KER 1 - The EOSC Compute Platform

Period 1 Review meeting, 24/05/2022

Enol Fernández
EGI-ACE WP3-WP4 leader
Cloud Solutions Manager at EGI Foundation

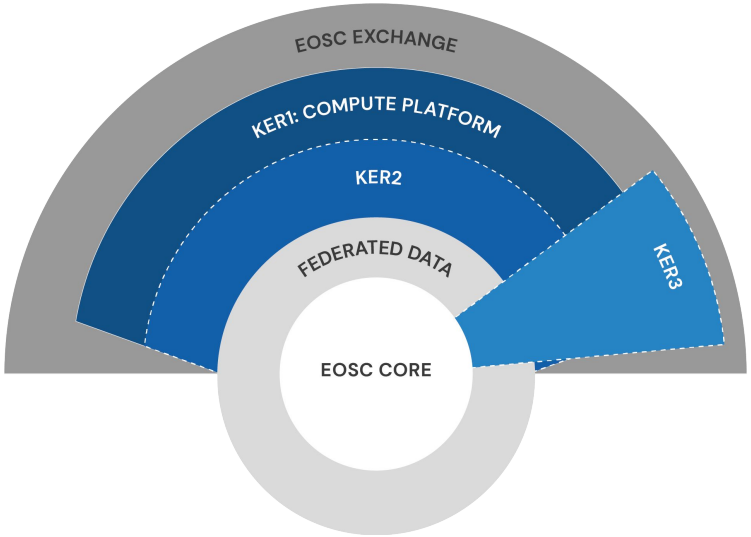
Dissemination level: Public

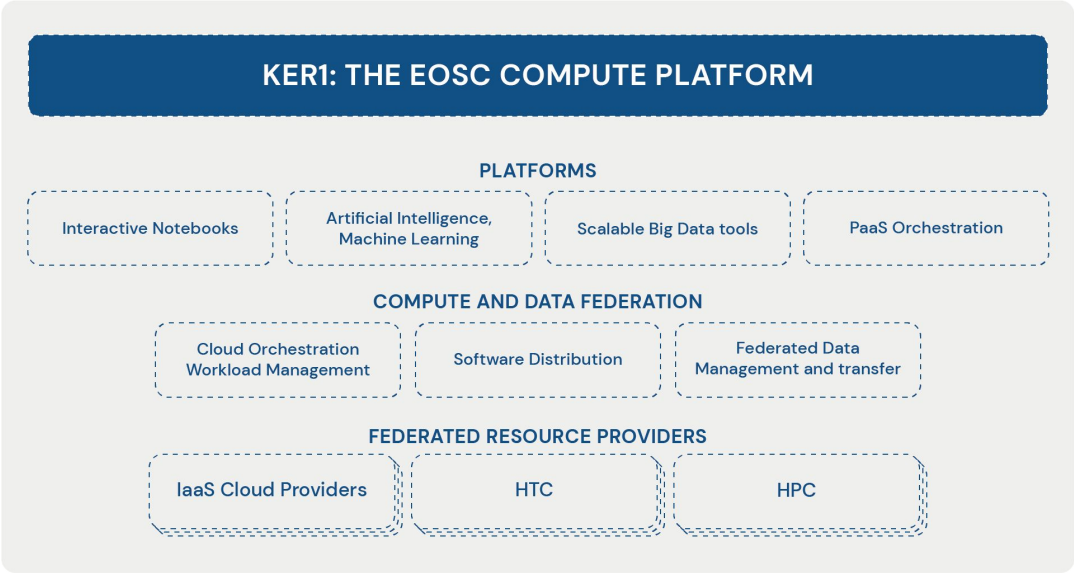
Disclosing Party: Project Consortium

Recipient Party: European Commission

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

KER Overview





Provided by: WP3, WP4, WP6, T7.3

KER1 The EOSC Compute Platform

ABOUT THE KER

The EOSC Compute Platform is an **integrated** and **distributed** computing environment built on a **hybrid** infrastructure composed of **cloud** computing resources, **HTC** sites and **HPC** centres, alongside a wide range of **compute and data management** services supporting research workloads for EOSC users.

