EGI-Engage

CANFAR federation roadmap

D4.1

Date	31 August 2015
Activity	WP4
Lead Partner	INFN
Document Status	FINAL
Document Link	https://documents.egi.eu/document/2549

Abstract

CANFAR is an integrated cloud infrastructure to support collaborative astronomy (A&A) teams, especially those with data intensive projects. This document summarizes the first 12 months of activity of the implementation plan for CANFAR and EGI federated cloud in support of data-intensive collaborative astronomy research. The activity described is expected to promote interoperability between EGI and CANFAR e-Infrastructure. The federation model proposed in this document is based on the assumption that the two clouds will remain independent and independently managed but users and projects will be able to use both e-Infrastructures for data sharing and computing. The e-Infrastructure federation model follows the *Open Science Commons* vision where researchers from all disciplines have easy and open access to the innovative digital services, data, knowledge and expertise they need for collaborative and excellent research.

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program under Grant number 654142 http://go.egi.eu/eng

COPYRIGHT NOTICE

DELIVERY SLIP

	Name	*Partner/Activity*	*Date*
From:	Giuliano Taffoni	INAF/WP4	10/08/2015
Moderated by:	Patrick Mann	Western Canada Research Grid	14/08/2015
Reviewed by	Andre` Schaaff Roberto Pizzo	Observatoire Astronomique de Strasbourg Astron	14/08/2015 14/08/2015
Approved by:	AMB and PMB		31/08/2015

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author/Partner*
v1.0	10/08/2015	Version ready for external review	Giuliano Taffoni/INFN Enol Fernandez/EGI.eu Severin Gaudet/IVOA
v2.0	27/08/2015	Version after external review	Giuliano Taffoni/INFN
FINAL	31/08/2015	Final version	Giuliano Taffoni/INFN

TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/

Here we report the document specific glossary

Acronym	
A&A	Astronomy and Astrophysics
AAI	Authentication and Authorization Infrastructure
ADASS	Astronomical Data Analysis Software and Systems
CADC	Canadian Astronomy Data Centre
CANARIE	Canadian Network for the Advancement of Research, Industry and Education
CANFAR	Canadian Advanced Network for Astronomical Research
CSIRT	Computer Security Incident Response Team
CTA	Cherenkov Telescope Array

FUSE	Filesystem in User Space
GMS	Group Management Service
IaaS	Infrastructure-as-a-Service
IGTF	Interoperable Global Trust Federation
INAF	Italian National Institute of Astrophyiscs
IVOA	International Virtual Observatory Alliance
RESTFul	REST (Representational State Transfer) system that communicates over HTTP with the same HTTP verbs (GET, POST, PUT, DELETE, etc.) that web browsers use to retrieve web pages and to send data to remote servers.
SKA	Square Kilometer Array
VM	Virtual Machine
VO	Virtual Organization
VObs	Virtual Observatory
VOSI	IVOA Support Interfaces
VOSpace	Virtual Observatory Storage System

Contents

1 Introduction 5
2 Cloud Federation Roadmap Definition 7
2.1 Federation sustainability 9
3 CANFAR cloud infrastructure 10
3.1 Authentication and Authorization 11
3.1.1 Group Management Service 11
3.1.2 Credential Delegation 11
3.2 Distributed storage 12
3.2.1 The VOSpace service 12
3.2.2 Data transfer service 12
3.3 Monitoring service 12
3.4 Computing Capabilities 12
4 Federation Roadmap 13
4.1 Core Infrastructure platform 14
4.1.1 Authentication and authorization infrastructure. 14
4.1.2 Service Registry and Marketplace 16
4.1.3 Accounting 17
4.1.4 Monitoring 18
4.1.5 Security monitoring 19
4.2 Cloud Realm 19
4.2.1 VM Image Catalogue and Management 19
4.3 Community Cloud 20
4.3.1 Data and storage federation 20
4.4 Risk analysis 22
5 Data policies 24
6 Partners involved in the Federation activity 25
7 References 26

1 Introduction

This document summarizes the implementation plan for CANFAR and EGI federated cloud in support of data-intensive collaborative astronomy research. The Canadian Advanced Network for Astronomical Research[1] (CANFAR) is a digital infrastructure for Astronomy and Astrophysics (A&A) based on cloud storage and cloud processing middleware and on tools and services developed by the International Virtual Observatory Alliance (IVOA)[2].

A&A community has gathered rich experiences in cloud computing within the CANFAR federated cloud deployed on Compute Canada resources and operated by National Research Council Canada. However, Europe is currently missing a cloud infrastructure for A&A. The Canadian cloud infrastructure represents a unique example of an A&A oriented infrastructure that joins together the IaaS and the standards and services developed by the IVOA e.g. for user authentication and authorization, data sharing, access to data and archives, and finally data processing.

Scope of the cloud federation is to

- **Extend the portfolio** of EGI federated cloud capabilities, through integration of new unique services based on IVOA standards and customization of generic EGI services (in particular clouds) to A&A requirements.
- Provide a new **innovative cloud infrastructure** built for European astronomers and Astronomical Data Centers.
- Provide **close collaboration** of e-science infrastructure between EU and Canada. One of the basic additional requirements to engage the community at large, is to interface other non-European infrastructures considering that the A&A ESFRI projects are world wide collaborations, to provide a uniform and seamless access to heterogeneous resources necessary to organize and process high data volumes.

