The User Support Programme And The Training Infrastructure Of The EGI Federated Cloud

Enol Fernandez

EGI.eu, Amsterdam, NL
IFCA (CSIC-UC), Santander, Spain
enol.fernandez@egi.eu

Gergely Sipos

EGI.eu, Amsterdam, NL
MTA SZTAKI Budapest, HU
gergely.sipos@egi.eu

Diego Scardaci

EGI.eu, Amsterdam, NL
INFN, Catania, Italy
diego.scardaci@egi.eu

David C.H. Wallom

University of Oxford, Oxford, UK
david.wallom@oerc.ox.ac.uk

Yin Chen

EGI.eu, Amsterdam, NL
yin.chen@egi.eu

Abstract—The EGI Federated Cloud is a standards-based, open cloud system as well as its enabling technologies that federates institutional clouds to offer a scalable computing platform for data and/or compute driven applications and services. The EGI Federated Cloud is based on open standards and open source Cloud Management Frameworks and offers to its users IaaS, PaaS and SaaS capabilities and interfaces tuned towards the needs of users in research and education. The federation enables scientific data, workloads, simulations and services to span across multiple administrative locations, allowing researchers and educators to access and exploit the distributed resources as an integrated system. The EGI Federated Cloud collaboration established a user support model and a training infrastructure to raise visibility of this service within European scientific communities with the overarching goal to increase adoption and, ultimately increase the usage of e-infrastructures for the benefit of the whole European Research Area. The paper describes this scalable user support and training infrastructure models. The training infrastructure is built on top of the production sites to reduce costs and increase its sustainability. Appropriate design solutions were implemented to reduce the security risks due to the cohabitation of production and training resources on the same sites. The EGI Federated Cloud educational program foresees different kind of training events from basic tutorials to spread the knowledge of this new infrastructure to events devoted to specific scientific disciplines teaching how to use tools already integrated in the infrastructure with the assistance of experts identified in the EGI community. The main success metric of this educational program is the number of researchers willing to try the Federated Cloud, which are steered into the EGI world by the EGI Federated Cloud Support Team through a formal process that brings them from the initial tests to fully exploit the production resources.

Keywords— Cloud Computing; e-infrastructure; Federation of Resources; support; training; tutorial.

I. Introduction

Science today is no longer exclusively produced in single research labs or within national boundaries. Modern scientific challenges call for integrated solutions, cross-country collaborations and computing power with flexible usage to analyse vast amounts of data. E-infrastructures allow scientists to share information securely, analyse data efficiently and collaborate with colleagues worldwide.

EGI operates one of the largest, collaborative e-infrastructures in the world. EGI supports the digital European Research Area (ERA) through this pan-European infrastructure, its innovative technological building blocks, and related support teams and networks for users. These all together offer reliable ICT services, which provide uniform, cost effective, user oriented and collaborative access to computing and data storage resources in more than 30 countries, from EGI's National Grid Infrastructures (NGIs). EGI's mission is to help scientists to make the most of the latest computing technologies, such as clouds, big data and grids by facilitating interactions between them and the NGIs.

The EGI community started developing a new type of infrastructure, the EGI Federated Cloud in 2011. The system reached production level in 2014. The EGI Federated Cloud (FedCloud) [1] is a standards-based, open cloud system as well as its enabling technologies that federate institutional clouds to offer a scalable computing platform for data and/or compute driven applications and services. The EGI Federated Cloud is already deployed on nearly 20 academic institutes across Europe who together offer 6000 CPU cores and 300 TB storage for researchers in academia and industry. This capacity is available for free at the point of access through IaaS capabilities and interfaces that are tuned towards the needs of technologists

from research and education. These technologists can define high-level platforms and environments – with the cloud terminology PaaS and SaaS systems - on top of the EGI IaaS cloud. The technologies that enable the EGI cloud federation are developed and maintained by the EGI community, and integrate open standards and open source Cloud Management Frameworks. These integrated technologies are available for institutes and communities who wish to setup their own federated cloud infrastructures.

The EGI Federated Cloud integrates the core capabilities of individual cloud deployments to enable workloads, simulations and services that may span across multiple administrative locations. The federation provides harmonised views across the individual cloud instances for end users, who are typically platform developers (PaaS, SaaS), scientific programmers, community/project coordinators and system administrators. The EGI Federated Cloud currently supports 26 scientific communities and 50 use cases coming from different scientific disciplines: bioinformatics, physics, earth sciences, basic medicine, arts, language and architecture, mathematics, computer sciences, etc. Furthermore, between 2015-2017 several research infrastructures from the ESFRI roadmap [2] (BBMRI, EPOS, ELIXIR, DARIAH, EISCAT-3D, INSTRUCT and LifeWatch) will define and implement community-specific capabilities on this platform in the recently started H2020 EGI-Engage project [3]. The adoption of the EGI Federated Cloud is ongoing within industry too, through early adopter SMEs from Spain, France and the UK.

