

Access EGI resources through the ESCAPE developed ESFRI Science Analysis Platform

S. Bertocco, M. Parra-Royón, D. Morris, K. Kliffen, F. Tinarelli, M. Stagni, V. Galluzzi
F. Bedosti, G. Taffoni, M. Molinaro, J. Swinbank, S. Sanchez Exposito

ABSTRACT

The EU ESCAPE project is developing ESAP, ESFRI's scientific analysis platform, as an API gateway that enables the seamless integration of independent services accessing distributed data and computing resources. At ESCAPE we are exploring the possibility of exploiting OpenStack EGI's cloud computing services through ESAP. As a use case, we are considering one of the studies known as Data Challenges used to prepare the community to work with the data to be generated by the Square Kilometer Array (SKA).

In our contribution, we describe the technical steps performed: we registered to the Virtual Organisation `vo.access.egi.eu` to count on the necessary development and test resources and we automated the creation of a Virtual Machine through the EGI `fedcloud` client. We automated the installation on a cloud virtual machine instance of the suitable analysis software through a software framework developed at IRA-INAF, called `ira-init`. We plan to provide ESAP's users with resources access writing an ESAP connector.

In this first prototype data access is simplified through NFS mounted storage or a cloud data volume. Data transfer tests are being conducted using `storm-webdav` to provide users with the ability to analyze the data both locally and remotely.

The ESCAPE EU project



ESCAPE stands for European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures. The project aims to address the Open Science challenges shared by ESFRI facilities (SKA, CTA, KM3Net, EST, ELT, HL-LHC, FAIR) and pan-European research infrastructures (CERN, ESO, JIVE) in astronomy and particle physics.

Our interest is focused on two of the main actions of the project:

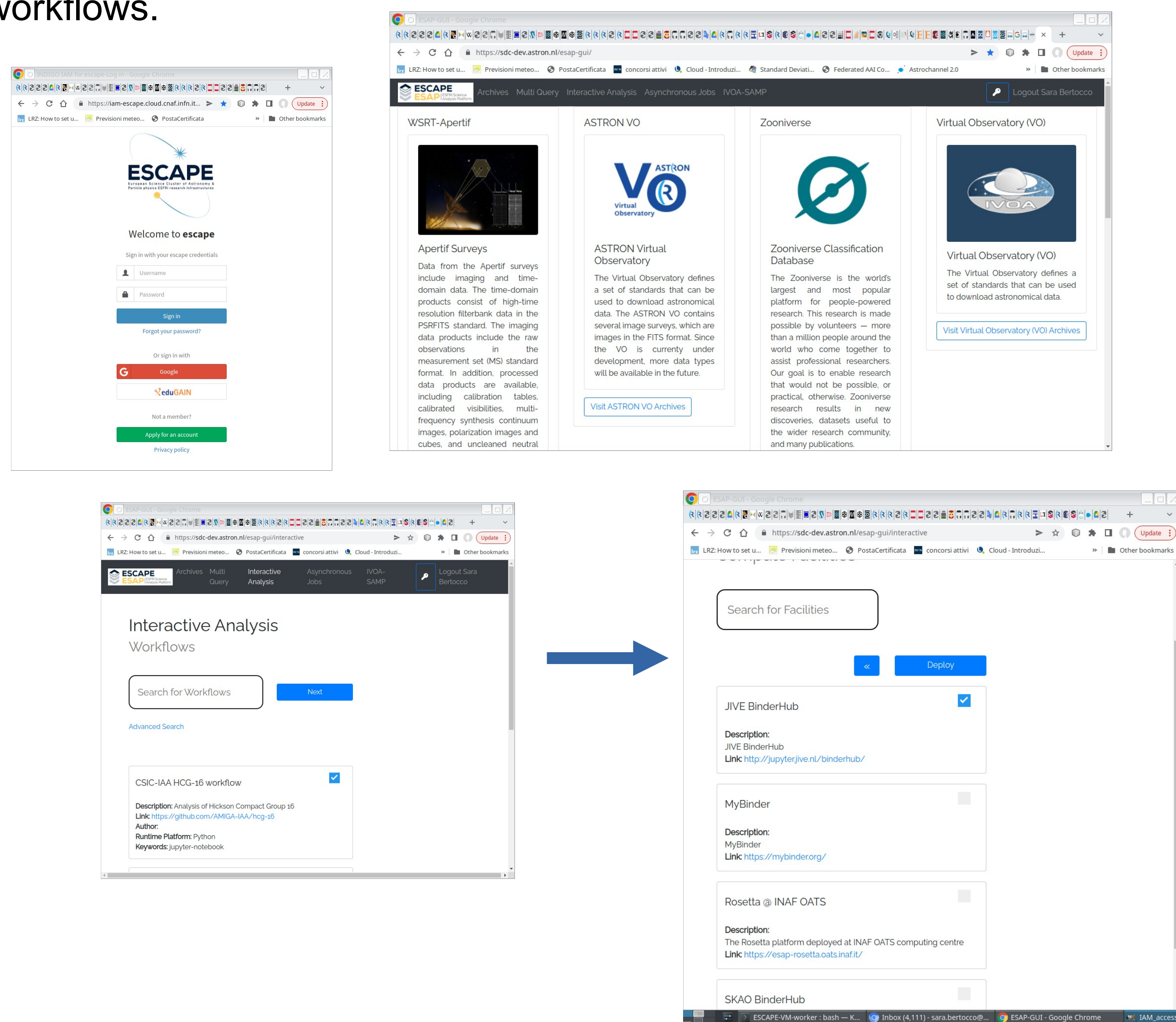
- connect ESFRI projects to EOSC ensuring integration of data and tools
- establish interoperability within EOSC as an integrated multi-messenger facility for fundamental science.