Traditionally A&A has been at the forefront of implementing digital repositories, e.g. for sky observations with ground and/or space based telescopes. Typically such repositories are maintained within data centres with appropriate provision for tools/services for access and analysis. Data archives are expanding rapidly, e.g. through flagship high data volume generating A&A projects, and EuroVO[3] has recently identified over 70 EU data centres. A&A data centres are employing Virtual Observatory (VObs) to provide seamless unified access to distributed and highly heterogeneous data archives. The use of IaaS cloud computing facilities is thus becoming increasingly important. e.g. EuroVO emphasizes that relation to the VObs should be taken into

[1] http://www.canfar.phys.uvic.ca/canfar/
[2] http://ivoa.net
[3] Census of European DataCentres http://goo.gl/zOWykc

appropriate consideration in order to provide a complete data usage ecosystem for A&A communities[4].

The activity described will ensure interoperability between EGI and CANFAR e-Infrastructure, the federation model proposed in this document is based on the assumption that the two clouds will remain independent and independently managed but users and projects will be able to use both e-Infrastructures for data sharing and computing. The e-Infrastructure federation model follows the *Open Science Commons* vision where researchers from all disciplines have easy and open access to the innovative digital services, data, knowledge and expertise they need for collaborative and excellent research.

The roadmap presented in this document, has been defined through a collaboration process between EGI and the Canadian Astronomy Data Centre that operates CANFAR. The Italian National Institute of Astrophysics (INAF) is the Italian NGI partner and EGI A&A community coordinator. INAF is working on defining the federation roadmap and implementation in Europe and it will provide the storage and cloud resources for the federation.

This roadmap will cover 12 month of activity starting from month 6 from the beginning of the EGI-Engage project (here after M6). At the end of M18 the roadmap will be revisited and integrated with the development and integration activities identified during the first 12 months.

Furthermore, a well-defined procedure will allow us to periodically (4 Months cycles) revise this roadmap accordingly to new requirements from CANFAR and EGI.

[4] EuroVO- CoSADIE Project http://www.eurovo.org

2 Cloud Federation Roadmap Definition

The roadmap presented in this document has been defined taking into account the requirements collected from European and Canadian resource providers and technology providers and from the A&A user community.

The planned activities will allow us to expand the EGI capacity and capabilities by integrating its technical solutions with those offered by CANFAR and to provide to A&A users and projects a uniform and seamless access to heterogeneous resources necessary to organize and process large data volumes (in the range of few hundreds of TB up to 1PB). This activity will also enrich the EGI Open Data Access platform with A&A services that will provide capabilities to publish, use and reuse openly A&A accessible data.

The cloud federation approach offers highly flexible and scalable storage and networking resources that allows managing also extremely large amount of data (many Petabytes). It allows to easily expanding the infrastructure on-demand not only increasing the single provider storage capabilities but also involving new data providers that expand the overall federation capabilities. In fact, A&A ESFRI projects, SKA or CTA, are currently evaluating the suitability of the federated IaaS paradigm in fulfilling their large-scale data and computational needs. Moreover, the A&A communities are increasingly recognizing that an approach along the lines of the proposed federation is useful for satellite data analysis where well-defined computing resources and software are necessary (such as specific OS vendor and version and library applications), e.g. the EUCLID[5] project is developing a community cloud based on open standards.

CANFAR is a cloud processing and cloud storage infrastructure that integrates authentication and authorization, monitoring, virtual storage environment and computing capabilities. The Federation model we propose in this document is based on two use cases:

- Authentication and Authorization Infrastructure (AAI) federation. CANFAR is now offering resources also to European users and groups. Users are registered in the CANFAR AAI service and they are issued CANFAR credentials. An interoperable AAI will allow European users to access CANFAR resources (data and computing) using their European (EGI Federated Cloud) credential. At the same time Canadian astronomers will be able to access EGI Federated Cloud resources using their Canadian credential. This will encourage collaboration between Canadian and European astronomers, projects and Data Centers. By implementing the AAI federation, users will exploit transparently resources provided via the integration with EGI Federated Cloud (EGI Federated Cloud).
- Data federation. CANFAR is offering virtual storage based on IVOA standards (VOSpace [R2]), which is used by Astronomers and data centers (CADC) to store and share data. We would allow data access and sharing from A&A community and offer new capabilities to European

[5] http://www.euclid-ec.org

data centers to share open data to astronomers and citizens using EGI Federated Cloud. It will allow data sharing (e.g. to replicate open and private data for data availability and preservation) and Virtual Machines (VM) sharing between Canada and Europe. Finally, it allows A&A community to move computation close to data rather than moving large amount of data.

The final goal of this activity is to provide interoperable access to storage resources for both European and Canadian users.