The EGI Federated Cloud was launched in production in May 2014 and can be seen as a new capability offered by the EGI infrastructure, which considerably widens the usage of existing, academic Distributed Computing Infrastructures. Now, web services and interactive applications can be easily integrated in the infrastructure, the computing environments can be finely tuned to satisfy user's needs in term of software (OSs and software packages) and hardware (number of cores, amount of RAM, etc.). These new opportunities offered by EGI hugely extended the potential user base of the infrastructure opening the doors to new research communities with minimal or none knowledge of distributed computing. The need for an extensive user support and training programme, which

facilitates the uptake of the new infrastructure within the NGIs and scientific communities, was imminent. The aim of this was to (1) increase the NGIs and scientific communities' awareness about the new technology and infrastructure, (2) raise interest of scientific communities that do not know the EGI infrastructure or even distributed/cloud computing yet and, (3) inform current and existing EGI users (grid users) about the new possibilities and support them migrate to more appropriate EGI usage models offered by the FedCloud. The details of the user support programme and the training infrastructure are later described in the paper.

This paper is organised as follows: section II describes related work on training programs of other similar cloud infrastructures. A description of the architecture of the EGI Federated Cloud is illustrated in section III. Section IV introduces the FedCloud training infrastructure and its educational program. Section V describes how new potential users identified during a training events are supported by EGI with dedicated technical consultancy and customised assistance until they are able to fully exploit the production environment. In section VI and VII, plans for the future and conclusions are depicted.

II. RELATED WORK

Many cloud federations were designed and deployed in Europe in the last years, but the two that have most similarities with the EGI Federated Cloud are Helix Nebula [4] and CONTRAIL [5], which both targeted multiple European institutions, and investigated how to provide access to multiple different resource providers by means of a unified interface.

Initially funded through the EC 7th Framework program (EC FP7), Helix Nebula is an on going activity, started in June 2012. Its models is based on the so called 'blue box', a single central broker deployed to support all the proprietary interfaces and cloud instance formats offered by the federation. Currently, the 'blue box' is a commercial technology based on SixSq Slipstream [6] exposing a web interface alongside a unique proprietary interface to provide also API access to the broker.

The Helix Nebula Marketplace (HNX) training program is based on courses to teach how to use and fully exploit the unique interface to access the different cloud providers
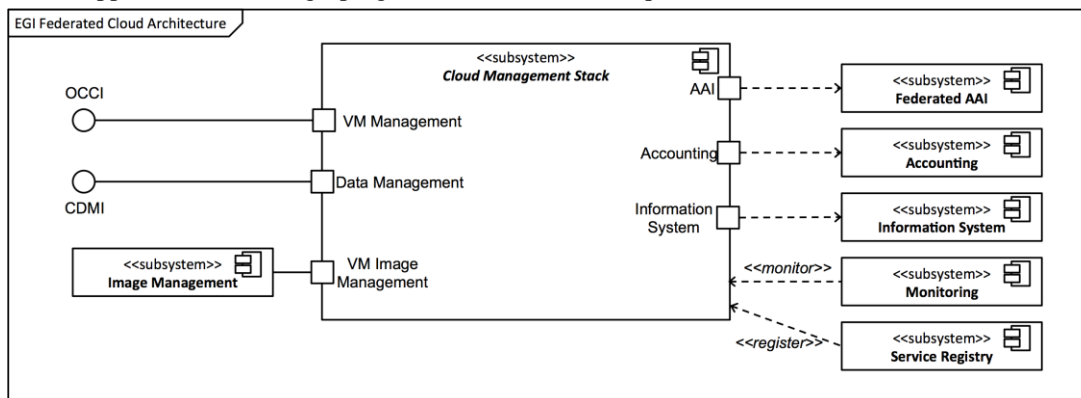

Figure 1. Technical architecture of the EGI Federated Cloud

connected to it through the SlipStream broker. The training events are organised by SixSq, the SME developing SlipStream and running HNX. At the moment, only paid trainings are available, therefore user communities potentially interested should consider that the cost should be added on top of the cost to access the connected IaaS cloud providers.

The CONTRAIL project, started in 2010 and concluded in 2014, developed a full-fledged cloud federation software stack offering identity management, federated Service Level Agreements (SLA), a dedicated cloud file system, a Platform-as-a-Service (PaaS) layer, and a unified administrative layer. The development of a complete new set of software tools required an extensive training programme [7]. The CONTRAIL community set up a modular design for the training components where each training module closely followed the respective CONTRAIL software component. The training modules are available on the web through slides, hands-on and video lessons. In addition, CONTRAIL organised three annual schools designed for junior researchers in computer systems willing to expand their knowledge in the area of Cloud computing.

III. THE EGI FEDERATED CLOUD

The EGI Federated Cloud architecture is developed from high-level capabilities extracted from users' requirements taking also into account the needs and expertise of the existing heterogeneous cloud management software locally installed at EGI resource providers. The architecture tackles this by considering each local Cloud deployment as an autonomous and abstract subsystem that integrates with the federation through well-defined interfaces (see Figure 1). Open standards are employed for these interfaces wherever possible. Where this was not feasible, community-accepted non-standardised solutions are used. It is the responsibility of resource providers to identify and deploy the solution that fits best their individual needs whilst ensuring that the offered services implement the required interfaces.

The architecture defines cloud specific capabilities and interfaces: VM Management and Block storage management, provided with Open Cloud Computing Interface (OCCI) [8][9][10]; Data Management, provided with Cloud Data Management Interface (CDMI) [11]; and the integration with the Image Management subsystem of EGI, provided with the Open Virtualization Format (OVF) [12] and HEPiX image lists format [13]. The architecture also defines interaction ports with a number of services in the EGI Core Infrastructure [14]: Federated AAI, Accounting, Information System, Monitoring, and Service Registry. The IaaS Cloud capabilities, defined in the architecture, are integrated with the Image Management subsystem, provided as part of the Federated Cloud infrastructure.

