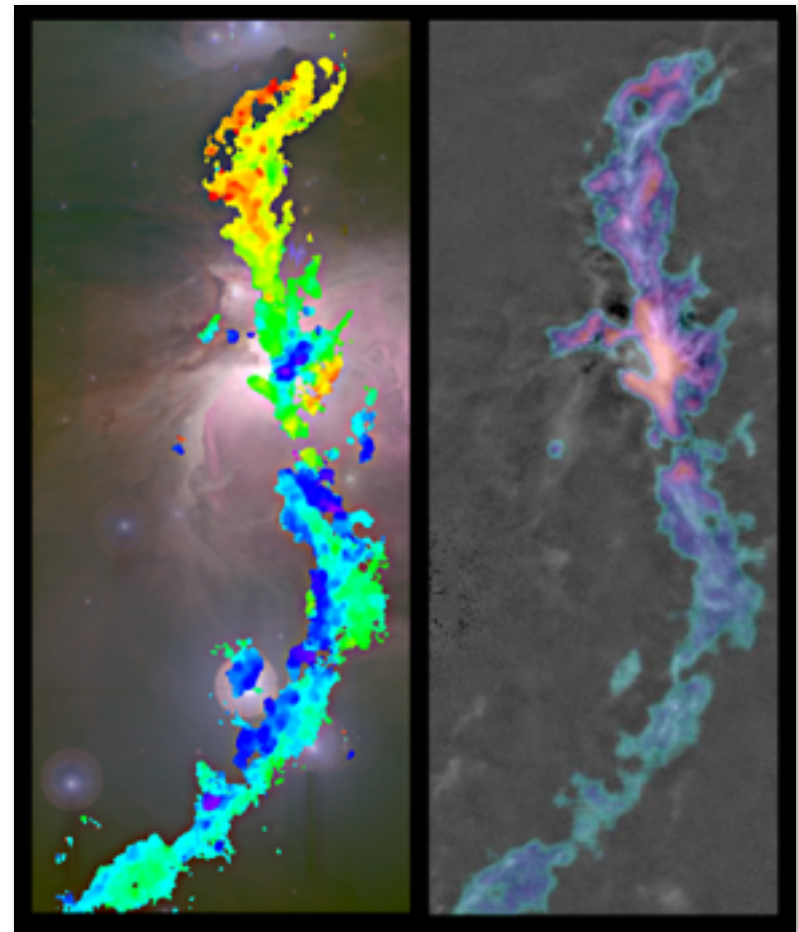
Moving code to data



Courtesy of G. Davis

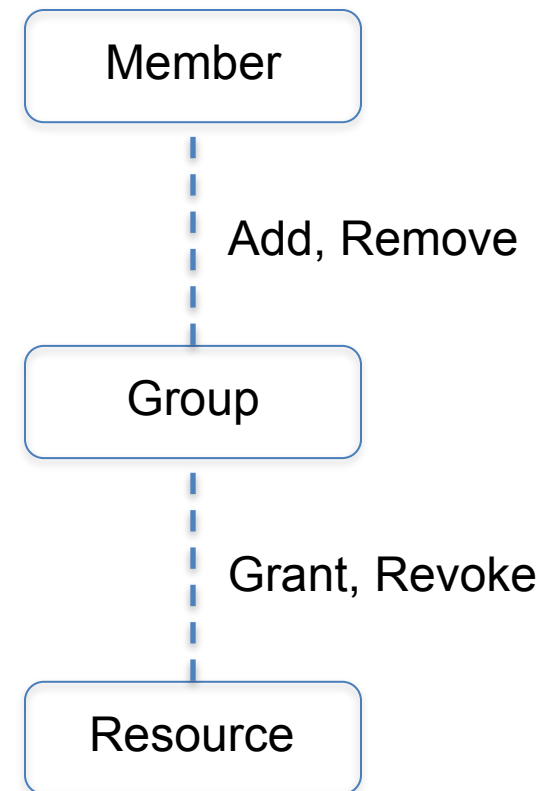
Interoperability for distributed resources

- Resources are no longer simple
- International teams can bring regional resources
- Big data -> moving code to data
- Resources
 - Data (files), databases (tables, rows)
 - Storage allocations, processing allocations, virtual machines, containers
 - Services, jobs, user info, group info
- Can these be integrated and interoperate?