The EGI Federated Cloud is a layered infrastructure as shown in Figure 1. CANFAR is a Community Cloud Platform that implements also some specific Core Infrastructure Services: it has a specific AAI that is based on IVOA specifications, a monitoring service and a federated management. It is then build on top of Compute Canada Infrastructure that provides the Physical Layer, some Core services and the Cloud Realm.

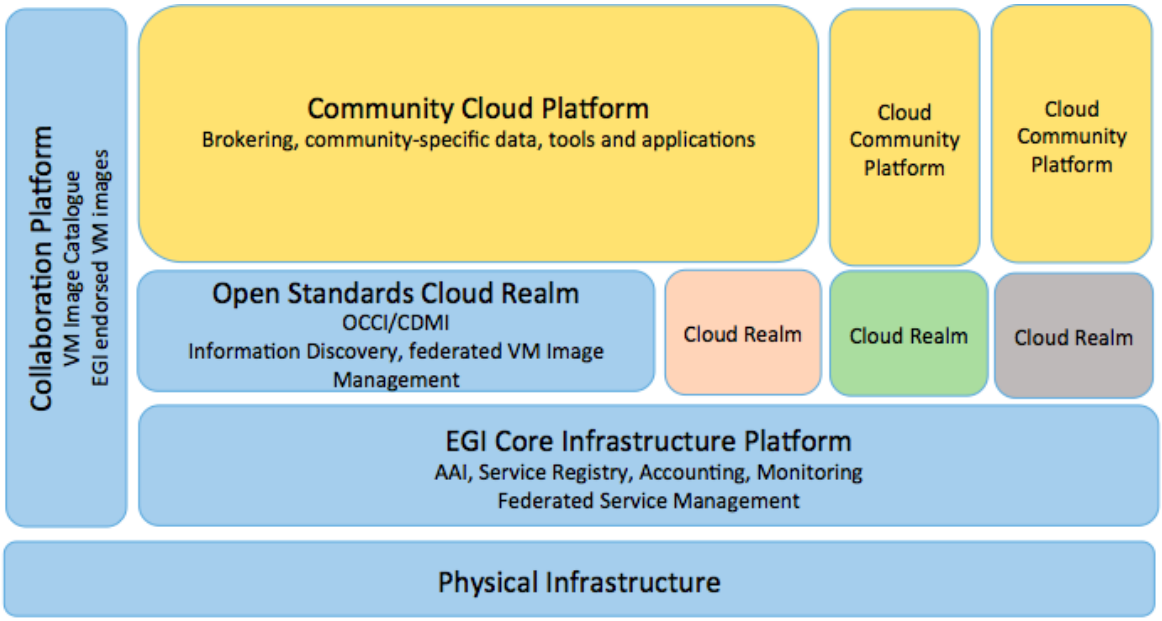

Figure 1 EGI Federate Cloud Layered Infrastructure

Federating the two infrastructures will imply the discussion and analysis of the different layers both at CANFAR and EGI. This roadmap will present the steps that drive the achievement of the federated environment. Some activities require services analysis and policy definition, some other service development on top of EGI Federated Cloud Core Infrastructure Platform and Cloud Realms.

2.1 Federation sustainability

This roadmap describes the activities to implement a federated cloud that involves two existing cloud infrastructures that provides production environments in Canada and Europe. While EGI-Engage project is providing enabling solutions to build the federation, it allows its member to find their own sustainability both for what regards the infrastructure procurement and maintenance and community services sustainability. The infrastructures involved in the federation are both sustained by their funding agencies and institutions. For what regards the European research centres involved in the federation activities, in Italy INAF provides and maintains the storage and computing capabilities. Initially two clusters for a total of 350 Cores and 100 TB of storage will be available, that will grow up to 1000 Cores and 1 PB in the next 3 years.

EGI.eu will maintain tools and services developed by EGI-Engage and necessary to support EGI Federated Cloud thanks to the EGI participants' early fees that contribute to the recovery of the running costs sustained to operate the federation services beyond the project lifetime. This assures sustainability to the project outputs.

For what regards A&A community cloud services developed during the EGI CANFAR federation activities, their sustainability will depend mainly on their relevance for the A&A community. In fact, they are largely used by CANFAR that demonstrate the importance for A&A community of a cloud-based infrastructure. INAF Astronomical Data Center (IA2) will adopt the services and tools developed and maintain them. Moreover, as discussed in the previous section, the integration herein described is a pathfinding activity for long term initiatives like SKA, CTA, EUCLID[6] etc. If adopted, these will contribute to their sustainability. This is also a factor of sustainability together with the dissemination activities that will be done by the partners participating to conferences and workshops of reference for the A&A community (e.g. IVOA meetings, ADASS[7] conference).

[6] http://www.euclid-ec.org
[7] http://www.adass.org

3 CANFAR cloud infrastructure

The Canadian Advanced Network for Astronomical Research (CANFAR) is a digital infrastructure combining the Canadian national research network (CANARIE), cloud processing and storage resources (Compute Canada) and an astronomy data centre (Canadian Astronomy Data Centre – CADC) into a unified ecosystem for storage and processing. It is an operational system for the delivery, processing, storage, analysis, and sharing of very large astronomical datasets: CANFAR stores about 2 PBs of data and 900 Million of files. The project has combined the best features of the grid and cloud processing models by providing a self-configuring virtual cluster deployed on multiple cloud clusters. CANFAR processed about 2.7 Million of jobs for about 500 cores per year. The project has provided users a robust and secure virtual storage environment layered on distributed storage resources. The CANFAR services make use of many technologies from the grid, cloud and International Virtual Observatory (IVOA) communities. Initially deployed in 2011, CANFAR has continuously evolved in response to operational needs and user input.