The EGI Federated Cloud currently integrates resources from realisations of the abstract cloud management stack for OpenNebula, OpenStack and Synnefo. The EGI Federated Cloud Task gives Resource Providers a platform to share their implementation solutions and build communities around the deployed Cloud Management Frameworks.

A. Cloud-specific capabilities and interfaces

To be fully compliant with the EGI Cloud Platform resource providers have to fully implement all of the following services or interfaces.

1) VM Management

The Open Cloud Computing Interface (OCCI) is a RESTful Protocol and API designed to facilitate interoperable access to, and query of, cloud-based resources across multiple resource providers and heterogeneous environments. The EGI Federated Cloud resource providers implement the Infrastructure specification of OCCI in version 1.1 with extensions to support contextualisation and the use of a consistent nomenclature for the resource templates (also known as "flavors").

Contextualisation is supported by combining the use of cloud-init [15] -a de-facto standard- from within the VM to retrieve contextualisation data, and a FedCloud specific OCCI extension that allows the user to define the data that is to be presented to cloud-init for the contextualisation.

OCCI is exposed through dedicated add-on components for the Cloud Management stacks. OCCI-OS [16] for OpenStack, snf-occi [17] for Synnefo and rOCCI[18] for OpenNebula. rOCCI is designed to support multiple back-ends and can be used for providing OCCI in CloudStack and AWS.

2) Data Management

Data management is enabled with the support of the Cloud Data Management Interface (CDMI), which defines a RESTful open standard for operations on storage objects. CDMI offers clients a way to operate both on a storage management system (i.e. data containers) and on single data items within a given data container. CDMI is built around the concept of objects and capabilities, allowing for partial implementations, where the exact level of support for each capability depends on the concrete implementation and is exposed to the client as part of the protocol.

In the EGI Federated Cloud environment, the primary goal of CDMI is to offer a standard interface for operating with blob data. Therefore the most relevant capabilities in this context are the CRUD (Create, Read, Update, Delete) operations on containers and data items. Currently, the reference implementation provided by Synnefo is utilized within the EGI federated cloud to support the standard.

3) Image Management

In a distributed, federated Cloud infrastructure, users need mechanisms to efficiently managing and distributing their VM Images across multiple and heterogeneous Cloud resource providers. The VM Image management subsystem enables users to register new VM Images and appliances for automatic distribution over the federated resource providers. The EGI Applications Database (AppDB) service [19] and HEPiX image lists subscribers at the Resource Providers provide the functionality of this subsystem.

AppDB provides a web interface for the user to register and manage VM image metadata and to endorse those VM images as part of VO image lists. The endorsement triggers the subscribers at the providers that, examine and cryptographically verify the provided information, and, pending their decision,

pool the new or updated VM Image locally for instantiation. Images are distributed using the DMTF Open Virtualization Format (OVF) standard packaging format which (if needed) subscribers convert to a suitable format supported by the Cloud Management stack.

B. Integration with the EGI Core Infrastructure

The resource providers must integrate with EGI core services, which are not cloud-specific but are needed for a successful operation of a production infrastructure.

1) AAI

From the user's perspective, identity federation is one of the core parts of any interoperable infrastructure (even when the users are not aware of that federation). If identity federation does not exist, users cannot access the infrastructure transparently, as they would need to be concerned with the specific details on how to access any given service provider. EGI controls access to resources employing personal X.509 certificates and the concept of "Virtual Organisation" (VO). A VOMS [20] server acts as an Attribute Authority that issue VO attributes which are contained in an augmented proxy certificate signed with the user's personal certificate. The proxy and the different attributes determine at the resource provider level if the user is granted to access to the resources. EGI Federated Cloud uses PERUN [21] in OpenNebula and Keystone-VOMS [22] in OpenStack for the user management process.

2) Accounting

Usage of resources must be recorded and integrated in the EGI Core Accounting system. EGI Federated Cloud has agreed on a Cloud Usage Record (UR) -which inherits from the OGF Usage Record [23]- that defines the data that resource providers must send to EGI's central Accounting repository. A local harvester generates those records from the proprietary data sources of the Cloud management stacks. UR data is encrypted with the Accounting DB's public key and signed with the local configured host certificate's private key, and sent as text payload in secure messages. A portal allows users, operators and administrators to access accounting data in a structured way.

3) Information Discovery

Users and tools can discover the available resource in the infrastructure by querying EGI information discovery services. The common information system deployed at EGI is based on the Berkeley Database Information Index (BDII) with a hierarchical structure distributed over the whole infrastructure. Each resource provider publishes information through the cloud BDII information provider [24] including the available endpoints and their characteristics. AppDB queries the EGI Information service to display the Resource Providers providing specific VM images and the usage details for instantiating them. This information is also used for operational purposes. For example, the Monitoring subsystem depends on certain data to be present in order to properly operate.

4) Monitoring

Services in the EGI infrastructure are monitored via SAM (Service Availability Monitoring). Specific probes to check functionality and availability of services must be provided by service developers to ensure both user satisfaction and compliance with the Operational Level Agreements (OLA) that are put in place once a federated resource provider becomes operational in the infrastructure. Cloud services are monitored in a variety of ways: (i) Basic "ping" probes to check service reachability (ii) VM Management probe to check OCCI compliance, and (iii) Accounting probe to check consistency and freshness of accounting data. CDMI and VM Image Management probes are under development.

