

**EGI-Engage**

CANFAR integration roadmap

D4.1

|  |  |
| --- | --- |
| **Date** | 07 July 2015 |
| **Activity** | WP4 |
| **Lead Partner** | EGI.eu |
| **Document Status** | DRAFT |
| **Document Link** | https://documents.egi.eu/document/XXX |

Abstract

CANFAR is a cloud for A&A, it has build to support astronomy projects, data tec. We present the idea. The infrastructure, the services and the tools needs to be federated for a complete federation of the two clouds. Finally we will present what will be done in this projects in term of the level of integration that will not involve all services. (about 120 words)

**COPYRIGHT NOTICE**



This work by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). The EGI-Engage project is co-funded by the European Union Horizon 2020 programme under grant number 654142.

**DELIVERY SLIP**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Name*** | ***Partner/Activity*** | ***Date*** |
| **From:** | Giuliano Taffoni | INAF/WP4 |  |
| **Moderated by:** |  |  |  |
| **Reviewed by** |  |  |  |
| **Approved by:** |  |  |  |

**DOCUMENT LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Issue*** | ***Date*** | ***Comment*** | ***Author/Partner*** |
| **v0.1** | 1/07/2015 | Document Structure and TOC | Giuliano Taffoni |
| **V0.2** | 9/07/2015 | Update on TOC and description of the Roadmap | GT |
| **...** |  |  |  |
| **v.n** |  |  |  |

**TERMINOLOGY**

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

**Contents**

1 Introduction 4

2 Cloud Federation Roadmap Definition 6

3 CANFAR cloud infrastructure 7

3.1 Authentication and Authorization 8

3.1.1 Group Management Service 8

3.1.2 Credential Delegation 8

3.2 Distributed storage 8

3.2.1 The VOSpace service 8

3.2.2 Data transfer service 8

3.3 Monitoring service 8

4 Federation Roadmap 9

4.1 Phase 1: Authentication and authorization infrastructure. 9

4.2 Phase 2: Interoperable VOSpace on EGI fedcloud 10

4.3 Phase 3: Server-to-server transfers 11

4.4 Roadmap implementation 12

5 Open data policies 14

6 Partners involved in the Federation activity 15

7 References 16

Appendix I. Appendix example 17

# Introduction

This document summarizes the development plan for CANFAR and EGI federated cloud federation until the end of the EGI/Engage project. The Canadian Advanced Network for Astronomical Research[[1]](#footnote-1) (CANFAR) is a digital infrastructure for Astronomy and Astrophysics (A&A) based on the OpenStack cloud middleware and on tools and services developed by the International Virtual Observatory Alliance (IVOA).

A&A community have gathered rich experiences in cloud computing within the CANFAR federated cloud operated in Canada by National Research Council. 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

* **extended portfolio** for 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 also 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[[2]](#footnote-2) has recently identified over 70 EU data centres. A&A data centres are employing 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 appropriate consideration in order to provide a complete data usage ecosystem for A&A communities[[3]](#footnote-3).

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.eu A&A community coordinator, it is working on defining the federation roadmap and implementation in Europe. Furthermore, a well-defined procedure will allow us to periodically (4 Months cycles) revise this roadmap accordingly to new requirements from CANAR and EGI.eu.

# 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 of the EGI capacity and capabilities by integrating its technical solutions with those offered by CANFAR e-Infrastructure and to provide to A&A users and projects a uniform and seamless access to heterogeneous resources necessary to organize and process high data volumes. 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.

CANFAR is an OpenStack based cloud infrastructure that integrates authentication and authorization, monitoring, virtual storage environment and computing capabilities. The Federation model we propose in this document is based 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 of CANFAR credentials. An interoperable AAI will allow European users to access CANFAR resources (data and computing) using their European (EGI federate cloud) credential. At the same time Canadian astronomers will be able to access EGI Federated Cloud resources using their Canadian credential. This will enforce 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 FedCloud.
* CANFAR Virtual Storage integration in EGI federated cloud. CANFAR is offering Virtual Storage based on IVOA standards (VOSpace), it is used by Astronomers and data centers (CADC) to store and share data. This integration will allow data access and sharing using the reference standard of the A&A community. It will offer new capabilities to European data centers to share open data to astronomers and citizens using EGI fedcloud. It will allow data sharing between data centers (e.g. to replicate open and private data for data availability and preservation). 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 transparent access to data and resources for both European and Canadian users.

Our federation model is …

# CANFAR cloud infrastructure

The Canadian Advanced Network for Astronomical Research[[4]](#footnote-4) (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. 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. 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.



CANFAR has four principal user-facing services: user storage (VOSpace), virtual machines on demand (VMOD), batch processing 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 this section we will briefly describe CANFAR services we are going to integrate and federate.