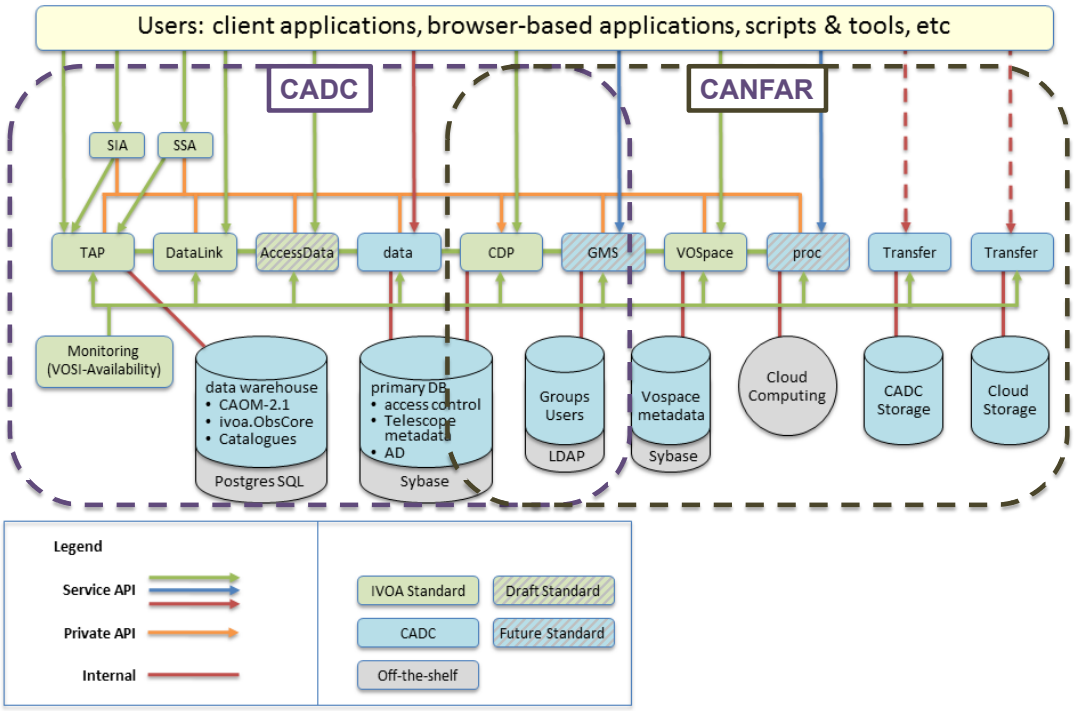

Figure 2 CANFAR infrastructure schema. This picture shows the main components of the infrastructure and the connection between the IVOA services offered by CADC and the CANFAR computing services.

CANFAR has four principal user-facing services: user storage (VOSpace), virtual machines on demand (Horizon/NOVA), batch processing (proc/Condor/CloudScheduler/NOVA) and group management (GMS), all RESTful web services accessible via http or browser user interfaces. CANFAR is developed, maintained and operated by the CADC with contributions from several Canadian universities. In Figure 2 we provide a description of CANFAR infrastructure and the connection between the IVOA services offered by CADC and the CANFAR computing services. CADC is offering data access to astronomical archive data using IVOA standards. Both proprietary and public data can be accessed. An authentication and authorization infrastructure is used to provide access to proprietary data. The same infrastructure is used by CANFAR to allow access to computing and storage resources providing the computing and storage services for proprietary data analysis.

In this section we will briefly describe CANFAR services we are proposing to integrate and federate.

3.1 Authentication and Authorization

The Authentication and Authorization infrastructure is used to provide access to data and resources to users (see Figure 2). Each user is identified by username/password and a X.509 certificate[8] [R8]. If the user has a certificate it can be uploaded to the service otherwise a CANFAR certificate is automatically issued. Any certificate that has been issued by a Certification Authority (CA) from a member of the IGTF[9] can be accepted by CANFAR.

A credential delegation protocol [R4] (based on the x.509 user certificate) allows applications and services to delegate a user's credentials to other services such that that service may make requests of other services in the name of that user.

3.1.1 Group Management Service

To manage authorization CANFAR implements a RESTful [R6] web service to manage groups and membership. It supports three classes of operations:

- Creating, getting, updating and deleting a group.
- Adding and deleting a user to a group.
- Adding and deleting a group to a group.

3.1.2 Credential Delegation

Credential delegation is an IVOA RESTful web service that enables users to create temporary credentials at a site and thus enable that site to perform web service actions on their behalf [R4]. This feature is the key to allow services from different sites to interact (e.g. CANFAR VOSpace using an EU Group Management Service for authorization).