5) Service Registry

EGI's central service catalogue is used to catalogue the static information of the production infrastructure topology. A set of service types are defined in the registry to allow Resource Providers to expose Cloud resources to the production infrastructure. The catalogue also tracks information for scheduled downtimes, which is important for accurate and correct reliability calculation performed by the Monitoring subsystem.

C. PaaS and SaaS services in the EGI Federated Cloud

The EGI Federated Cloud offers to its users a set of PaaS and SaaS deployed on top of the IaaS capabilities. These services are developed by external contributors, which decided to interface them with the FedCloud to exploit its resources.

They provide users with an alternative way to access the FedCloud, hiding the IaaS layer and making simpler the infrastructure usage.

Table 1 and Table 2 list, respectively, PaaS and SaaS currently available in the FedCloud. This set will be further extended in the next future as described in the Future Works section.

Table 1. PaaS currently available in the EGI Federated Cloud.

PaaS	Description
Catania Science Gateway Framework [25]	Visual interface for VMs creation and deployment on different cloud providers based on OCCI standard
Slipstream	VM image creation and deployment, contextualisation, VM cluster automated deployments, orchestration.
Vcycle [26]	VM lifecycle manager

Table 2. SaaS currently available in the EGI Federated Cloud.

SaaS	Description
Catania Science Gateway Framework	It provides users with intuitive web interface to execute applications on the Cloud as jobs and to manage these jobs during their running
COMPSs [27]	VM image creation and deployment, contextualisation, VM cluster automated deployments, orchestration.
VMDIRAC [28]	VM lifecycle manager
WS-PGRADE [29]	Job, parameter study and workflow submission tool to clouds

IV. THE TRAINING INFRASTRUCTURE AND THE EGI FEDERATED CLOUD EDUCATIONAL PROGRAM

The EGI Federated Cloud training infrastructure is built on top of the cloud production sites. Training resources are offered to students and isolated from the production environment through an ad-hoc virtual organisation used only for educational

aims. This VO is mapped to a special tenant on the cloud sites with a limited set of assigned resources. In this way, cloud resource providers, willing to offer resources to the training infrastructure, have only to create a tenant, assign to it a quota of resources and map the training VO to the new tenant. This solution removes the need to create dedicated sites for training, which would mean deploying and maintaining a parallel infrastructure, although of minor size. In this way, the cost to keep up & running a training infrastructure are greatly reduced, making it more sustainable.

In any case, the cohabitation in the same cloud sites of production and training resources implies the need of a solid security system, which prevents unintended interactions between the two environments and the exploit of any security flaws that could deteriorate the availability and reliability of the production infrastructure. For this reason, the tenant associated to the training VO is configured on the cloud sites to be completely isolated by the other tenants. Moreover, we decided to grant students the training infrastructure access through short-term robot certificates [30]. A robot certificate is created for each training event, it is registered on the training VO and its lifetime is limited to the tutorial duration. Students cannot directly access the robot certificates, which are securely stored and managed by the eToken server, a standard-based service for central provisioning of robot credentials [31]. They can get credentials to access the EGI Federated Cloud training infrastructure querying the e-Token server through a simple REST interface that allows creating and retrieving short-term proxies from a robot certificate, valid only for few hours. These proxies can be used as security token to access the Federated Cloud resources. As additional security mechanism, these proxies contain a unique user id, associated to each student during the registration phase. A proxy containing this information is named Per-User Sub-Proxy [32] and allows the full traceability of the user operations on the infrastructure. The standard security monitoring system is also enabled on the training infrastructure identifying any suspicious activity during the VMs run-time.

Beyond the security issues, this mechanism provides students with a simple access to the EGI Federated Cloud. In this way, students without a Grid certificate can really test the infrastructure without the need to complete the complex procedure to obtain the certificate before the tutorial. Furthermore, in this way, the training event can focus its agenda on core cloud topics and not on the details of the pre-requisites to access the infrastructure.

A. Training events

The EGI Federated Cloud training program includes different types of events from basic tutorial highlighting the main features offered by the FedCloud, with the aim to spread the knowledge of our infrastructure in the scientific world, to tutorials for specific disciplines teaching how to use an application set already available in the EGI infrastructure.

In this section, three different types of training events are described:

- Basic tutorials: to introduce the EGI Federated Cloud describing its main services and capabilities;

- Tutorials to introduce high level tools and platform: each of these events is dedicated to a specific high level tool or platform integrated on top of the FedCloud IaaS;

- Community oriented training events: specialised training for a scientific discipline or specific services.

The target audience of these events is mainly composed by scientific communities, single researchers and IT technicians willing to learn more on Cloud.

Each tutorial will be integrated with an additional session where the instructions to join the EGI community and to exploit the Federated Cloud capabilities are described. An EGI certification and verification process guarantees for the quality of the training events as described below.

1) Basic Tutorials

The tutorial will train attendees on how to obtain access and how to use the cloud services (IaaS, PaaS and SaaS) offered by the EGI Federated Cloud:

- The tutorial begins with a short introduction part that describes the main capabilities of the EGI Federated Cloud and the main steps of the typical user workflow.