The ESAP platform

ESAP stands for ESFRI Science Analysis Platform.

ESAP is a flexible science platform for the analysis of open access data available through the EOSC environment. It allows EOSC researchers

- to identify and stage existing data collections for analysis,
- to select among software tools and packages developed by the ESFRI or bring their own custom workflows to the platform,
- to exploit the underlying computing infrastructure to execute those workflows.



Researchers can select datasets from different archives.

A set of workflows is provided, the researchers can select the workflow of interest and then choose to deploy it in a suitable infrastructure.

The EGI federated cloud



The EGI Core is a federation and management platform that pool together various resources. The EGI Federated Cloud Platform (FedCloud) is a federated cloud infrastructure.

The EGI cloud compute service gives users the ability to deploy and scale virtual machines on-demand.

The use cases

The workflows provided in ESAP are relative to use cases of interest for the ESFRI and big research infrastructures users.

We considered specifically:

- To provide IVOA services in a container
<https://github.com/zarquan/Oligia>
- To provide an environment to run the workflow to process HI data cubes produced by radio interferometers, in particular large data cubes produced by future instruments like the SKA.
See: HI-FRIENDS

<https://hi-friends-sdc2.readthedocs.io/en/latest/>
<https://github.com/HI-FRIENDS-SDC2/hi-friends>