## Authentication and Authorization

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nec odio. Praesent libero. Sed cursus ante dapibus diam. Sed nisi. Nulla quis sem at nibh elementum imperdiet. Duis sagittis ipsum. Praesent mauris. Fusce nec tellus sed augue semper porta. Mauris massa. Vestibulum lacinia arcu eget nulla. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Curabitur sodales ligula in libero.

### Group Management Service

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nec odio. Praesent libero.

### Credential Delegation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nec odio. Praesent libero. Sed.

## Distributed storage

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nec odio. Praesent libero. Sed.

### The VOSpace service

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nec odio. Praesent libero. Sed.

### Data transfer service

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nec odio. Praesent libero. Sed.

## Monitoring service

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer nec odio. Praesent libero. Sed.

# Federation Roadmap

Here we discuss the roadmap. It will be based on 3 phases each phase will be divided in cycles each cycle will end with verification. Each phase will end with a test that verifies the

**Descrivi come operi: che tipo di implemetazione fai, le fasi, I clicli, I tests uso dei dati nella fase finale.**

## Phase 1: Authentication and authorization infrastructure.

This activity will explore how to federate EGI FedCloud and CANFAR AAI methods and it will:

* Enable users to access the EGI FedCloud (and services) and CANFAR with the same credentials.
* Enable A&A community to collaborate using both EGI and CANFAR infrastructure
* Enable access to interoperable services for data sharing and computing
* Verify the possibility to implement a common identity federation (e.g. EduGain)

This activity will be done in collaboration with other EGI-Engage WPs, in particular WP3 TJRA1.1 that is in charge of explore and integrate AA methods and SA2.1 and SA2.2 (Training & User support) for what regards A&A community requirements.

EGI FedCloud and CANFAR use X.509 certificate to identify users and to delegate user credential when using services. CANFAR users are also resisted to a Group Management Service (GMS) that is charge of authorization procedures. GMS is queried by CANFAR services to verify user capabilities and capacities.

The focus of phase one of the roadmap is to implement an interoperable GMS service on top of the EGI FedCloud at INAF and to achieve group membership resolution from CANFAR to EGI GMS.

A new implementation of the GMS must be developed and deployed at INAF site. Once the GMS is active EGI FedCloud users may reference EGI FedCloud group, capabilities and capacities on CANFAR services for access control. In a second step of the implementation 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 CDP to achieve VO membership resolution from EGI FedCloud to CANFAR storage and computing services. EGI FedCloud users must delegate their certificate to CANFAR CDP before using CANFAR services (e.g. VOSpace)

In the last step of this activity we will analyse the use of eduGAIN identity federation.

The service we will use for testing and verification will be CANFAR VOSpace.

Table 1 – Technology gap analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SERVICE**  | ***Type*** | ***CANFAR*** | ***EGI*** | ***Activity*** |
| **GMS** | Authorization | Implement new capabilities to allow users from other GMS to access CANFAR resources | Implement a new GMS for EGI  | Development  |
| **CDP** | Authentication | Implement VOMS extensions | use the CANFAR CDS | Development  |
| **VOSpace** | Data  | Allow users from two or more different GMS to access |  | Development |
| **Transfer Service** | Data | Allow users from two or more different GMS to access |  | Development |
| **VOMS** | Authorization  |  | VOMS backend for GMS  | Feasibility study |
| **Identity Federation**  | Authentication | Use of eduGAIN | Feasibility study |

At the end of Phase one we will test the implementation: one EGI FedCloud registered user will access CANFAR VOSpace, upload/download data, grant access to her data to other users registerd both in CANAF and EGI.

This phase will last for 8 months, we will implement two cycles of 4 months with verification. At the end of each cycle we verify both status of activity and eventual correction in the roadmap due to new technology implementations both at EGI and CANFAR.

## Phase 2: Interoperable VOSpace on EGI fedcloud

This activity focuses on the implementation of a federated virtual storage for EGI FedCloud and CANFAR and it will:

* Enable users to seamless access, move and process data on CANFAR and EGI FedCloud
* Enable an data platform for A&A community and projects based on the reference standard in Astronomy (VOSpace)
* Verify the possibility to implement a VOSpace implementation for the EGI Open Data platform based on oneData.

This activity will be done in collaboration with other EGI-Engage WPs, in particular [OPEN DATA PLATFORM] that will provide support to study a possible integration of VOSpace with EGI Open Data platform.

The final goal of this activity is to develop and deploy a fully functioning VOSpace implementation on top of EGI FedCloud that can reference both CANFAR and INAF GMS for access control.

The implementation of a VOSpace in EGI FedCloud requires to develop two core services: the VOSpace file catalogue and the Transfer service. The user credential must be delegated using a credential delegation protocol for X509 certificates. Once these 3 services will be in place both a CANFAR user and an EGI FedCloud user will be able to manage files in the EGI VOSpace and to grant access to their files also to other CANAFAR or EGI users.

In a second step of the implementation we will analyse the use EGI Open Data platform based on oneData as a VOSpace backend taking into account VOSpace and oneData requirements in terms of capabilities. The scope of this analysis is to verify the possibility of an astronomer to access her data from both VOSpace and oneData interfaces.