- In the second part, introductive information on the different functions of the cloud infrastructure are given. The features are described with the aim to highlight the benefit for the end users:

  - VM image preparation: how to create an image that can be used in the Federated Cloud;

  - VM image management: registration in the cloud marketplace, image endorsement work-flow and deployment on the cloud sites;

  - short introduction on the EGI core services: monitoring, information system, accounting, etc.

  - CLI and/or API to access to the IaaS layer;

  - high level access through PaaS services;

  - graphical access exercises through SaaS portals.

According to the duration of the event and the level of expertise of the attendees, some of these items can be deeply analysed or skipped.

- The third part will introduce strategies of integrating new platforms and applications with the EGI Federated Cloud platform: different solutions are depicted highlighting advantages and disadvantages and for what kind of services are recommended.

These events are directly organised by the EGI FedCloud user support team that guarantees for their quality, which should be in line with the EGI standard.

2) Tutorials to introduce high level tools and platforms

The EGI Federated Cloud offers to its users a set of high level tools, integrated on top of the IaaS interface, that provide

users with PaaS and SaaS services [33]. Each tutorial of this category is focused on a specific tool and its format strictly depends on the technology introduced. Basic information on how to use the high-level services with real examples are provided by the users. Experts and developers of the tools will compose the team of tutors.

Since external contributors developed most of these tools, the EGI FedCloud user support team certifies the quality of the training events through a verification process that includes inspection of (1) the training material and (2) the tutorial agenda. Furthermore, the support team provides trainers with guidelines to organise the events that should be followed to obtain the EGI certification.

3) Community oriented training events

These events are specialised training for a scientific discipline or specific services aiming to provide students with information on services and applications already available in the EGI infrastructure for this community. The target of the tutorial is teaching how to use these tools, describing the steps to fully exploit their integration in the EGI Federated Cloud. Cloud technical details are not covered. The team of tutors will be composed mainly by experts in the same discipline covered by the event. This kind of events should also satisfy the quality process above described to be included in the official EGI training calendar.

B. Next steps to become an EGI user

The final aim of each training event is promoting the EGI infrastructure in the scientific world and invite new research communities to join the EGI infrastructure. Considering the complexity and the several alternatives offered to the user to become part of the EGI eco-system, each training event is completed with a description of the main steps to become an EGI users giving to each scientist the opportunity to follow the way that better satisfies his needs.

In the Table 3, the actions that should be done by the users to join the EGI community are described.

Table 3. How to join the EGI community.

How to join the EGI community	User's actions
User wants to access resources	• User can join one of the existing groups/VOs • User can find relevant VO in the Ops Portal [34]
User wants to request an allocation on the EGI resource providers	• User wants dedicated resources • User should submit request via the EGI Resource Allocation tool, e-Grant [35]
A community wants to join the FedCloud as resource provider	• A community wants to turn his site into a community federate cloud; • The community should contact EGI Operations to have further information.

V. SUPPORT USERS FROM THE TRAINING TO THE PRODUCTION ENVIRONMENT

The main objective of each training event is spreading the knowledge of the EGI Federated Cloud in the scientific communities and increasing the number of researchers that take advantage of its services. The main success metric of these events is the number of researchers willing to test the FedCloud to verify if it can help them in their daily work. Once these possible new users are identified, the EGI Federated Cloud Support Team takes care of them with a dedicated technical consultancy and assistance customised for each community.

The EGI Federated Cloud support team consists of members of EGI.eu [36] and representatives of technical expert coming from the NGIs [37] belonging to the EGI collaboration.

When a new scientific community expresses interest on the EGI FedCloud, the support team organises a face-to-face or web meeting to gather requirements, identify the better strategies to integrate community applications and services and define a complete work-plan. A team of technical experts is assigned to the communities to assist them during the activities. Furthermore, EGI offers to the new potential users access to its whole documentation [38] containing guide describing the cloud services, tutorials that guide users to better appreciate the Cloud paradigm and defining step-to-step procedures to understand how to maximise value through using the Cloud. EGI has been also documenting the application models of successful use cases so to facilitate the ease of adoption by future users [39]. Moreover, a VM set, secure endorsed by the EGI CSIRT [40], is available in the Cloud Marketplace [41] to speed-up the testing work. These VMs are already configured to exploit the main FedCloud capabilities are the contextualisation and span both basic OS images, where users' application can be installed, and image with pre-installed common applications and services.

To allow scientists to immediately try and evaluate the Cloud in an easy way, we decided to create a catchall virtual organisation (VO), acting as an incubator for any new use cases, reducing the time between the first contact with a community and their first test of the cloud e-Infrastructure. This catchall VO has been enabled in all certified sites, to engage all resource providers in this effort. The stated mission of the VO is to support application prototyping and validation for up to six months. The catchall VO greatly lowered the barrier to entry by removing the need for a large administrative setup process. This is particularly useful for those prospective user communities that mainly want to understand whether the infrastructure can be useful for their use cases. This also allowed for those communities who have found that the model will work for them, to directly, focus on the integration work without further administrative overheads.

After the first meeting, the support team starts a continuous tracking and support giving technical aid and evaluating the community progress accordingly to the work plan. Periodic meetings are organised to discuss about the work status and remove hindrances.

When the research community successfully completes its integration work and express interest to start to use the EGI FedCloud in production, the user support team provides assistance: (1) to obtain a commitment from a set of resource providers to support the community with resources, (2) to setup a production VO, (3) to define a SLA with EGI.eu.