EGI `fedcloud` is a suitable infrastructure to satisfy such use cases

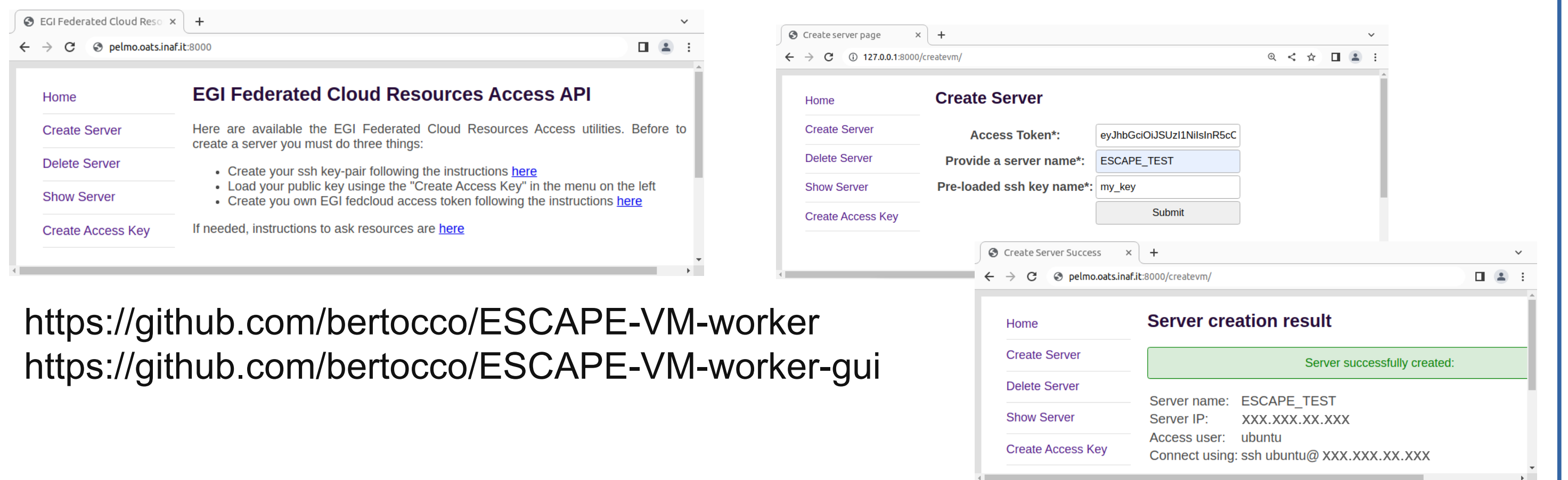
The integration work

The integration work is done through different steps:

- we registered to the Virtual Organisation `vo.access.egi.eu`
- we ask to CESGA site support to be able to allocate needed resources (particularly a public IP) following the EGI provided guide (<https://docs.egi.eu/users/getting-started>)

and using the EGI federated cloud python client (<https://github.com/tdviet/fedcloudclient>)

we wrote a simple django web application, to be integrated in ESAP, to automate the virtual machine creation process in the EGI `fedcloud`.



<https://github.com/bertocco/ESCAPE-VM-worker>
<https://github.com/bertocco/ESCAPE-VM-worker-gui>

The lesson learned

We have identified some show stopper in the virtual machine automation process:

1) Public network. Each site has a slightly different network configuration, and there is no standard way to tell which one to use and how to setup the router for it. There is not a convention for the Networks naming → need some guess (human intervention) to understand which is a public network.

2) Supported Virtual Organizations. There is not a command in `openstack` or in the `Fedcloud` tool that lists the sites that support a certain virtual organisation.

3) Flavours and images: it is needed to guess (human intervention) the resources associated to the flavors or the Operative System in an image. A naming convention could be useful. Possible solution for flavors: to use 'list' and 'show' commands in association to create a configuration information file.

ESCAPE & EOSC Integration

We worked in the scope of WP4 (Connecting ESFRI projects to EOSC through VO framework) and WP5 (ESFRI Science Analysis Platform) of the ESCAPE (European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures) European H2020 project to integrate IVOA compliant Virtual Observatory (VO) services within the EOSC hybrid cloud scenario, to test containerization of VO aware applications, to provide run pipeline for the SKA data challenges.

Acknowledgment to CESGA that provides resources for our tests.

ESCAPE & EOSC related content of this poster benefits support from the ESCAPE project, funded by the European Commission, Horizon2020 programme (grant n. 824064).



Contact:
sara.bertocco@inaf.it