USER GROUPS

- Researcher communities and individual researchers
- Providers

KEY VALUE PROPOSITIONS

- Single allocation process for a wide range of computing resources: cloud, HTC and HPC with streamlined support and consultancy
- Distributed processing on an integrated platform with higher level services to facilitate usage and adoption

DISSEMINATION & COMMUNICATION

- 12 webinars
- 17 presentations

EXPLOITATION

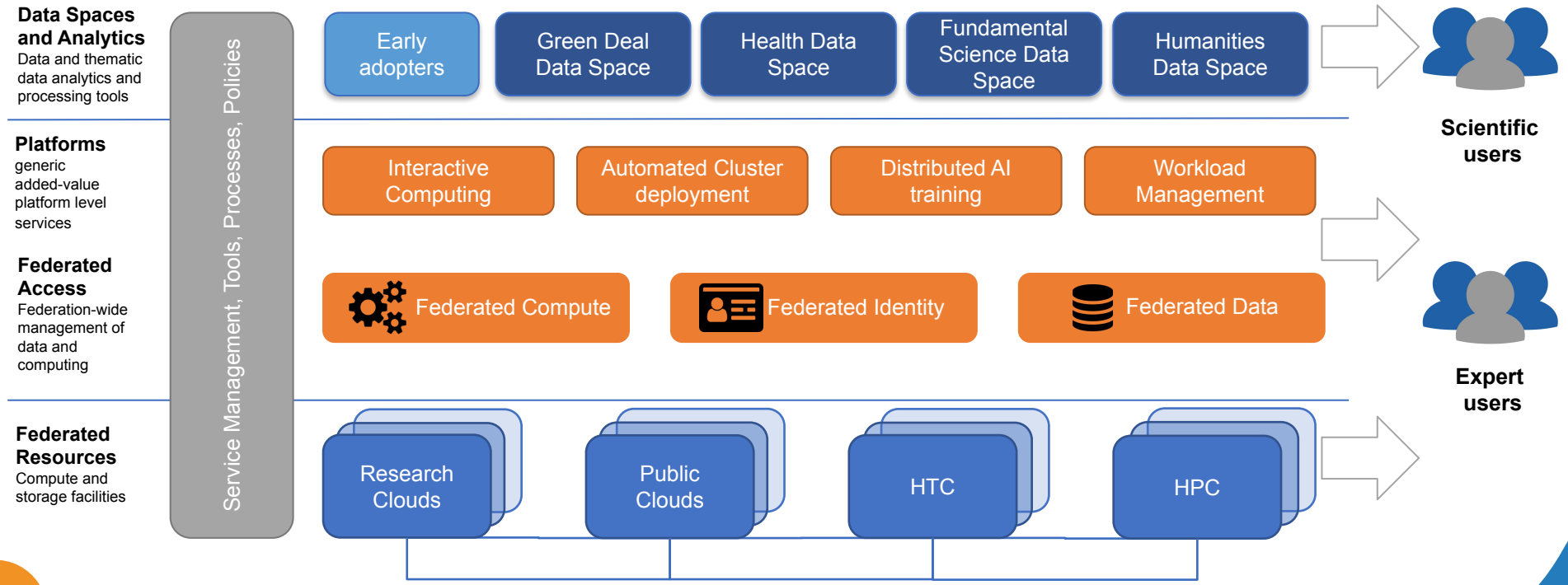
- 7,748 users in 84 communities
- 27 cloud providers, +200 HTC providers, 4 pilot HPC centres

IPR APPROACH

Open Source license for software

CC-BY license for documentation

EGI-ACE tiered service architecture



Federated Resources

Distributed compute and storage facilities deliver CPU, GPU and Storage

- 27 Cloud providers: 15 funded with VA, 23 supporting EGI-ACE use cases
- 200+ HTC providers: 1 funded with VA, 58 supporting EGI-ACE use cases
- 4 pilot HPC centres