The work of the support team is completed when the new community is able to run its services in the EGI production infrastructure in an autonomous way.

The Table 4 summarises all the service offered by EGI during the integration work.

Table 4. Services to support new EGI Federated Cloud users.

Services	Description
Face-2-Face/Web Meetings	• Initial requirements • Allocation of technical experts • Milestones
Continuous tracking and support	• Technical integration support from EGI/NGI team • Periodic meetings
Incubator VO	• Resources for application prototyping • Enabled in all sites • Up to 6 months (renewable)
Documentation	• Step by step guides • Tutorials • Examples
Basic VM Images	• Main OS available • Secure endorsed • Contextualisation
Migration into production	• Resource providers commitment • Support to create a production VO

As described above, the user support process includes a few manual labour intensive steps, therefore it was essential to define a clear process to bring the use cases from the initial tests to a full production status identifying all the activities that should be accomplished during each steps. This end-to-end process consists of eight phases defining actions and responsible parties for each phase (see Table 5 and Figure 2).

Table 5. EGI Federated Cloud use case statuses.

Status	Actions and responsible parties
Pre-assessment	The user support team identified a potential use case that can profit from the EGI Federated Cloud services.
Assessed	The use case requirements are assessed with relevant information added to a dedicated wiki page so to manage the full porting and integration process.
Preparatory	The users and the support team setup the environment to execute tests on the Federated Cloud: configuration of client environment, identification of resource providers, uploading of VM images, etc.
Test and integration	Users access the EGI Federated Cloud through the incubator VO to execute tests while integrating their applications to the infrastructure.
Pre-production	Test and integration are successfully completed. The users create a production virtual organization or join an existing VO.
Production	The users negotiate and agree on a VO SLA with EGI. EGI agrees on a VO OLA with the resource providers supporting the VO. The use case completed all tests and is regularly making use of the Federated Cloud using a production level VO.
Closed	Test and integration are successfully completed. Use case does not foresee moving into production or the use case ends its life cycle (e.g. all the computations have been completed).
Cancelled	Test and integration are completed unsuccessfully. The user or the support team cancelled the use case.

VI. FUTURE WORKS

Nowadays, as described in the previous sections, EGI has already created a training infrastructure for its Federated Cloud and has defined a training programme taking into account the different levels of access that a cloud infrastructure offer to its users: IaaS as a core capability, and PaaS and SaaS layers through external contributors. Dedicated event for each type of access are foreseen, which have the common objective to raise the awareness of the EGI FedCloud and increase the size of its user base with the final goal to advance science. Now, the next steps is the implementation of the envisaged educational program organising the first events and evaluating their outcome in function of that is considered the main success metric, the number of new researchers willing to try the FedCloud and then proceed to become self-sufficient user of the production system. Feedback collected after these tutorials will be used to fine-tune the training programme trying to better focus it according the user requirements.

The EGI Federated Cloud user support team already scheduled the first two training events. Both events will be basic tutorials that, in this initial phase, are better structured to spread the FedCloud paradigm, and in a short format introduce the main features offered by the infrastructure. The first event will be held at the SAP offices in Feltham in the context of a software carpentry event (July 17-19, Feltham, United Kingdom) [42] organised in collaboration with the Harness project. The second will be part of the HPCS-2015 conference programme (July 20-24, Amsterdam, The Netherlands) [43].

After this initial phase, we will start to organise also the other kinds of events, the tutorials to introduce high-level tools and platforms integrated in the FedCloud and the community oriented training events. Probably, the first training events of these types will be held in the last months of 2015 or first 2016. These will benefit of the major numbers of services, platform and tools that will be integrated in the infrastructure. Indeed, currently, several applications are being ported and the FedCloud service portfolio should be considerably enriched for the end of this year. Then, the community oriented training events could cover a wider set of services and so increasing their appeal for scientists. Currently, bio-informatics, earth science and physics are the scientific disciplines most promising for organising a community oriented tutorials. Furthermore, the EGI-Engage [44] work plan foresees the integration of several platforms in the infrastructure. Some of the more relevant are listed below:

- a uniform platform for international astronomy research collaboration developed in collaboration with

Figure 2. Support workflow for the full use-case life cycle.

the Canadian Advanced Network for Astronomical Research (CANFAR) [45];

- the D4Science infrastructure [46] that hosts more than 25 Virtual Research Environments to serve the biological, ecological, environmental, and statistical communities world-wide;

- the EPOS thematic core services (TCS) [47];

- the e-Collaboration for Earth Observation (e-CEO) platform [48] developed by the European Space Agency (ESA)[49].

A wider portfolio of high-level platform and tools available will make the organisation of the tutorials to introduce high-level tools and platforms easier. The integration of these platforms require the investigation and integration of additional capabilities to the IaaS layer of the EGI Federated Cloud. Particularly:

- Broadening support for the Cloud Data Management Interface (CDMI) into the federation through the availability of implementations for the other CMFs beyond Synnefo.

- Enabling larger customizability within community-specific Virtual Organisations that are created on the EGI Federated Cloud. Greater level of customizability is required to enable features that are necessary for certain communities, for example username-based login (for example in a bioinformatics Virtual Organisation), or enable technology-specific services, for example OpenStack Heat (in an OpenStack-specific Virtual Organisation).