[8] X.509 is an ITU-T standard for a public key infrastructure and Privilege Management Infrastructure. It is a digital certificate that can be used also to authenticate users.
[9] www.igtf.net

3.2 Distributed storage

The distributed storage layer allows users to access, process and share their data. It is based on VOSpace IVOA service [R1] and on a data transfer service. CANFAR provides also a FUSE [R7] VOSpace implementation to allow users to mount their VOspace on virtual machines

3.2.1 The VOSpace service

VOSpace is an IVOA standard RESTful web service to organise and share data. Data sharing can be public (allowing anonymous access) or to a project team (via groups of users from a Group Management Service). The service is based on the specification provided by IVOA document [R1].

3.2.2 Data transfer service

The data transfer service is responsible for handling the upload and download of data to and from physical storage locations for VOSpace. The VOSpace service will point users to preferred data transfer service based on the requirements of the user's file transfer.

3.3 Monitoring service

Each web service supports a RESTful monitoring interface that complies with the VOSI-availability (IVOA) standard. These interfaces can be monitored via periodic requests from many standard monitoring packages (e.g. Nagios) to provide a central view of a system. The IVOA Support Interfaces (VOSI) is discussed in [R2].

3.4 Computing Capabilities

CANFAR is offering a cloud computing facility based on OpenStack that allow the creation of on demand virtual machines and data-location-aware virtual clusters to process and analyse data. Web interfaces allow users to create both single machines and virtual clusters to execute batch jobs.

The system is based on standard technologies (proc/Condor/CloudScheduler/NOVA) and IVOA standards that allow the execution and management of jobs in the cloud (the IVOA Universal Worker Service [R3]).

CANFAR is processing 2.7 million of jobs with a request of about 500 cores per year.

4 Federation Roadmap

The Roadmap for a federated cloud between CANFAR and EGI Federated Cloud will cover the following aspects:

- Core Infrastructure platform
  - Authentication and authorization infrastructure
  - Service registry and market place
  - Accounting
  - Monitoring and Security Monitoring
  - Operations tools and services
  - Cloud RealmFederated VM management
  - VM Image Catalogue and Management
- Community Cloud
  - Interoperable virtual storage: VOSpace
  - Cloud2Cloud data transfer.

Table 1 – Services and tools available at EGI and CANFAR and federation activity

SERVICE	*Activity*	*EGI Service involved*	*CANFAR Service Involved*
AAI	Analysis and Development	X.509 and VOMS	X.509 and GMS
Service registry	Analysis, requirement collection and policy definition	Service Registry (GOCDB)	IVOA Registry service
Accounting	Analysis and requirement collection	APEL, accounting portal	None
Monitoring	Analysis, requirement collection and policy definition	ARGO monitoring tools	VOSI-availability monitoring service

Security Monitoring	Analysis and policy definition	EGI Computer Security Incident Response Team (CSIRT)	None

The federation roadmap will imply technical discussions and policy definitions to identify the services the EGI and CANFAR are interested in federating.

If technical development of community, realm or infrastructure service is required, the activity will conclude with a new software release and a verification and validation test to measure the progress in terms of capabilities of the software running on the infrastructures.

The federation activity will conclude with a demonstrator based on the use cases discussed in the introduction.

EGI and/or CANFAR bug-tracking systems will be used allow users to reports bugs that may arise at later stage when services are in production. For example, a new Support Unit can be register in the GGUS[10] system to create incident records, to record the request and tracks the ticket from creation through to solve.

The development and implementation activity will be done in collaboration with the other tasks and sub tasks of WP4 and with other WPs, in particular WP3 and SA2 that gathers the requirements and coordinates the provisioning of services for scientific communities.

The Roadmap will be updated during the project lifetime according to:

- Federation Requirements gathered through the technical discussions and/or implementation activities and stored in the EGI ticket system;
- New services or tools developed by EGI and/or CANFAR;
- Upgrades of the infrastructures implemented during the project lifetime.

Roadmap updates will cover prioritization of actual federation activities or definition of new activities (each 4 months). A major roadmap revision and integration will happen at M18 and will cover the activity of the next 12 months.

4.1 Core Infrastructure platform

4.1.1 Authentication and authorization infrastructure.

This activity will explore the federation of EGI Federated Cloud and CANFAR AAI methods and will:

[10] https://ggus.eu/

- Enable users to access the EGI Federated Cloud (and services) and CANFAR with the same credentials.
- Enable A&A community to collaborate using both EGI and CANFAR infrastructures
- Enable access to interoperable services for data sharing and computing
- Verify the possibility of implementing a common identity federation (e.g. EduGain)

This activity will be done in collaboration with other EGI-Engage WPs, in particular WP3 JRA1.1 that is in charge of exploring and integrating AA methods and SA2.

This activity already started at month 2 from the beginning of EGI-Engage project (M2) with a few technical discussions and a face-to-face meeting at M4. Some technical aspects have been identified. The focus is to interoperate authorization. Interoperating Authentication is out of scope but could be layered on interoperable identity services provided by EGI.eu and Compute Canada once they exist.

