Cloud and HTC integration Workshop

Dissemination level: Public

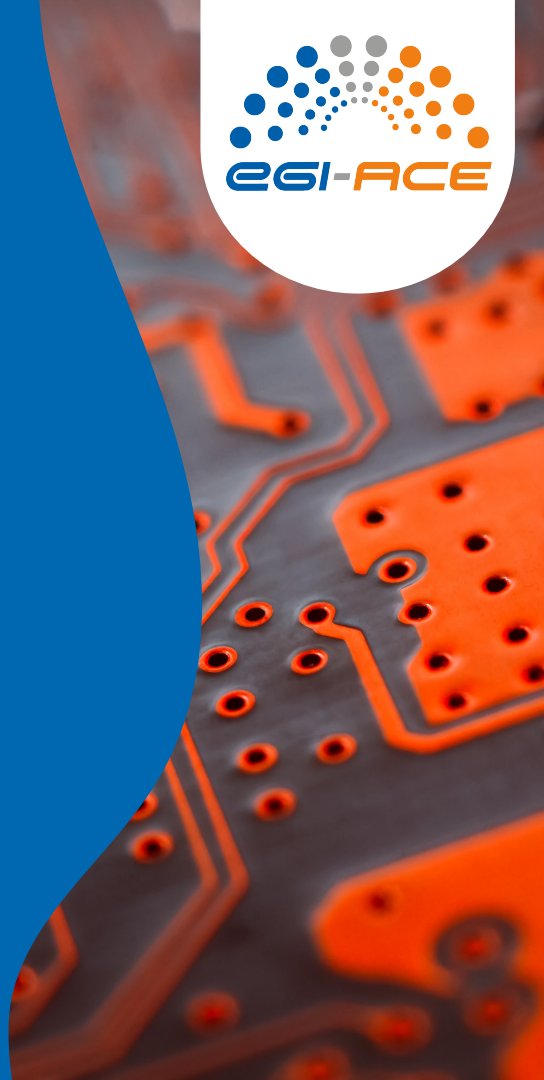
Disclosing Party:

Recipient Party:

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

Cloud & HTC integration workshop
Welcome!

Starting at 9.00 CET



Agenda overview

(all times CET)

- 9:00 – 9:50 Welcome and infrastructure objectives and status overview

- 4 blocks with presentations from partner infrastructures:
  - 10:00 – 10:50 Block 1
  - 11:00 – 11:50 Block 2
  - 12:00 – 12:50 Block 3
  - (Lunch)
  - 14:15 – 15:15 Block 4

- 15:25 – 16:30 Discussion block

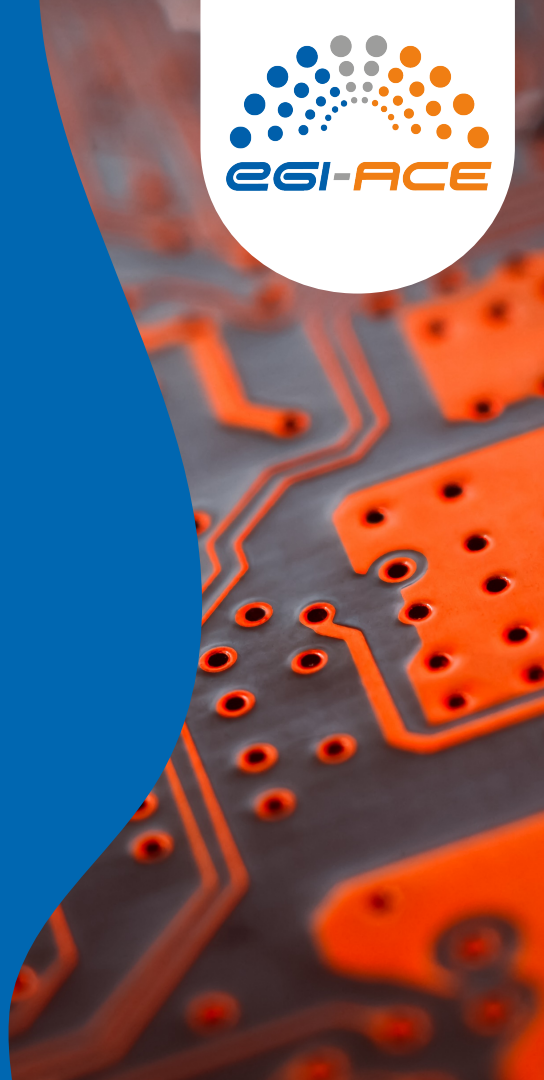
Introduction

EGI-ACE

EOSC Compute Platform

Onboarding

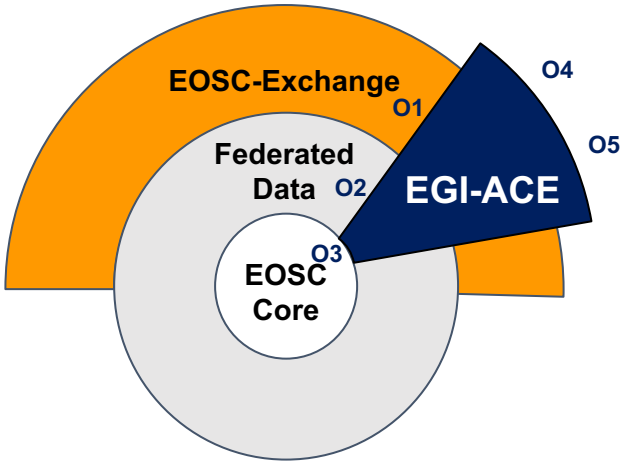
Capacity Allocation



EGI-ACE objective

Implement the **Compute Platform of the European Open Science Cloud** and contribute to the **EOSC Data Commons** by delivering integrated computing, platforms, data spaces and tools as an integrated solution that is **aligned with major European cloud federation projects and HPC initiatives**.