Cloud Compute

Cloud Container Compute

High-Throughput Compute

Online Storage

HPC coming soon

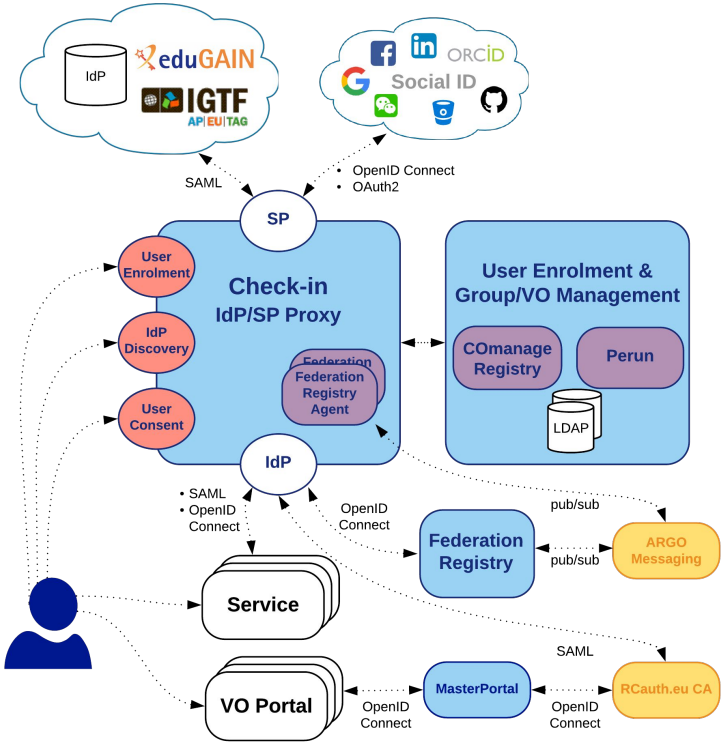
Federated identity

Check-in provides authentication, authorization and user management for the EOSC Compute Platform

Standards based:
- SAML 2.0 / OpenID Connect 1.0 / OAuth 2.0 / LDAP

Interoperable:
- AARC and EOSC AAI compliant
- Support for legacy X.509 services via MasterPortal

Community management:
- Comanage and Perun supported
- Other Community AAIs pluggable



Compute and Data Federation

End-user services and tools for federation-wide management of data and computing

Software distribution: sharing applications across the infrastructure

AppDB

CMVFS

DynamicDNS: hostnames for VMs and services on the EOSC Cloud Compute Platform providers

Compute orchestration: management of compute workloads on cloud, HTC and HPC

Workload Manager

Infrastructure Manager

Data management and transfer: organise and make data available for computing

Data Transfer

Datahub

RUCIO

openRDM .swiss

Platforms

Generic added-value platform level services

Notebooks: Interactive Jupyter Notebooks and reproducible computing environments with Binder

INDIGO-PaaS: Orchestration cloud resources using TOSCA standard with automatic selection of providers

DODAS: On-demand Distributed and customisable data analytics platform

DEEPaaS: Automated training of ML/AI models on the computing resources of the EOSC Compute Platform

Value Proposition

Researchers

BEFORE

- Lack of consistent offer of computing resources in EOSC with fragmented and incompatible services
- Lack of expertise for dealing with distributed computing workloads
- Lack of knowledge about available national resources.
- Not enough computing capacity at a single provider/need to process distributed data.

AFTER

- Single allocation process for a wide range of computing resources: cloud, HTC and HPC
- Streamlined support and consultancy to realise a use case
- Integrated platform with higher level services to facilitate usage
- Federated Compute and data management services enable distributed processing

Innovation highlights

An integrated infrastructure for research:

- Common identity with Check-in
- Interoperable services
- Integrated with Federating Services (KER2) and EOSC

Expanded infrastructure:

- +4 HPC providers pilots
- +1 new HTC flavour
- + 5 GPU providers (2 with VA)
- +2 non-EU providers

Higher level services:

- +11 new services for compute and data management
- + 3 improved services reaching TRL8-9

Service delivery:

- +4 new services in EOSC Marketplace
- Combination of VA, local funds & external projects funds
- Demo and piloting allocations for selected services

Expanding the infrastructure: HPC

First version of the HPC integration handbook (M7.3, 23 downloads since Feb)

Support for hybrid HPC-cloud workloads with:

- Federated identity – use same account in all EOSC Compute Platform systems
- Portable execution with udocker
- Operational integration: accounting

D7.3 HPC integration handbook (June)

- Data Transfer
- Operational integration: monitoring
- Presence in EOSC marketplace