Interoperability Through Authorization

- Authorization based on **Groups**
 - A sets 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
- **Granting** action associates a group to a resource
 - Any member of group X can query this row (TAP)
 - Any member of group Y can read this file (VOSpace)
 - Any member of group Z can execute this VM (CANFAR)
- Granting read-only to public also



The experiment

- EGI funded (Data Commons)
- Interoperability with other A&A infrastructures
 - At to IVOA level:
 - Osservatorio Astronomico di Trieste and CADC
 - At the regional infrastructure level:
 - EGI.eu and Compute Canada
- Identities and groups defined at INAF-OAT to be used in CANFAR
- Identities and groups defined in CANFAR to be used at INAF-OAT



The Starting Point – CANFAR and...

- A cloud ecosystem:
 - Data intensive astronomy
 - Collaboration platform
- Users request allocations of resources
- A collection of integrated services
 - Storage
 - Interactive VMs
 - Persistent VMs
 - Batch processing with VMs
- Built with IVOA services
- Using research cloud resources



Operational since 2011!



University
of Victoria



University of
British Columbia

canarie



compute+calcul
CANADA

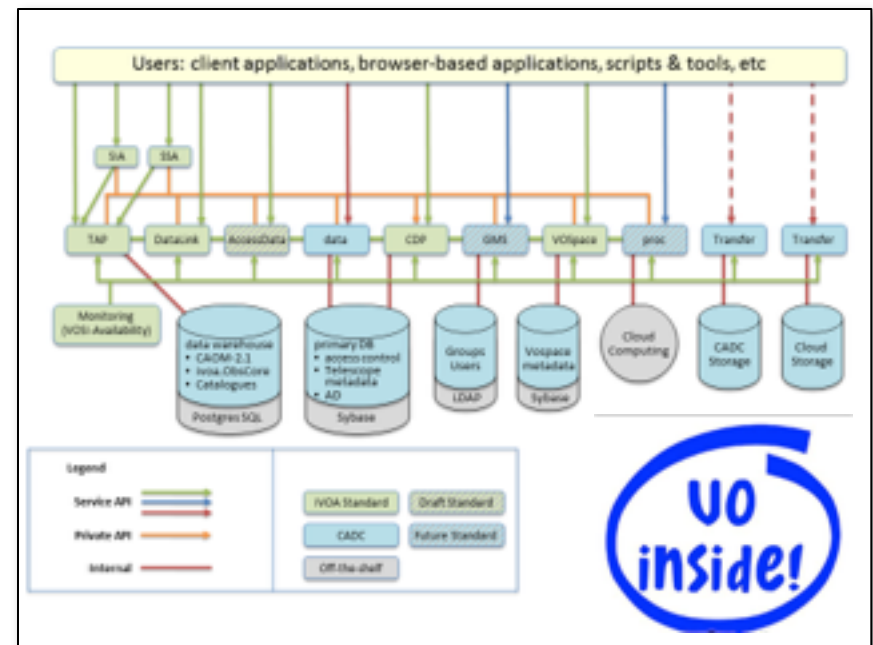


The Starting Point – Integrated VO services

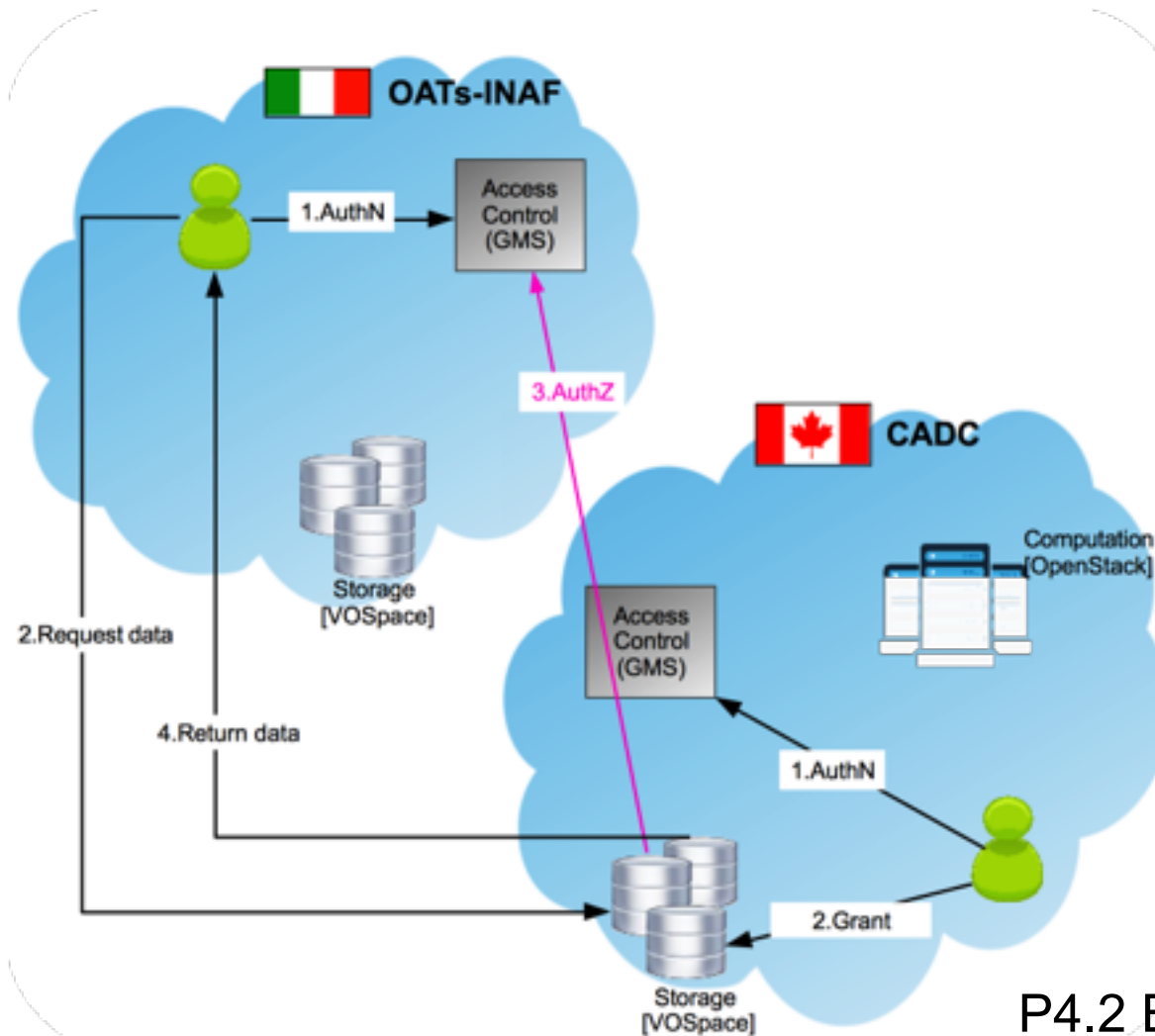
- VOSpace: User storage
- Single-Sign-On Profile
- Credential Delegation Protocol
- Universal Worker Service
- VO Support Interfaces

+

- In CANFAR
 - Group Management Service



Interoperating VOSpaces



P4.2 Bertocco et al.

Next Steps: IVOA

- “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.”



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



Next Steps

- Resource providers (telescopes, archives, data centres, processing centres) to interoperate authorizations
 - Integrate authorization across all resources
 - Allow external trusted group information providers
 - Better to allow users to create and manage groups
 - Better to allow users to grant authorization to their allocated resources
- Working with Compute Canada and EGI
- The exploratory project AENEAS