EGI-ACE specific objectives



Objective 1. Deliver the European Open Science Cloud Compute Platform and expand the supply-side

Objective 2: Contribute to the implementation of the EU Data Strategy and the EOSC Data Commons to support the Green Deal, Health and Fundamental Research

Objective 3: Integrate the EOSC Compute Platform with the EOSC Portal and the EOSC Core

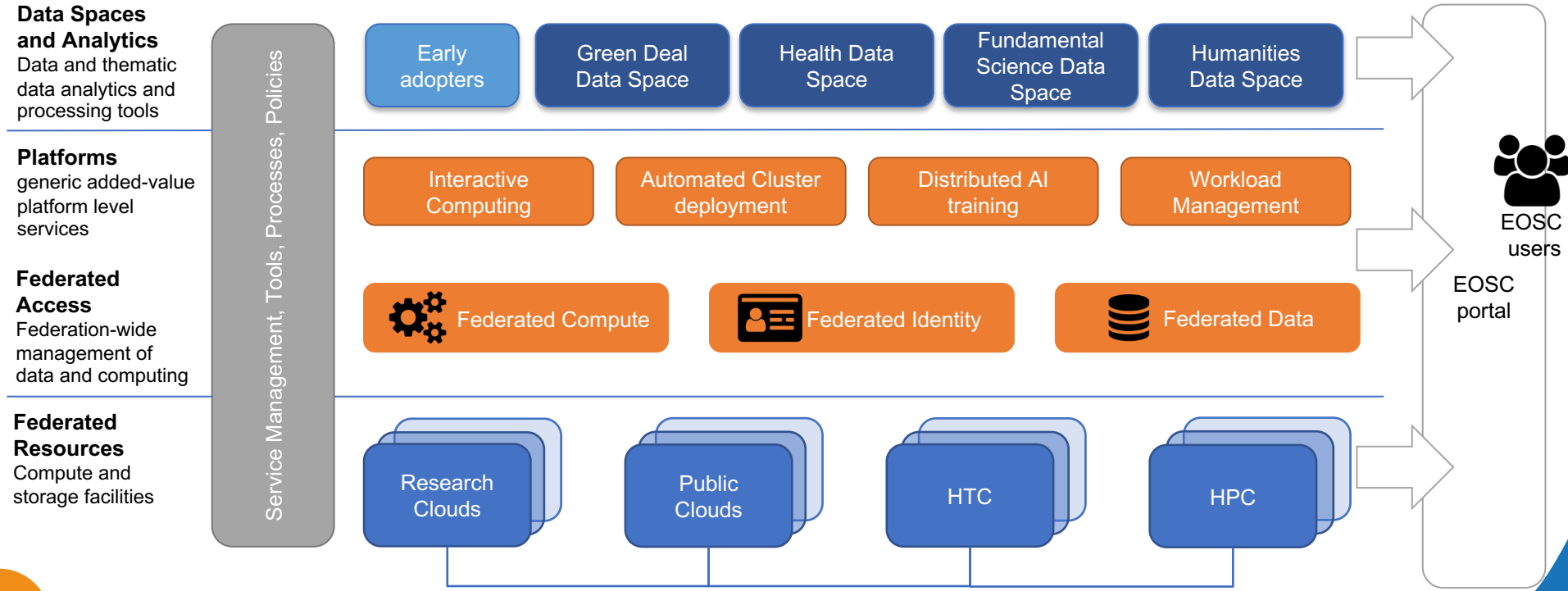
Objective 4: Contribute to the realization of a global Open Science Cloud

Objective 5: Expand the demand-side of EOSC across sectors and disciplines

Solutions for a sustainable EOSC

report from the EOSC Sustainability WG

EGI-ACE concept and methodology: Tier service architecture



Infrastructure layer – Cloud, HTC, HPC

For users and communities within the project **AND** for new EOSC users

- Increasing capacity (Cloud and HTC):
  - **Phase 1 (month 1-10)** ⇒ 13.8 Million CPU hours/41,300 GPU hours/7.5 PB/month
  - **Phase 2 (month 11-20)** ⇒ 27.6 Million CPU hours/83,000 GPU hours/15 PB/month
  - **Phase 3 (month 21-30)** ⇒ 41.4 Million CPU hours/124,000 GPU hours/23 PB/month
- **Mix of funding mechanisms:**
  - Virtual Access:

Diverse locations AND compute flavors

  - Local grants enabling policy-based access (EGI Cloud Federation + new partners):

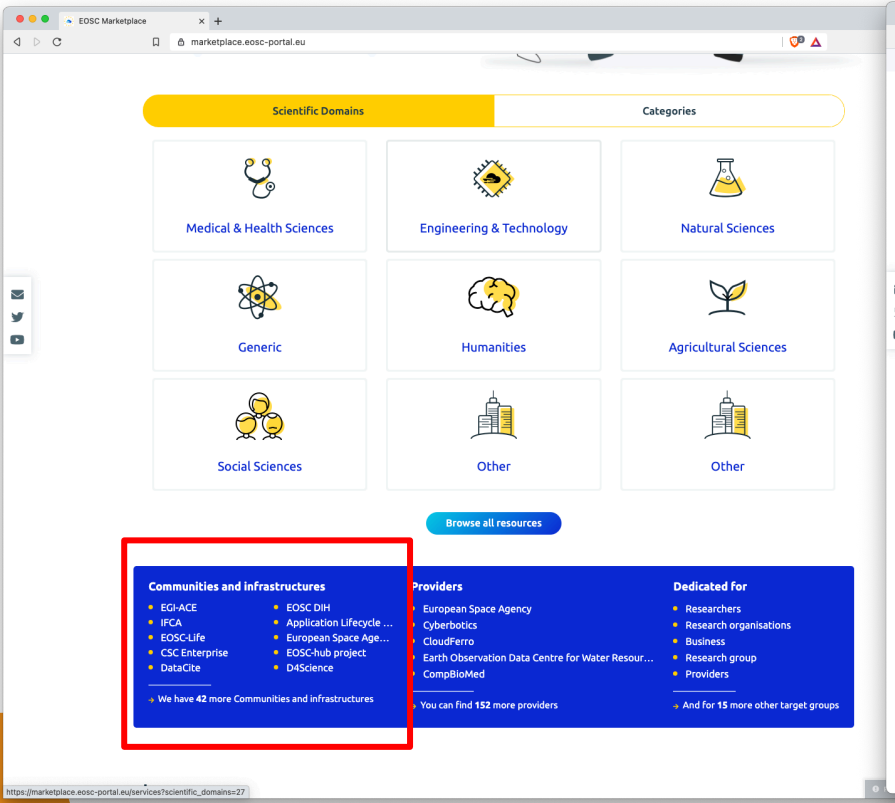
Moldova, Latvia, Georgia, Hungary, Austria, USA, China, South Africa