Dissemination and Communication

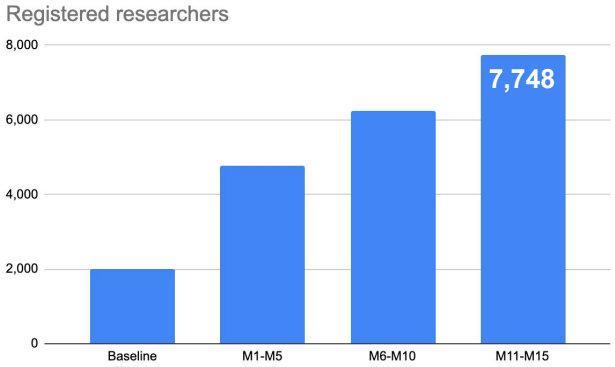
12 Webinars - 370 participants

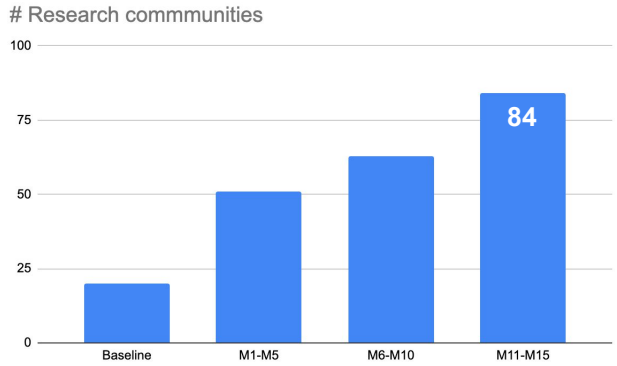
- 3 on Federated Resources layer
- 5 on Check-in & Compute and Data Federation layer
- 4 on Platforms layer

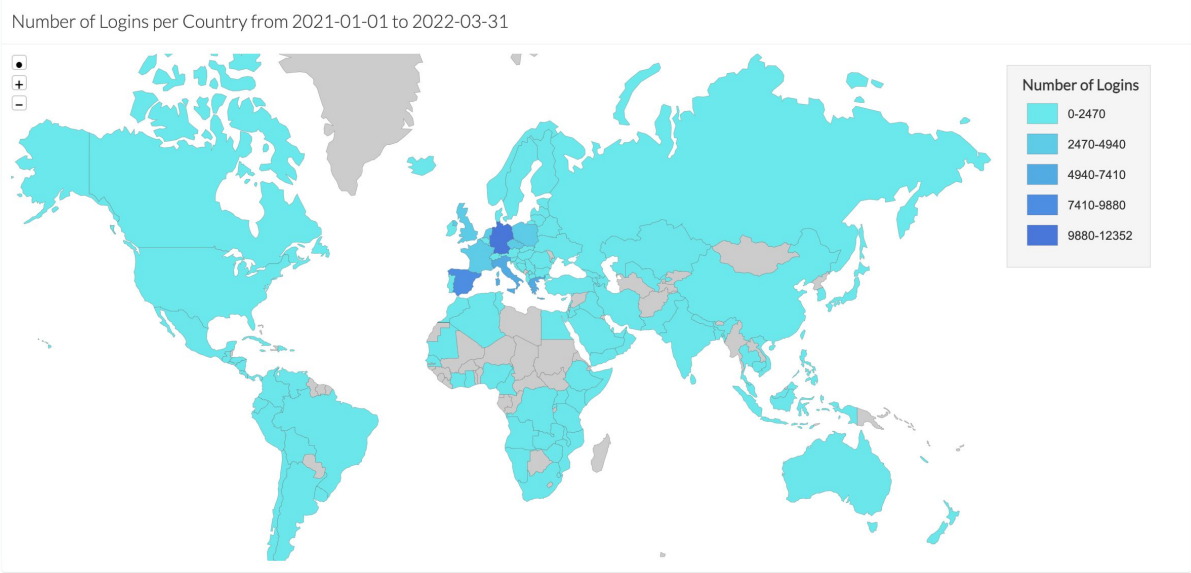
17 presentations at conferences and workshops – estimated reach +900 attendees

- EGI Conference 2021
- EGI-ACE Communities Workshops (one in 2021, one in 2022)
- EOSC Future Ask me anything sessions (3 events)