EGI Federated Cloud and CANFAR use X.509 certificate to identify users and to delegate user credential when using services. CANFAR users are also registered to a Group Management Service (GMS) that is in charge of authorization procedures. GMS is queried by CANFAR services to verify user capabilities and capacities. The focus of AAI roadmap is to implement an interoperable GMS service on top of the EGI Federated Cloud at INAF and to achieve group membership resolution from CANFAR to EGI GMS. We will analyse the use of VOMS as a backend for EGI GMS taking into account GMS requirements in terms of capabilities and flexibility.

CANFAR will implement VOMS extensions on the Credential Delegation protocol [R4] to achieve VO membership resolution from EGI Federated Cloud to CANFAR storage and computing services

Table 2 – AAI roadmap

Activity	***Start***	***Release***	
Technical Analysis	M2	M6	• Identification of AAI at CANFAR and EGI • Technical requirements collections • Service development strategy Mean of verification: • Monthly Technical meetings and teleconferences • Participation at 2015 EGI conference
Requirements collection and policy definition	M2	M18	• Contribution to technical requirements for EGI AAI and to the identification of common Identity providers for user accounting. • Contribution to the definition of EGI AAI. Mean of verification: • Monthly Technical meetings and

			teleconferences • Participation to EGI conferences and user forums
Technical architecture and development	M7	M16	• GMS development activities Mean of verification: • Software release
Testing	M16	M17	• Implementation and Test Mean of verification: • Verification and testing complete

At the end of AAI federation we will test the implementation: EGI Federated Cloud registered users will access CANFAR VOSpace, upload/download data, and grant access to her data to other users registered both in CANFAR and EGI. Group of users will be created and different roles will be assigned to group users (member, leader, and administrator). GMS capabilities will be tested (e.g. a group administrator will add new users to the group and new capabilities to a user). The execution of Virtual Machines will be also tested.

4.1.2 Service Registry and Marketplace

Sharing and discovering research resources (instruments, computing, software, data, etc.) and services (consulting, sample preparation, etc.) are essential. EGI provides a service registry for configuration management of e-Infrastructure and is developing a “marketplace” concept where free and paid resources can be listed and discovered.

Thanks to the experience of the IVOA resource registry, CANFAR will participate in technical requirements collection and further discussions for the design of the EGI marketplace service. This activity aims at:

- Discussing if and at which level CANFAR federated resources will populate the service and resource catalogue of EGI and the related policies.
- Collecting requirements for EGI marketplace and service registry to interoperate with IVOA resource registry tools.

Table 3 – Service Registry roadmap

Activity	*Start*	*Release*	
Technical Analysis	M8	M10	• Identification of service registry tools at EGI and CANFAR Mean of verification: • Monthly Technical meeting and teleconferences
Requirements collection and policy definition	M9	M10	• Federation strategy: define if marketplace service will be implemented as part of the Federation. • Contribution to technical requirements for EGI marketplace and service registry. • Possible contribution to registry and marketplace developments to interoperate with IVOA tools. Mean of verification: • Services registered in EGI service registry. • Services listed in EGI marketplace if decided. • Monthly Technical meetings and teleconferences • Participation in EGI conferences and user forums.

4.1.3 Accounting

The EGI accounting tool collects accounting data from sites participating in the EGI and WLCG infrastructures as well as from sites belonging to Grid resources, about 350 resource centres are participating to the accounting infrastructure. The aim of a federated accounting service is to collect, aggregate and display usage information across the whole federation. The accounting information is gathered from different sensors into a central repository where statistics are generated and displayed. Statistics regarding the number of VM instantiated and consumed, the total amount of CPU consumed, and storage information, GPU accountings are gathered. They are displayed from different point of view: user, Virtual Organization (VO), site, etc. Using this information a user or a VO can actively monitor the use of its resources and know in case the resources he/she requested are underused.

The aim of this activity is to discuss the inclusion of a federated accounting system in the EGI-CANFAR federation and eventually how to contribute to the evolution of the EGI accounting system.

Table 4 – Accounting roadmap

Activity	*Start*	*Release*	
Technical Analysis	M10	M12	• Accounting service analysis: current solutions at EGI and CANFAR. Mean of verification: • Monthly Technical meeting and teleconferences
Development activity	M10	M18	• Federation strategy: define if accounting service will be implemented as part of the Federation. • Contribution to technical requirements for EGI accounting tools. • If federation will involve accounting, integrate accounting at CANFAR and EGI. Mean of verification: • Accounting implementation if decided

4.1.4 Monitoring

Monitoring a federated cloud infrastructure is a complex task as it must integrate the existing cloud ecosystems and provide a modular and scalable framework. EGI Federated Cloud is developing the ARGO platform and CANFAR has a monitoring tool based on the VOSI-availability standard proposed by the IVOA (section 3.3). The aim of this activity is to discuss the inclusion of a federated monitoring system in the EGI-CANFAR federation and eventually define the implementation strategy.

Table 5 – Monitoring roadmap

Activity	*Start*	*Release*	
Technical Analysis	M12	M14	• Monitoring service analysis: current solutions at EGI and CANFAR. Mean of verification: • Monthly Technical meeting and teleconferences

Development activity	M12	M18	• Federation strategy: define if monitoring service will be implemented as part of the Federation • If federation will involve also monitoring, develop and implement a federated monitoring model. Mean of verification: • Federated Monitoring feasibility study if decided.