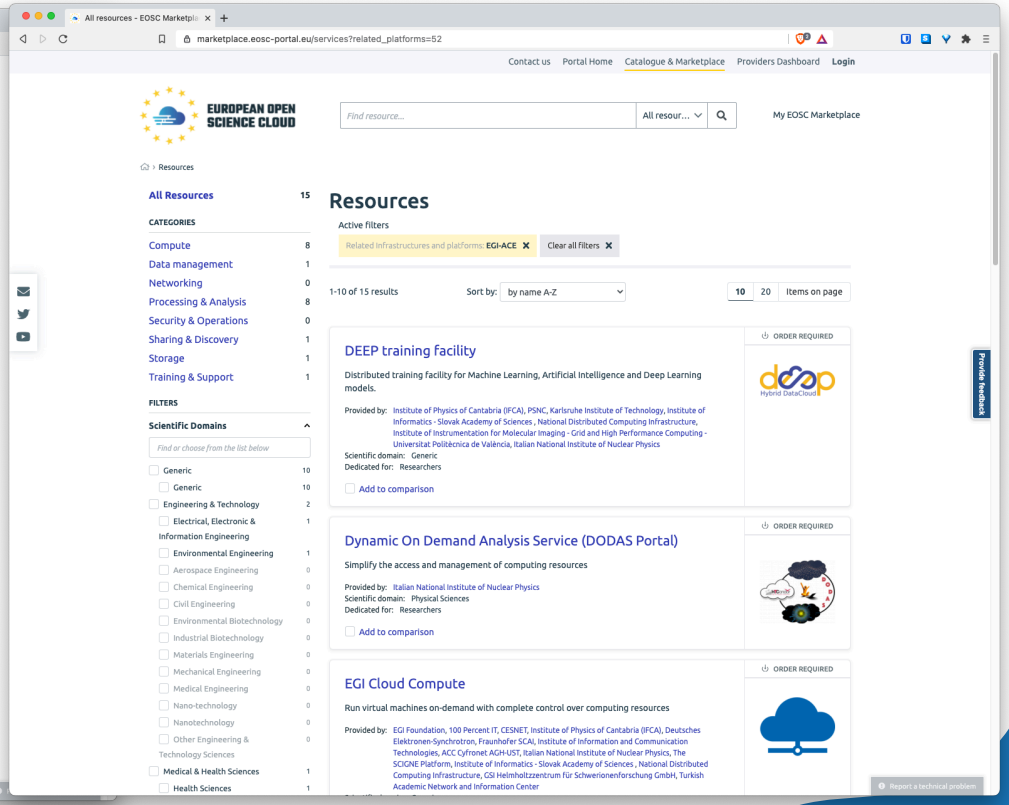
  - Pay-for-use:
  - Pilot HPC sites:

International cloud activities (WP7.4)

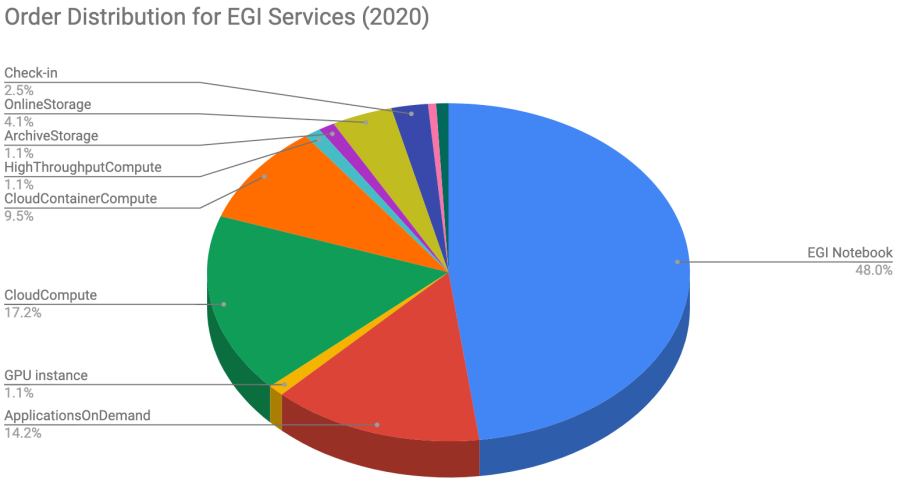
- **Coordinated delivery of cloud and user support**
  1. Onboarding into EOSC marketplace
  2. Integrate EGI Check-in
  3. Deliver user support and training for national and international communities
- **Exchange applications and data**
  1. Share applications with EGI AppDB (VM images)
  2. Share data with EGI DataHub
  3. Facilitate cross-infrastructure data processing and analytics workflows as demanded by user communities.
- **Impact of e-infrastructure services**
  1. To share EGI and EOSC user feedback and to feed this into their continuous improvement processes.
  2. To share success stories, to prepare joint articles, publications, presentations and demonstrations

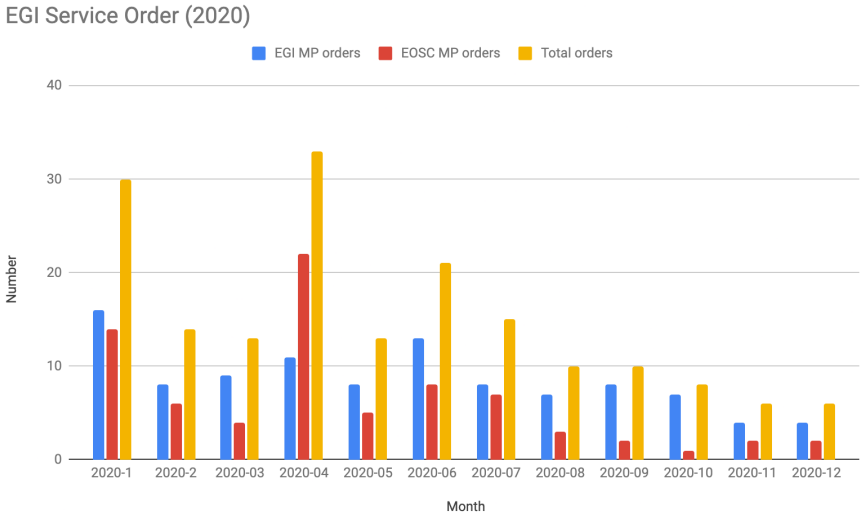
EGI-ACE portfolio in the EOSC Marketplace





Orders for EGI services



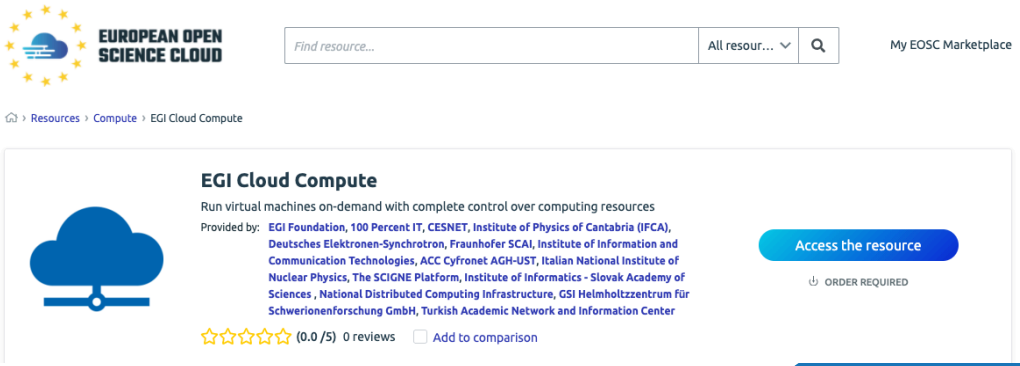


EGI services are among the most demanded from EOSC Market Place (32% during 2020)
EGI Cloud Compute is the most demanded service from EGI on EOSC Market Place (42%)