Exploitation: Users

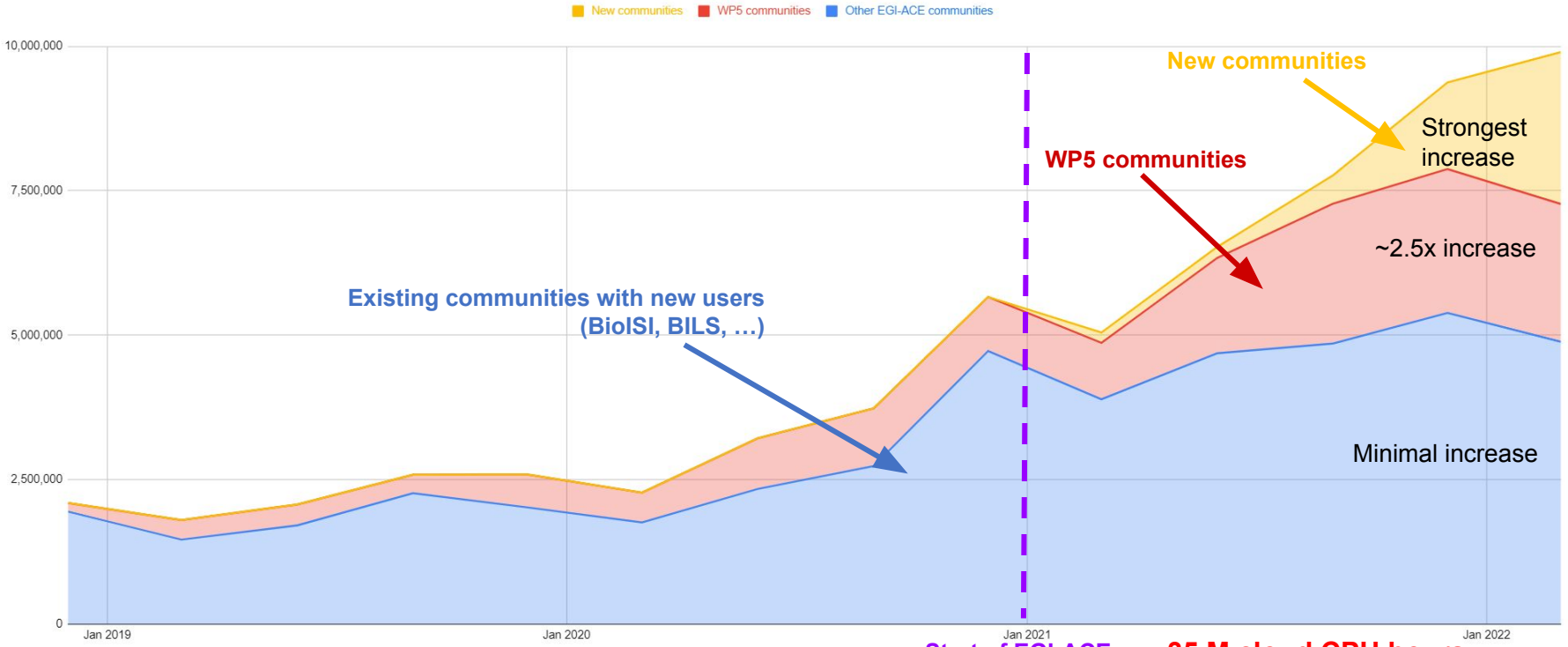






102 countries (+170%)
4484 logins per month (+299%)

Impact: Cloud use by EGI-ACE communities



35 M cloud CPU hours delivered since Jan 2021
(VA+national+pay4use)

Impact: Capacity requests and delivery (WP3)

Service	VA at M15	% VA consumption	Target at M30	Use cases requests
CPU Cloud Compute (15 installations)	19,684,800	26.9%	73,190,800	73,460,353
CPU HTC (1 installation)	22,505	0.41%	5,500,000	4,083,330
Storage (12 installations)	5,536	11.28%	49,100	122,363
GPU (2 installations)	29,213	11.77%	248,200	69,318

CPU

- 20 M CPU hours delivered with local and external project funds for supporting use cases

Storage

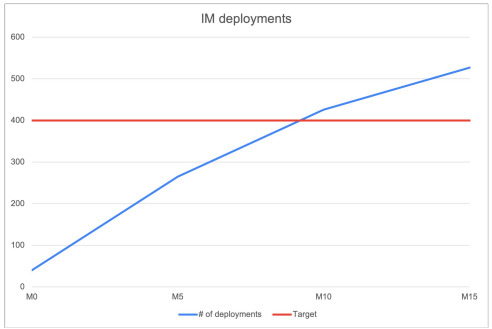
- Several additional data space will become operational in Y2, usage is expected to increase over the next period

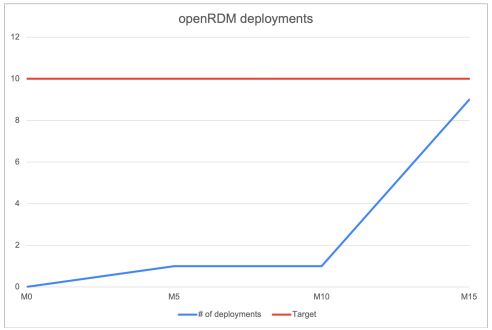
GPU

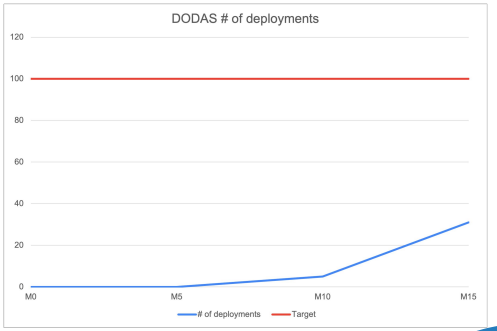
- Promotion campaign started recently to increase usage: several use cases in the pipeline with GPU needs that will increase the requested capacity