4.1.5 Security monitoring

The EGI security monitoring service is in charge of managing security incidents. This is crucial to prevent problems that affect users, sites and infrastructure providers. This activity will identify requirements from both EGI and CANFAR and define a common policy for security monitoring.

Table 6 – Security Monitoring roadmap

Activity	*Start*	*Release*	
Technical Analysis	M14	M16	• Security Monitoring Service analysis: current solutions at EGI and CANFAR. Mean of verification: • Monthly Technical meeting and teleconferences
Policy definition	M15	M18	• Requirements collection • Common policy definition Mean of verification: • Common policies

4.2 Cloud Realm

4.2.1 VM Image Catalogue and Management

EGI provides a VM image catalogue consisting on an open library of virtual appliances (bundle of one or more VM images) for use on clouds or for personal download, supporting VM image management operations including: registration of new instances, reuse of existing ones and contextualization. The catalogue also provides tools for automatically and securely distributing

the images to any resource provider. This activity will identify requirements from both EGI and CANFAR to use the VM image catalogue and integrate the distribution of images within the two infrastructures.

Table 7 – VM Image Catalogue roadmap

Activity	*Start*	*Release*	
Technical Analysis	M14	M16	• VM Image Catalogue: current solutions at EGI and CANFAR. Means of verification: • Monthly Technical meeting and teleconferences
Development activity	M15	M18	• Federation strategy: define if VM image catalogue service will be implemented as part of the Federation • If part of the federation, integrate VM Image management tools. Mean of verification: • Feasibility study of federated VM Image distribution and definition of development activity roadmap in M18-M26.

4.3 Community Cloud

This section discusses the federation of services offered by the CANFAR community cloud and the EGI Federated Cloud.

4.3.1 Data and storage federation

This activity focuses on identifying technical solutions for federating data between EGI Federated Cloud and CANFAR and will:

- Enable users to seamless access, move and process data on CANFAR and EGI Federated Cloud
- Enable a data platform for A&A community and projects;
- Enable user-initiated data transfers between CANFAR and EGI for replica or preservation.
- Enable data centers to replicate data for preservation
- Enable data centers, users or projects to optimize data distribution in order to improve data analysis and data reduction capabilities. Users will execute processing tasks close to data rather that move data close to their computing resources when necessary

This activity has already started at M2 with technical discussions and a face-to-face meeting at M4. Some technical aspects have been already identified.

EGI Federated Cloud technologies and services will be evaluated; we will verify different approaches to data federation that will be implemented during the second year of activity. Additionally this activity will develop and implement a VOSpace implementation on top of EGI FedCloud;

1) Develop and implement a VOSpace implementation on top of EGI Federated Cloud;
2) Study possible interoperability between CANFAR VOSpace and EGI Open Data platform based on OneData;
3) Implement a user-initialized server-to-server transfer based on gridftp based solution or the EGI File Transfer Service.

More will be discussed in M6-M18 and a suitable approach will be identified.

This activity will be done in collaboration with other EGI-Engage WPs, in particular WP4 JRA 2.1 and JRA2.2 that will provide support to study a possible integration of VOSpace with EGI Open Data platform.

Table 8 – Data federation roadmap

Activity	***Start***	***Release***	
Technical Analysis	M2	M10	• Technical requirements collections • Identification of Virtual storage service at CANFAR • Identification of currently available services at EGI for data storage and data transfer • Service development strategy Means of verification: • Monthly Technical meetings and teleconferences • Participation in the 2015/2016 EGI conference • Participation in the EGI user forum
Technical architecture	M11	M18	• Feasibility study for federated storage: implementation activities identification and development plan • Data Transfer development architecture definition Means of verification: • Development plan

Requirements and policies definition	M2	M18	• Contribution to technical requirements for EGI Data cloud • Identification of other EGI Federated Cloud sites to support data replication Mean of verification: • Monthly Technical meetings and teleconferences • Participation in EGI conferences and user forums

At the end of the EGI-Engage project we will perform a set of final tests. The tests will verify that users and groups will be able to access EGI Federated Cloud VOSpace upload and manage data, grant access to the data both for CANFAR and EGI users and groups.

Data that will be replicated in Europe will be the CADC Canada France Hawaii Telescope public data. Data will be stored at the INAF EGI Federated Cloud site. Other sites for data replication will be identified. Some Large Binocular Telescope data and/or Italian Galileo National Telescope data will be replicated in Canada. About 100TB of data will be exchanged between CANFAR and EGI FedCloud. INAF will provide at least 100TB of storage for testing and pre-production. Implementation and testing will be done during M18-M30 of the project.