Onboarding

1. Service is directly visible in the EOSC Portal & Marketplace
   - It's already there → No further action
   - It has to be registered → https://eosc-portal.eu/for-providers

2. Service is part of another service that's visible in EOSC Portal & Marketplace
   - It's already there → No further action
   - It has to be added
     - As a supplier
     - As a new option
     - As a new component



Resource allocation, user support

Single users
Small groups
Experimental users

International projects
Multi-national communities

EOSC USERS

Business-to-User

Business-to-Business

EOSC Portal

- Ready-to-use resources/services
- Self-service configuration
- Short term engagement

EGI-ACE Open calls
(3x per year, 1st in March 2021)

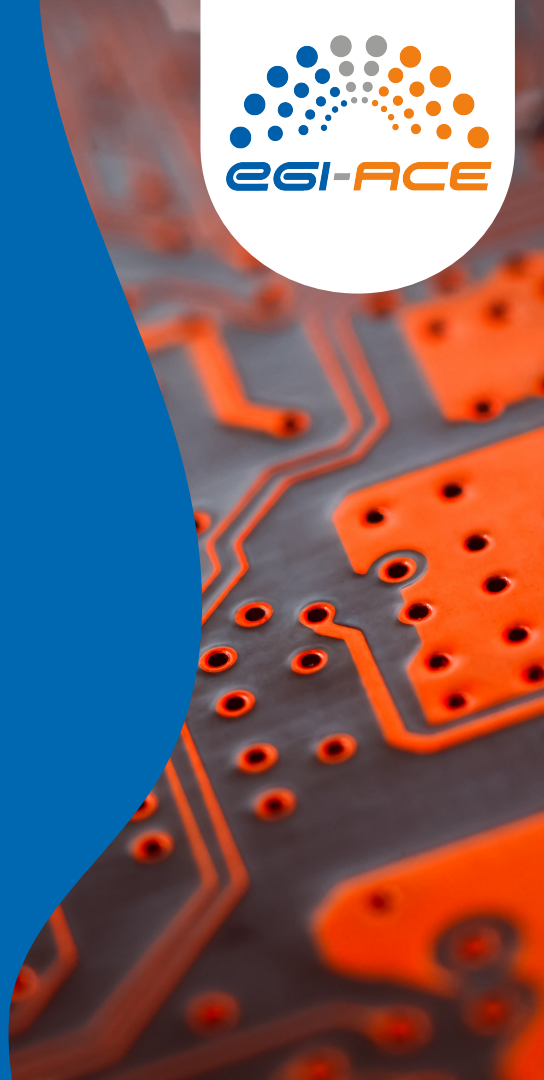
- High capacity demand
- Custom configurations
- Long term engagement

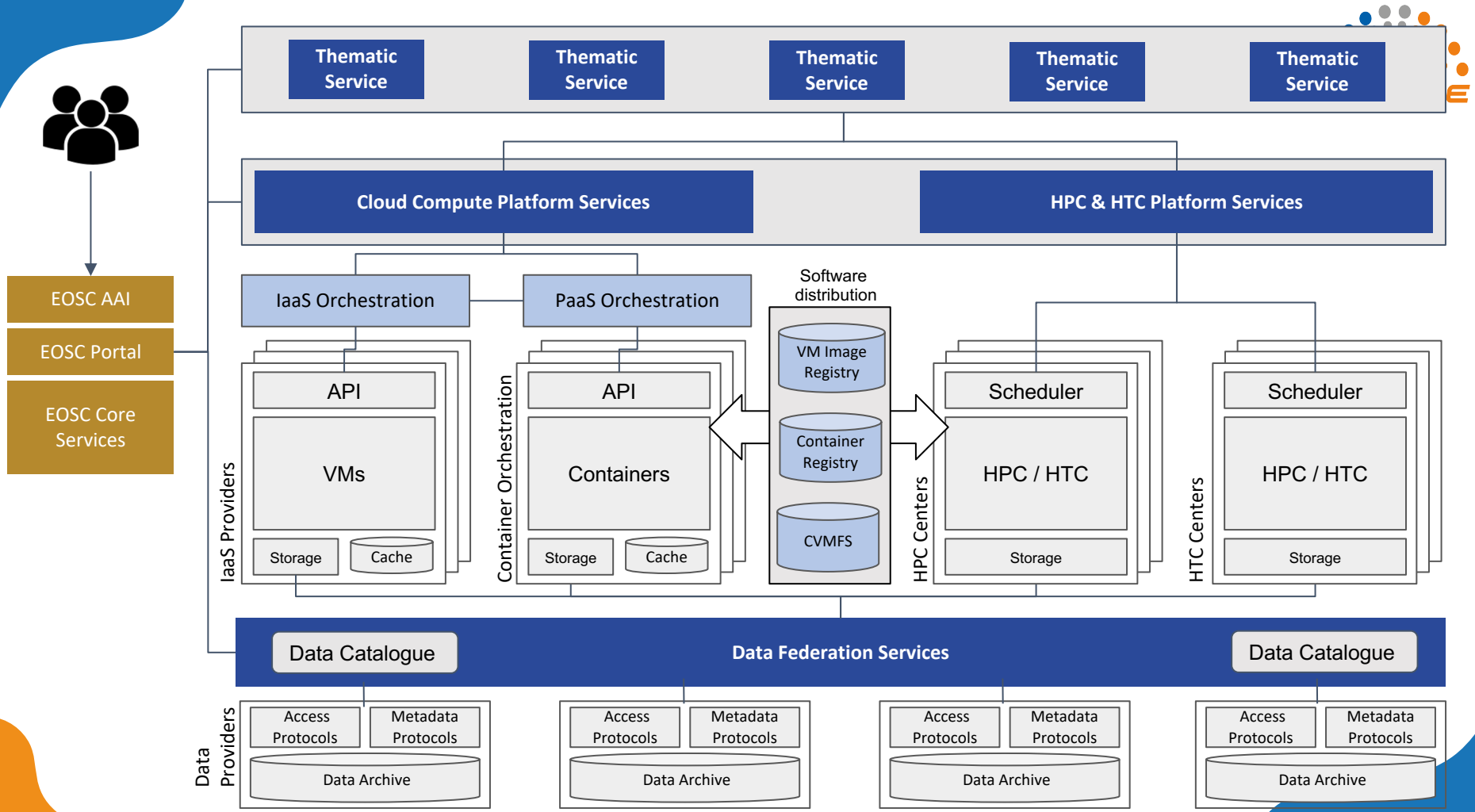
Resource allocation

2 phases:

- Assessment phase: evaluation of requests, identification and selection of providers, assign implementation shepherd
- Implementation phase: final mapping of services, create competence centre to support implementation, and monitor implementation

Technical integration




Thematic Service
Thematic Service
Thematic Service
Thematic Service
Thematic Service
Cloud Compute Platform Services
HPC & HTC Platform Services
EOSC AAI
EOSC Portal
EOSC Core Services
IaaS Orchestration
PaaS Orchestration
Software distribution
VM Image Registry
Container Registry
CVMFS
IaaS Providers
API
VMs
Storage
Cache
Container Orchestration
API
Containers
Storage
Cache
HPC Centers
Scheduler
HPC / HTC
Storage
HTC Centers
Scheduler
HPC / HTC
Storage
Data Catalogue
Data Federation Services
Data Catalogue
Data Providers
Access Protocols
Metadata Protocols
Data Archive
Access Protocols
Metadata Protocols
Data Archive
Access Protocols
Metadata Protocols
Data Archive
Access Protocols
Metadata Protocols
Data Archive

Integration areas / handbook

1. Access federation
2. Application federation
3. Data federation
4. Operation federation

http://go.egi.eu/egi-ace-integration

EGI-ACE integration handbook

Abstract:
EGI-ACE project delivers the **EOSC Compute Platform**, a distributed federated infrastructure that supports the present and future needs of data-centric research-focused computing in Europe. This handbook provides guidance for service providers that want to federate their resources into this platform. The document describes the various technical integration options and points to specific manuals for further details. The handbook is meant to be used as a document to understand the 'big picture' of EOSC Compute Platform onboarding and to choose the best approach for integrating into it.

URL:
http://go.egi.eu/egi-ace-integration

Version (date of last update):
v1.0 (2021-02-26)

Access Federation

Check-in

RCAuth.eu &
Master Portal

EGI Check-in

Proxy service that operates as a central hub to connect federated Identity Providers (IdPs) with EGI service providers

Perun

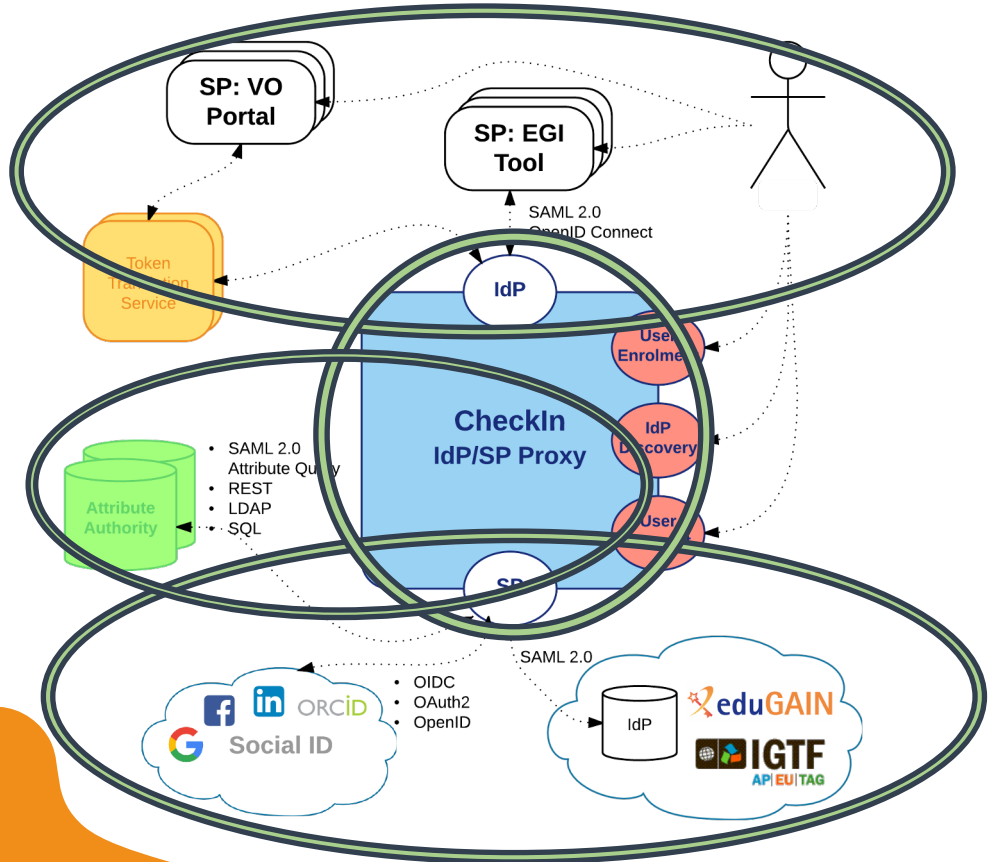
RCAuth.eu &
Master Portal

Main characteristics:

- Enables multiple federated authentication sources using different technologies
- Federated in eduGAIN as a service provider (REFEDS RnS and Sirtfi compliance)
- User registration portal to allow accounts-linking
- Combines user attributes originating from various authoritative sources (IdPs and attribute provider services)
  - Integrated with **PERUN**
- Compliant with the AARC Blueprint
- Support non-web use cases & delegated access
  - Integrated with RCAuth.eu and **Master Portal**