Considering the remarkable number of innovations in the FedCloud foreseen for the next months, it is not possible yet to define the full calendar of the future training events, but those to be organised in July will serve as blueprints for national and regional events in the future.

VII. CONCLUSIONS

Overall, the adoption of the EGI Federated Cloud has been significant, both in terms of research communities with which the EGI community already has a relationship but also those that it does not. A key part of this is having a well-defined and structured 'on-ramp' with which we can support a community that may have never considered using a national or international scale e-infrastructure and bringing them into the wider communities with which we are already engaged. A key part of this is the ability to provide training to all ability and knowledge infrastructure provided by the EGI is of fundamental importance when considering how the utilisation of e-infrastructures, which

is a key goal of the European Commission going forward in the European Research Area, can be encouraged.

Cloud computing and the EGI Federated Cloud is by it fundamental nature exceptionally suited to the provision of training environments. The ability to launch on demand multiple easily managed replicas of a single environment, which is self-contained and sandboxed from other users due to the use of virtualization, is an especially powerful tool. This ability also ensures for the point of view of the training deliverer the simplification of both pre-course setup as well as dealing with errors or problems during training delivery. Due to the presence of the complete environment in a controlled manner, we are also through the same infrastructure able to support multiple different groups from within a single community. An example may be the community specific IT support groups that need training in the underlying operation of their infrastructure and applications as well as the end user community who really only need to know about the end user facing services. They would both be able to be supported by the same infrastructure with different authorisation rights attached to the training accounts for each specific activity.

With the ability support multiple different paradigms also we have shown that we are able to tailor support and training requirements to different communities and, therefore, not only are we able to provide them custom environments for the training itself but, with the connection to the Federated Cloud, more generally the trainee is also able to take the environment, which they may have become extremely comfortable with away with them and just continue using it for their day to day activities. This gives training activities the power of almost one on one training programs with the cost advantage of a more generally available type of course after which a user may have to take what they have learnt and. before exploiting. it build a similar environment within their own institutional resources etc. This type of growth of capability is extremely important when we consider the new paradigms that are being more routinely used now including Platform and desktop as a Service which provide common environments for the end user irrespective of the underlying platforms and allow the disconnection for the first time between the user and the 'computers'.

The final significant advantage of the federated cloud is that it supports the greater mobility that is available within the European Union now for research communities and their participants. The availability of a single common e-infrastructure platform across Europe allows communities to provide a single common training methodology and material, which may be utilised by all of the participants irrespective of location in Europe. Further afield with growth in the FedCloud

internationally and into global projects will only increase this reach.

It is clear that IaaS on its own is not a solution for training but the EGI cloud platform, which supports all layers of the stack from the underpinning foundational core platform to the Community platform and application catalogue provided by the Application DB, which allow communities to share exemplar tools and services. This also greatly supports the drive for greater openness in publically supported research and development as well as provides a publishing location for these services which will have significant visibility both inside and outside their own communities.

REFERENCES

[1] EGI Federated Cloud, https://www.egi.eu/infrastructure/cloud/.

[2] The ESFRI roadmap, http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap.

[3] The EGI-Engage project - Engaging the Research Community towards an Open Science Commons, https://www.egi.eu/about/egi-engage/

[4] Helix Nebula project http://helix-nebula.eu

[5] Coppola, Massimo, et al. "The Contrail approach to cloud federations." *Proceedings of the International Symposium on Grids and Clouds (ISGC'12)*. Vol. 2. 2012.

[6] SixSq, SlipStream, http://sixsq.com/products/slipstream.html.

[7] CONTRAIL project training program, http://contrail-project.eu/training.

[8] R. Nyren, A. Edmonds, A. Papaspyrou, and T. Metsch, "Open Cloud Computing Interface - Core," GFD-P-R.183, April 2011. [Online]. Available: http://ogf.org/documents/GFD.183.pdf

[9] T. Metsch and A. Edmonds, "Open Cloud Computing Interface - RESTful HTTP rendering," GFD-P-R.185, June 2011. [Online]. Available: https://www.ogf.org/documents/GFD.185.pdf

[10] T. Metsch and A. Edmonds, "Open Cloud Computing Interface - Infrastructure," GFD-P-R.184, April 2011. [Online]. Available: http://ogf.org/documents/GFD.184.pdf

[11] SNIA Technical Position, "Cloud Data Management Interface (CDMI) 1.0.2", June 2012 http://snia.org/sites/default/files/CDMI%20v1.0.2.pdf

[12] DMTF, "Open Virtualization Format Specification v2.1.0", December 2013

[13] http://www.dmtf.org/sites/default/files/standards/documents/DSP0243_2.1.0.pdf

[13] HEPIX Virtualisation Working Group, "HEPiX Virtualisation Book", August 2012.

[14] D. Wallom & M. Drescher, "MS520 EGI Platform Roadmap", June 2014, https://documents.egi.eu/document/2232

[15] S Moser et al, "Configuration and customization of cloud instances", https://cloudinit.readthedocs.org/en/latest/index.html

[16] T. Metsch, A. Edmonds. OCCI-OS https://wiki.openstack.org/wiki/Occi

[17] A. Asiki, Ch. Loverdos. snf-occi https://code.grnet.gr/projects/snf-occi

[18] Parák, B., Šustr, Z., Feldhaus, F., Kasprzakc, P., & Srbac, M. "The rOCCI Project–Providing Cloud Interoperability with OCCI 1.1". In *International Symposium on Grids and Clouds (ISGC)* (Vol. 23, No. 28), March 2014.