4.4 Risk analysis

Table 9 – Critical risks for implementation

Description of risk	***Activity involved***	**Proposed risk-mitigation measures**
Shortfall of externally performed tasks: delay in providing the integration of CANFAR GMS on top of EGI services impacts cloud realm integration	AAI	CANFAR infrastructure in Canada will be available from M6 for porting, testing and verification.
Both CANFAR and EGI Federated Cloud move towards a different incompatible cloud implementation.	all	CANFAR is using the Openstack cloud open source software that is fully supported by EGI. The use of OCCI standards will be discussed for any new or unsupported cloud stacks.
A production GMS will not be ready	AAI	The CANFAR GMS can also be used by

at M18.		EGI users. Other EGI-Engage WP will be involved in the developing activity in particular JRA1 to complete the integration.
Common monitoring service will not be part of the federation	All	The European cloud sites that will support A&A will install both EGI Federated Cloud and CANFAR monitoring System
Common accounting service will not be part of the federation	Accounting	Only European cloud sites that support A&A will install EGI Federated Cloud Accounting system.
Security monitoring policy not in place at M18	Computing and VM distribution	A common security policy must be in place if we decide to share also VM. Otherwise only data sharing will be available. Discussion should also involve also Compute Canada.
No other EGI Federated Cloud sites will support the CANFAR Virtual Organization	ALL	INAF will provide a minimum set of cloud resources (100 cores with 8GB ram per core) and 100 TB of storage.

5 Data policies

Data that will be shared between CANFAR and EGI Federated Cloud is public astronomical data. CADC and INAF are also hosting private data from projects in their astronomical archives. At this stage no proprietary data will be shared however data policies will be analysed and discussed to verify compatibility between CANFAR and EGI Federated Cloud policies.

Table 9 – Data policy discussion

Activity	***Start***	***Release***	
Technical Analysis	M14	M16	• Data policies analysis: current solutions at EGI and CANFAR. Means of verification: • Monthly Technical meeting and teleconferences
Policy definition	M15	M18	• Data replication policy definition for data storage and federation (Section 4.3.1) • Common policy definition if necessary Means of verification: • Common policies

6 Partners involved in the Federation activity

The **Canadian Astronomy Data Centre** (CADC)[11] was formed in 1986 and has developed into the national data hub for Canadian astronomy. It is now one of the largest astronomy data centres in the world and certainly one of the most heterogeneous with data for multiple missions, facilities and wavelengths with pointed and survey observations. The archive houses 434 Terabytes of astronomy data from 12 telescopes and 6 advanced data product collections. During the last year 1.1TB of data has been handled and 91 Million of files.

In 2013, CADC services were accessed by 511 authenticated users and over 3,000 anonymous users and delivered 353 TB of data to those users. The CADC is also the technical hub of the Canadian Advanced Network for Astronomical Research (CANFAR).

The CADC is part of the NRC Herzberg, the Astronomy and Astrophysics program of the National Research Council of Canada[12]. It has a staff of 20 (6 scientists, 9 software developers and 5 operations). The CADC is also an active participant in the International Virtual Observatory Alliance and has deployed many Virtual Observatory services, both for CADC data collections and CANFAR services. In EGI-Engage CADC is playing the role of a technology provider of the CANFAR cloud infrastructure and it works on the preparation of this federation roadmap and on the implementation of the federated cloud.

Istituto Nazionale di Astrofisica (**INAF**[13]) will participate in this activity with one of its research institutes, namely the Astronomical Observatory of Trieste. INAF has been active for several years in the fields of Grid technologies (deployment of infrastructure and integration of domain-specific applications) and archives of astronomical data (from both ground-based and space-borne facilities). The INAF site of Trieste is an EGI node providing several dozens of CPUs. INAF is a partner of the Italian NGI (IGI) with an active participation in user support and middleware developing. INAF was also involved in Euro-VO projects (such as AIDA, DCA, TECH and currently ICE and CoSAIDE) and in IVOA working groups. INAF has also the main Astronomical Data Centre in Italy, IA2; it is involved in CTA (in particular the Astronomical Observatory of Catania and the Astronomical Observatory of Monteporzio are participating to the WP of computing and storage) and SKA (Astronomical Observatory of Catania and Astronomical Observatory of Trieste INAF will provide HW resources in kind for implementing cloud infrastructures and testing and to support all production activities.

[11] http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/

[12] http://www.nrc-cnrc.gc.ca/eng/rd/nsi/

[13] http://www.inaf.it

7 References

No	Description/Link
R1	VOSpace specification Version 2.0 http://www.ivoa.net/documents/VOSpace/20130329/index.html
R2	IVOA Support Interfaces http://www.ivoa.net/documents/VOSI/index.html
R3	The universal worker service http://www.ivoa.net/documents/UWS/index.html
R4	Credential delegation protocol http://www.ivoa.net/documents/CredentialDelegation/
R5	SSO profile http://www.ivoa.net/documents/latest/SSOAuthMech.html
R6	Principled Design of the Modern Web Architecture Roy T. Fielding e Richard N. Taylor, in ACM Transactions on Internet Technology (TOIT), vol. 2, nº 2, New York, Association for Computing Machinery, 2002-05, pp. 115–150, DOI:10.1145/514183.514185, ISSN 1533-5399
R7	Filesystem in Userspace (FUSE) http://fuse.sourceforge.net
R8	Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile https://tools.ietf.org/html/rfc5280