The EGI AAI Platform: Check-in

The IdP/SP Proxy

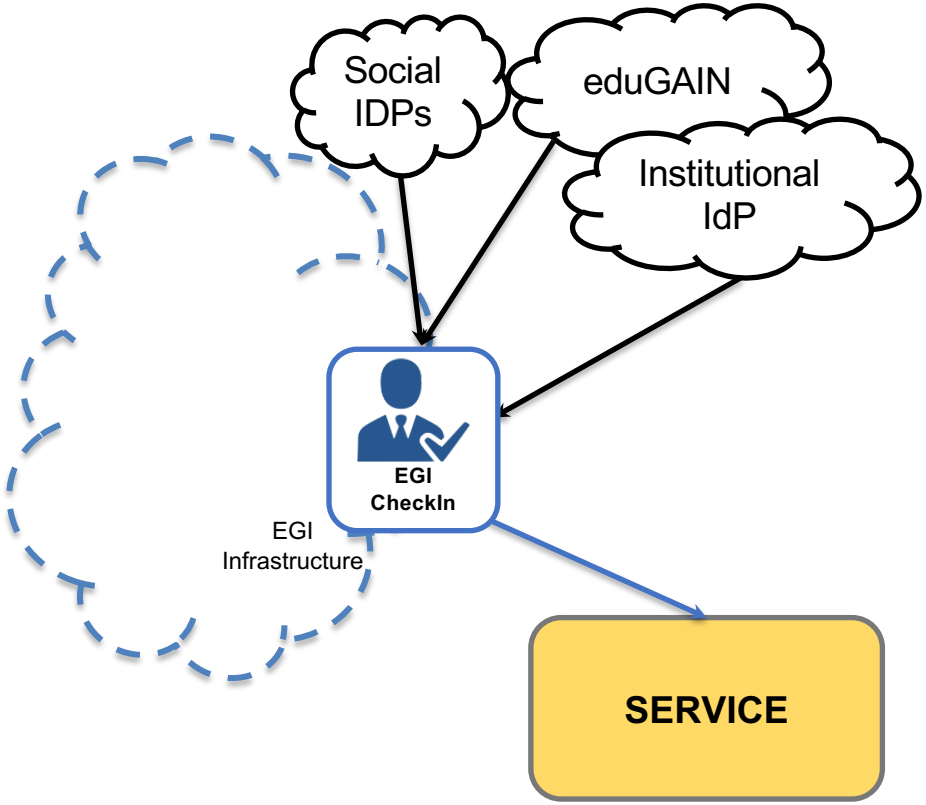


- Identity Providers:
  - SAML2.0: eduGAIN
  - OIDC/OAuth2: Google, Facebook, LinkedIn, ORCID, WeChat, BitBucket
  - X.509: IGTF
- Service Providers:
  - SAML2.0 & OIDC
- Attribute Authorities
  - SAML2.0 Attr. Query, REST, LDAP, SQL
- Token Translation Services
  - SAML2.0-to-X.509: Master Portal to RCauth.eu Online CA
- Support for Levels of Assurance
- User enrolment & account linking
- IdP Discovery
- User Consent

Use case: AAI as a Service

Check-in as an authentication proxy to allow user logins from institutional IdPs in eduGAIN and social media for non EGI services

- Minimal overhead for the service development
- Prerequisites:
  - Service provider must accept EGI policies on data protection

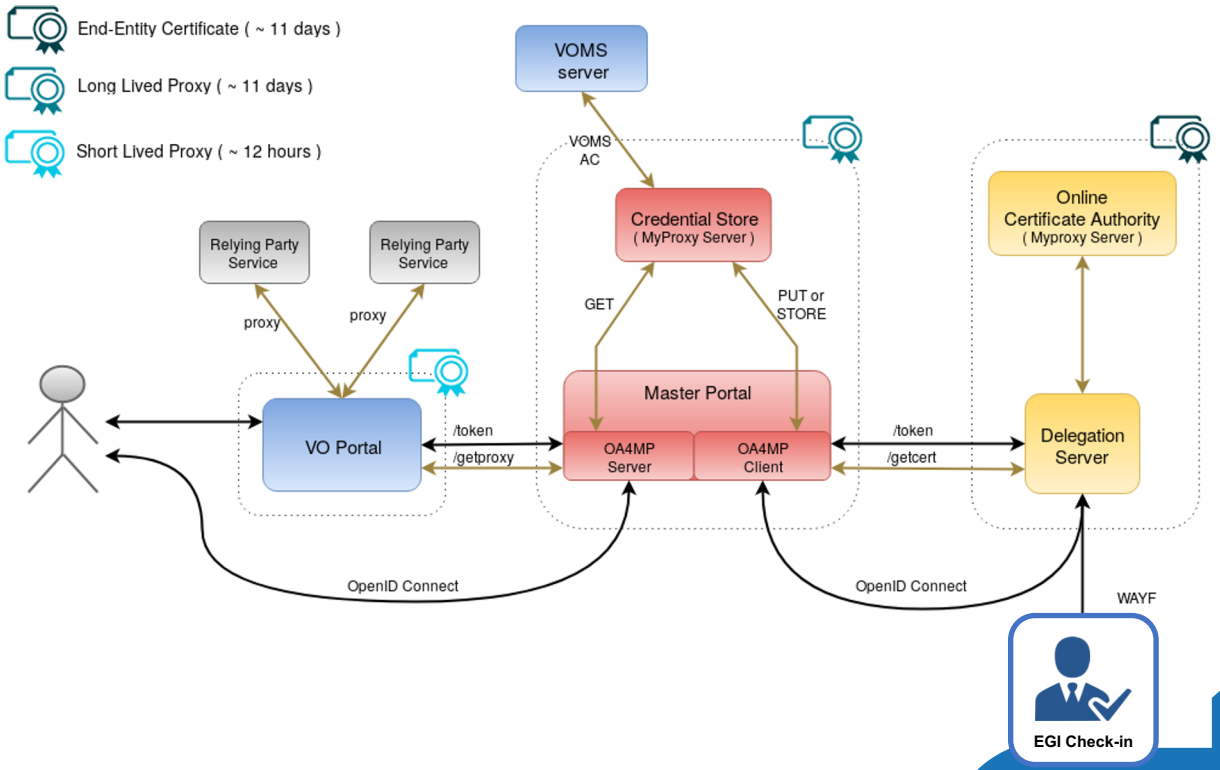


Non-web use cases & delegated access

RCauth.eu Online CA issued certificates

Check-in has been integrated with the production RCAuth.eu Online CA for allowing users to retrieve X.509 proxy certificates using their federated credentials

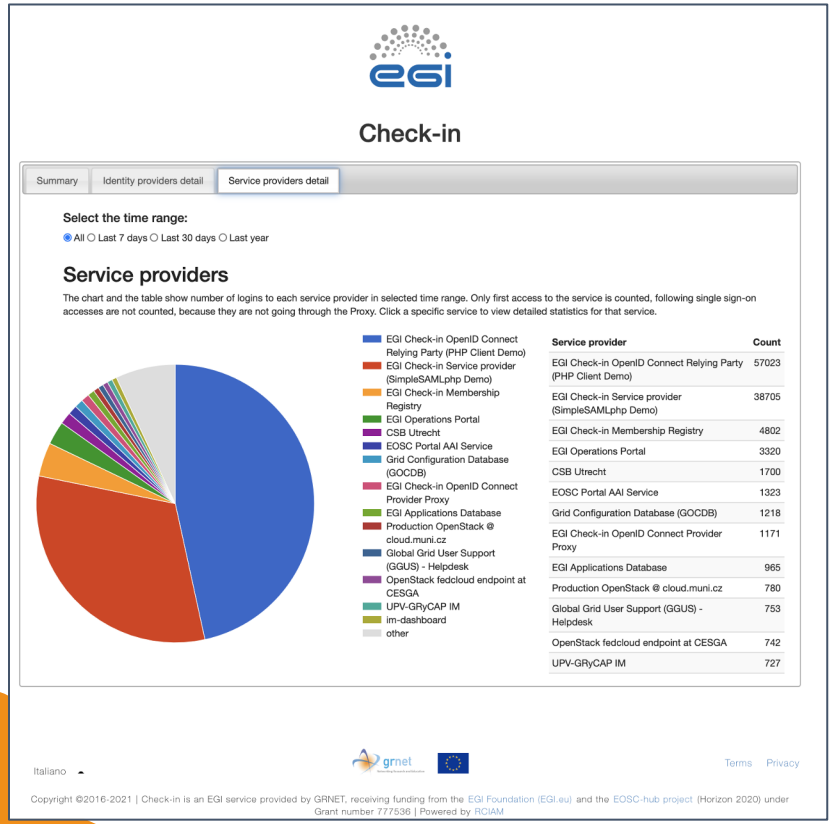
- Master Portal retrieves end-entity certificate from RCauth.eu
- Long-lived proxy certificate stored in backend MyProxy server
- Short-lived proxies provided via:
  - Science Gateways via OIDC (so-called VO-portals)
  - users e.g. via SSH key authentication