[19] EGI Applications Database, "AppDB," https://appdb.egi.eu/.

[20] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnello, A. Frohner, A. Gianoli, K. Lorentey, and F. Spataro, "Voms, an authorization system for virtual organizations," in *Grid computing* (pp. 33-40). Springer, 2004.

[21] Denemark, J., Matyska, L., Ruda, M., Jankowski, M., Meyer, N., & Wolniewicz, "User Management for Virtual Organizations". In *Integrated Research in GRID Computing* (pp. 135-146). Springer, 2007.

[22] López García, A. Fernández-del-Castillo, E., Puel, M, "Identity Federation with VOMS in Cloud Infrastructures". In Proceedings of the 2013 IEEE International Conference on Cloud Computing Technology and Science - Volume 01. IEEE, 2013.

[23] A. Cristofori, et al, "Usage Record – Format Recommendation", July 2013, http://www.ogf.org/documents/GFD.204.pdf

[24] A. López, S. Pinto, E. Fernández, B. Parak. cloud-bdii-provider. https://github.com/EGI-FCTF/cloud-bdii-provider

[25] Valeria Ardizzone, Roberto Barbera, Antonio Calanducci, Marco Fargetta, E. Ingrà, Ivan Porro, Giuseppe La Rocca, Salvatore Monforte, R. Ricceri, R. Rotondo, Diego Scardaci, Andrea Schenone: "The DECIDE Science Gateway". Journal of Grid Computing Vol. 10(4), pages 689-707, Editor Springer 2012.

[26] Vcycle: VM lifecycle management, http://www.gridpp.ac.uk/vcycle/

[27] Daniele Lezzi, Roger Rafanell, Abel Carrión, Ignacio Blanquer Espert, Vicente Hernández, Rosa M. Badia, "Enabling e-Science Applications on the Cloud with COMPSs". Euro-Par 2011: Parallel Processing Workshops, Lecture Notes in Computer Science Volume 7155, 2012, pp 25-34.

[28] Víctor Méndez Muñoz, Adrià Casajús Ramo, Ricardo Graciani Diaz, Víctor Fernández Albor, "How to Run Scientific Applications with DIRAC in Federated Hybrid Clouds". ADVCOMP 2013, The Seventh International Conference on Advanced Engineering Computing and Applications in Sciences.

[29] Peter Kacsuk, Zoltan Farkas, Miklos Kozlovszky, Gabor Hermann, Akos Balasko, Krisztian Karoczkai, Istvan Marton, "WS-PGRADE/gUSE Generic DCI Gateway Framework for a Large Variety of User Communities". Journal of Grid Computing, December 2012, Volume 10, Issue 4, pp 601-630.

[30] EUGridPMA Guidelines on Approved Robots, https://www.eugridpma.org/guidelines/robot/

[31] Roberto Barbera, Giuseppe Andronico, Giacinto Donvito, Alberto Falzone, J. J. Keijser, Giuseppe La Rocca, Luciano Milanesi, Giorgio Pietro Maggi, Saverio Vicario, "A grid portal with robot certificates for bioinformatics phylogenetic analyses". Concurrency and Computation: Practice and Experience 23(3): 246-255 (2011)

[32] Per-User-Sub-Proxies specification, https://wiki.egi.eu/wiki/Fedcloud-tf:WorkGroups:Federated_AAI:per-user_sub-proxy.

[33] EGI Federated Cloud High Level Tools, https://wiki.egi.eu/wiki/HOWTO10#4._Infrastructure_broker_2.

[34] EGI Operations Portal VO search engine, http://operations-portal.egi.eu/vo/search.

[35] The EGI Resource Allocation tool - e-Grant. https://e-grant.egi.eu.

[36] EGI.eu, http://www.egi.eu/.

[37] National Grid Initiatives (NGIs), https://www.egi.eu/community/ngis/.

[38] EGI Federated Cloud User Support wiki, https://wiki.egi.eu/wiki/Federated_Cloud_user_support.

[39] EGI Federated Cloud Use Cases, https://wiki.egi.eu/wiki/Federated_Cloud_Communities.

[40] EGI Computer Security Incident Response Team (CSIRT), https://www.egi.eu/about/policy/groups/EGI_Computer_Security_Incident_Response_Team_EGI_CSIRT.html

[41] EGI Federated Cloud Marketplace, https://appdb.egi.eu/browse/cloud.

[42] EGI Federated Cloud tutorial co-organised with the Harness project, July 17.19, Feltham, United Kingdom, https://harnesscloud.github.io/2015-07-15-feltham/

[43] EGI Federated Cloud tutortial at the HPCS 2015 conference, July 20-24, Amsterdam, The Netherlands, http://hpcs2015.cisedu.info/

[44] The EGI-Engage project, https://www.egi.eu/about/egi-engage/.

[45] Canadian Advanced Network for Astronomical Research (CANFAR), http://www.canfar.phys.uvic.ca/canfar/.

[46] The D4Science infrastrucutre, https://www.d4science.org/.

[47] The European Plate Observing System – EPOS, http://www.epos-eu.org/.

[48] The e-Collaboration for Earth Observation platform - e-CEO. http://www.ca3-uninova.org/project_eceo

[49] The European Space Agency – ESA, http://www.esa.int/ESA.