We will provide at least one implementation of the VOSpace at INAF during the fist step but at least 2 implementations will be provided for the end of the project.

Table 2 – Technology gap analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SERVICE**  | ***Type*** | ***CANFAR*** | ***EGI*** | ***Activity*** |
| **GMS** | Authorization | Add VOMS extentions | Uses GMS developed at phase 1 | Development  |
| **CDP** | Authentication |  | Develop a new CDS | Development  |
| **VOSpace** | Data  |  | Develop VOSpace metadata catalogue | Development |
| **Transfer Service** | Data |  | Develop a filesystem backend for VOSpace transfers | Development |
| **Open Data platform**  | Data | oneData backend for VOSpace  | Feasibility study |

This activity will last for 10 months divided in two cycles of 5 months each. At the end of each cycle we will verify the development. We verify both progress status of the activity and we eventually update the roadmap due to the availability of new technology for EGI FedCloud or CANFAR.

At the end of this phase we will perform a final test. A user will be able to access EGI FedCloud VOSpace upload, and manage his data and grant access to her data both to CANFAR and EGI users.

## Phase 3: Server-to-server transfers

 The goal of this phase is to enable user-initialed server-to-server transfers, as per the VOSpace specification, between CANFAR and EGI.

Will last for 6 months, two cycles of 3 months with verification. We verify both status of activity and eventual correction in the roadmap due to the implementation of new technology at EGI fedcloud or CANFAR.

Final test will be service to service data transfer from Canada to Europe

## Roadmap implementation

In the table each activity is identified with a number (phase.cycle.activity), a short descriptive name, a start date and release date, the status and eventual dependencies with other activities.

Table 1 – Roadmap

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Activity**  | ***Name***  | ***Start date*** | ***Release date*** | ***type*** | ***Dependencies*** |
| **1.1.1** | EGI GMS implementation | 9/15 | 2/16  | Development  |  |
| **1.1.2** | Verification | 2/16 | 2/16 | test | 1.1.1 |
| **1.2.1** | EGI GMS implementation  | 2/16 | 7/16 | Development  | 1.1.1 |
| **1.2.2** | VOMS and GMS analysis  | 2/16 | 4/16 | study | 1.1.1 |
| **1.2.3** | CANFAR VOMS certificate  | 2/16 | 6/16 | Development  |  |
| **1.2.4** | Identity Federation | 4/16 | 6/16 | study |  |
| **1.2.5** | Verification | 6/16 | 6/16 | test | 1.2.1 |
| **1.2.6** | Final test | 6/16 | 6/16 | test | 1.2.1, 1.2.5 |
| **2.1.1** | VOSpace metadata interface | 7/16 | 12/16 | Development  | 1.2.1 |
| **2.1.2** | CDS VOMS extensions | 7/16 | 10/16 | Development  | 1.2.3 |
| **2.1.3** | VOSpace file transfer | 9/16 | 12/16 | Development |  |
| **2.1.4** | CDC implementation | 10/16 | 12/16 | Development |  |
| **2.1.5** | Verification  | 1/17 | 1/17 |  | 2.1.1, 2.1.2, 2.1.3, 2.1.4 |
| **2.2.1** | VOSpace and transfer service | 2/17 | 6/17 | Development  | 2.1.1, 2.1.2, 2.1.3, 2.1.4 |
| **2.2.2** | VOSpace oneData analysis  | 4/17 | 7/17 | study |  |
| **2.2.3** | Verification  | 7/17 | 7/17 |  | 2.2.1 2.2.2 |
| **2.3** | test | 7/17 | 7/17 | test |  |

HERE a GANTT

# Open data policies

Here we discuss the policy to share data [input from Tiziana]

# Partners involved in the Federation activity

The **Canadian Astronomy Data Centre** (CADC)[[5]](#footnote-5) 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.

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[[6]](#footnote-6). 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[[7]](#footnote-7)**) 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.

# References

|  |  |
| --- | --- |
| ***No*** | ***Description/Link*** |
| R1 | Description<http://example.com>  |
| R2 |  |
|  |  |
|  |  |
|  |  |
|  |  |

1. Appendix example
1. <http://www.canfar.phys.uvic.ca/canfar/> [↑](#footnote-ref-1)
2. Census of European DataCentres <http://goo.gl/zOWykc> [↑](#footnote-ref-2)
3. EuroVO- CoSADIE Project http://www.eurovo.org [↑](#footnote-ref-3)
4. <http://www.canfar.phys.uvic.ca/canfar/> [↑](#footnote-ref-4)
5. <http://www.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/> [↑](#footnote-ref-5)
6. <http://www.nrc-cnrc.gc.ca/eng/rd/nsi/> [↑](#footnote-ref-6)
7. http://www.inaf.it [↑](#footnote-ref-7)