EGI Check-in Statistic viewer

Login statistics for services integrated

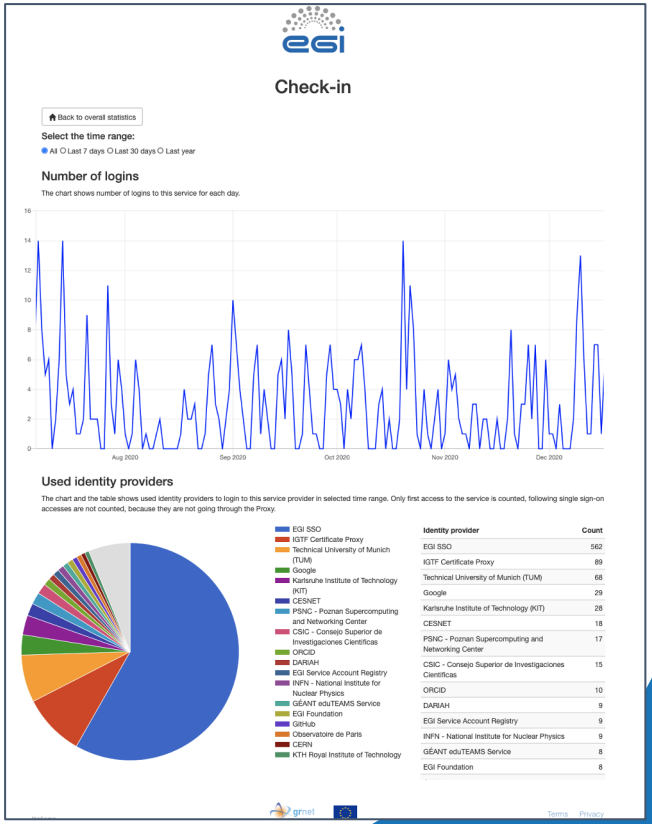
EGI Check-in statistics



General view of the login statistics (all services)

Dedicated service view for login statistics:

- time trend of user logins
- distribution of logins based on the IdPs used by users



Application federation

Software Distribution

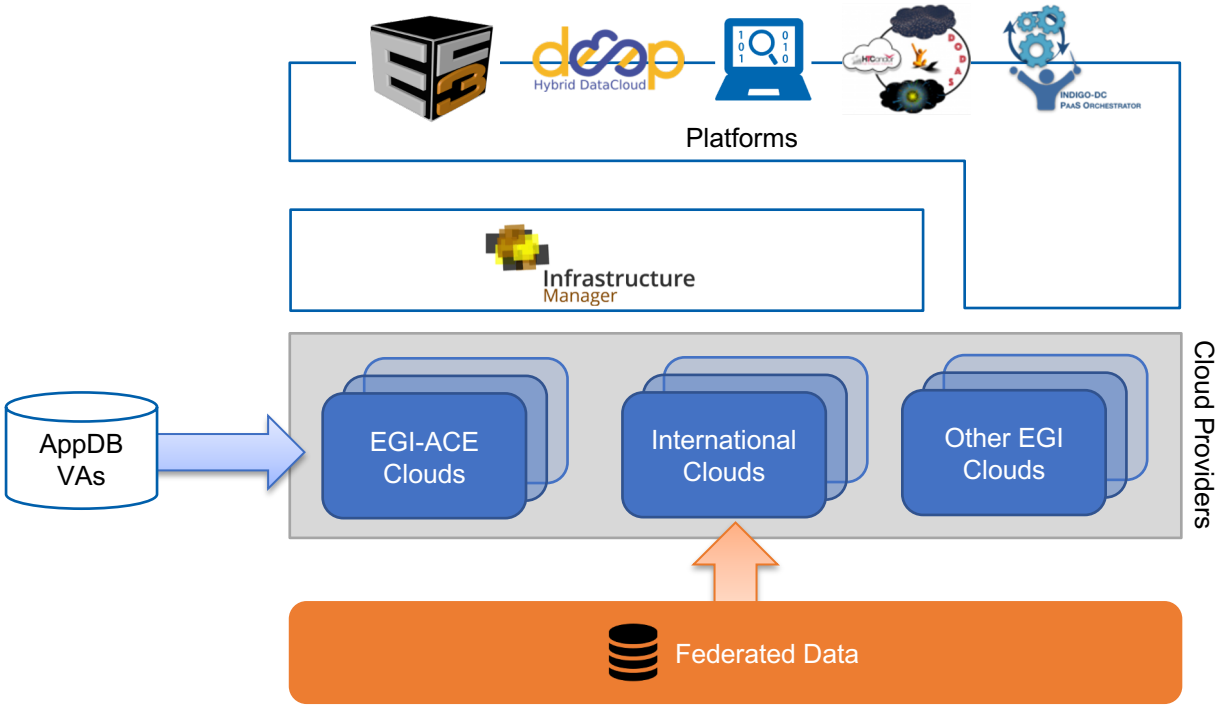
CernVM-File System (CVMFS) provides a POSIX read-only file system in user space

- Scalable software distribution, recently extended to handle container images
- Caching at providers for fast access to frequently used software
- Mostly used in HTC, but useful also for IaaS

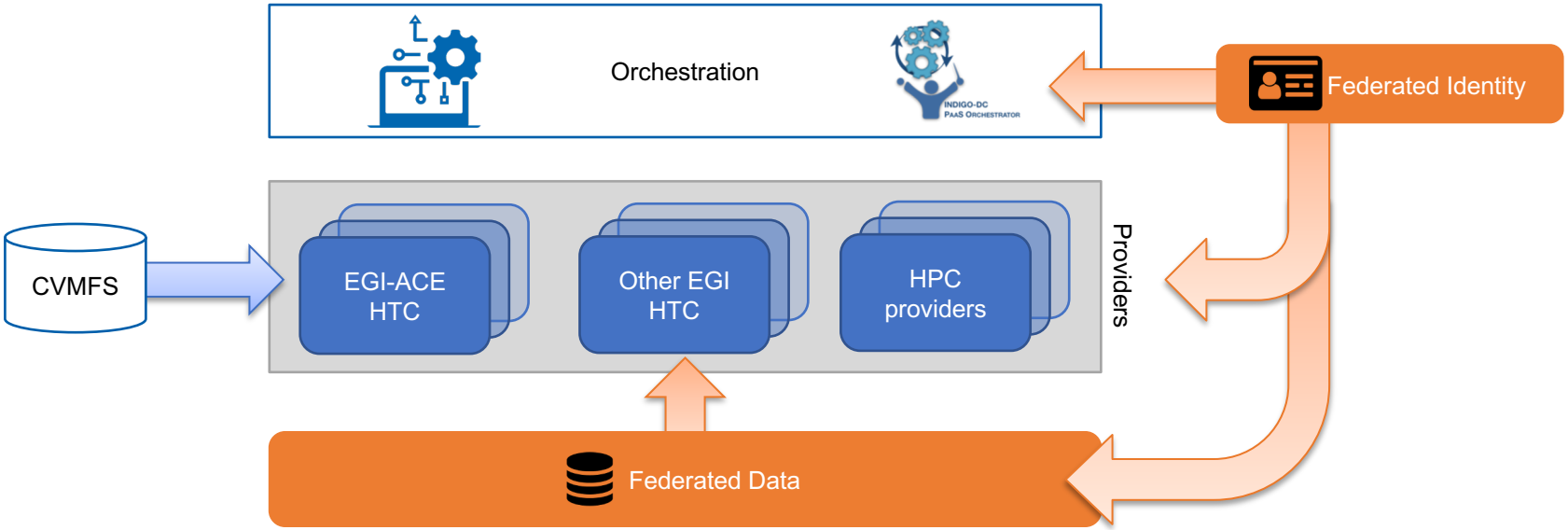
Application Database (AppDB) provides Virtual Appliance (VAs) Catalogue

- Public registry of images with detailed metadata
- Community-managed VA lists, fine control on what can be executed by users
- Automatic distribution as VM images to the cloud providers

EGI-ACE platforms for Cloud



Beyond IaaS Clouds: HTC and HPC



Data Federation

RUCIO

DataHub

Access key scientific datasets in a scalable way

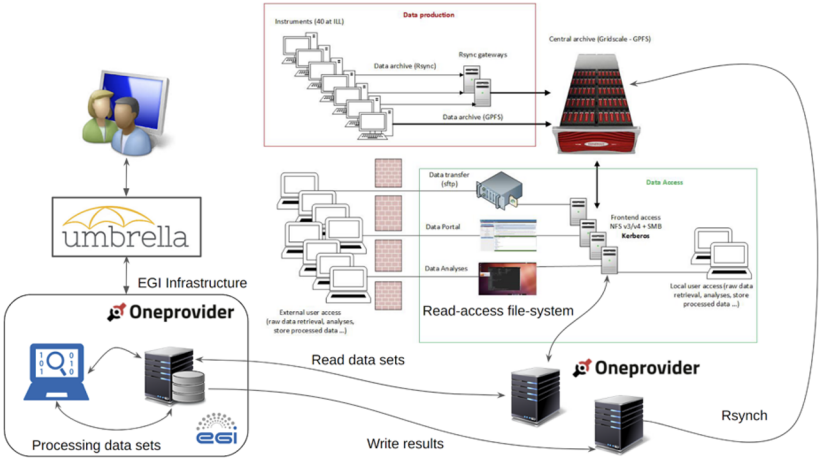
- Service based on **Onedata** technology

- It allows **transparent data access** under a **common namespace** regardless of the location
- Data can be accessed via a **GUI** or **APIs**
- Developed in EOSC-hub project and

  under integration by many communities

- 1 central managing component +

distributed component at providers



DataHub

Main functionalities

- Integrated with EGI-Checkin
- File/Share Management
  - Shares can eventually be published via Handle services and have a PID assigned
- Extensive backend support
  - Posix, S3, Swift, CEPH, GlusterFS
- Data import
  - Allows exposing existing datasets
- Metadata Management
  - Metadata associated to files (k-v pairs, json, rdf)
  - OAI-PMH interface
- Harvester Management
  - Possibility to automatic extract metadata from files published to a space
  - Metadata updates are pushed to external indexes (e.g. Elastic Search) and automatically synchronized

Data Transfer

Transfer large datasets between storages

FTS File Transfer Service

- Easy users interaction for submitting transfers
- Portal for end-users, Real Time monitoring and Web Admin
- Checksums and retries are provided per transfer
- Multiprotocol support
- Different clients to access the service (RESTFul APIs, python bindings)
- Transfers from/to different storages (including staging from tape)
- Parallel transfers scheduling and optimization to get the most from network without burning the storages
- Priorities/Activities support for transfers classification

Rucio

Scalable data management

Rucio is the software developed at CERN for the management of LHC experiments data

- provides a complete and generic scientific data management service
- Rucio manages multi-location data in a heterogeneous distributed environment
  - Creation, location, transfer, and deletion of replicas of data
  - Orchestration according to both low-level and high-level driven data management policies (usage policies, access control, and data lifetime)
  - Interfaces with workflow management systems
  - Large-scale and repetitive operational tasks can be automated
- Interfaced to FTS for transfer execution and possibly integrated with EGI Workload manager to implement data transfer orchestration linked to workflow executions

Operation federation

Services

	What it does	More info
Accounting	collects, stores, aggregates, and displays usage information about HTC e jobs, IaaS VMs and storage providers	https://accounting.egi.eu/
Monitoring	Periodically monitors IT services and calculates A/R metrics	https://argo.egi.eu/
Helpdesk	Entry point and ticketing system/request tracker for issues concerning EGI services.	https://ggus.eu

Expected Integration

1. **EGI Check-in to facilitate user access** – this will also enable EGI-ACE platform services!
2. Sharing applications via AppDB / CFMFS
3. Sharing data via DataHub / FTS / Rucio
4. Operations services: accounting, monitoring, helpdesk

Thank you!

Contact: **egi-ace-po@mailman.egi.eu**
Website: **www.egi.eu/projects/egi-ace**

EGI Foundation

@EGI_eInfra

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.