VA performance

	Above target	Within target	Below target
WP3	Infrastructure Manager DynamicDNS	AppDB	
WP4	Notebooks	DIRAC	DEEPaaS DODAS
WP6	Check-in	PERUN OpenRDM Onedata CVMFS EC3	MasterPortal RUCIO FTS PaaS Orchestrator





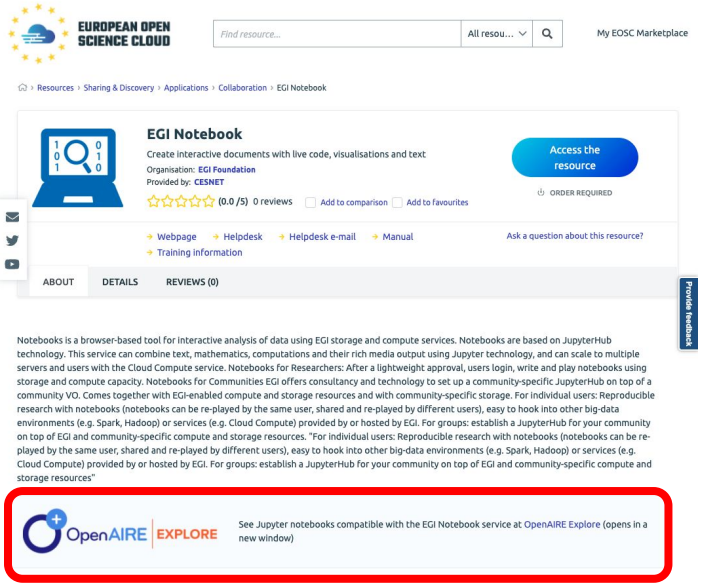


Impact: KPIs

KPI	Value (M15)	Target (M30)
# EGI-ACE services part of the EOSC Portal	43	42
# non-EU-providers integrated in the EOSC Compute Platform	2 (CNIC, IDIA)	7
# Commercial partners involved in use case support	0	2
# requests received	86 marketplace orders 30 use-cases from open call 2 EOSC Future	250

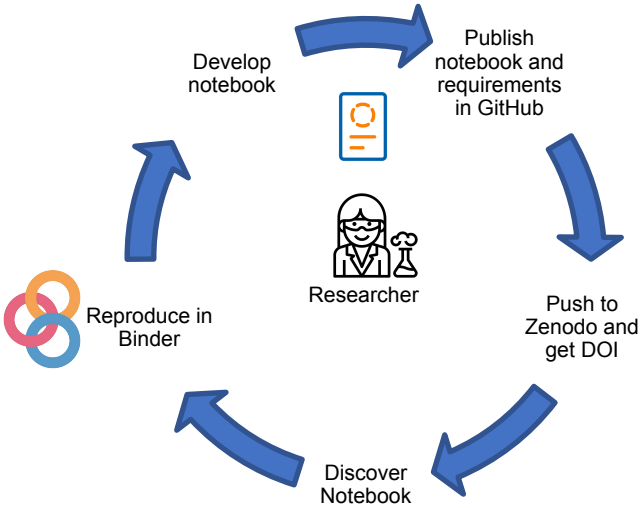
Success Stories: Notebooks

An EOSC Open science cycle by EGI-ACE, OpenAIRE, EUDAT, EOSC Future



Interactive, user friendly, browser-based platform

- Access data from **DataHub** and **B2DROP** seamlessly
- Cite and share notebooks via **Zenodo/B2SHARE** discover them in **OpenAIRE Explore**
- Reproducible computing environments with Binder



Future Plan

Improve integration and interoperability between services

Monitor VA consumption

- WP3 CPU/GPU/Storage
  - Reassessment of communities requests (ongoing) to ensure services uptake
  - Redistribution of load to ensure fair distribution among providers

Increase VA usage

- WP4
  - Targeted dissemination activities: DODAS for ESCAPE users, DEEP for AI/ML use cases. New use cases already in the pipeline.
- WP6
  - Engage with communities testing RUCIO and FTS on premises (e.g. EISCAT_3D, LOFAR), then switch to the available WP6 installations for their production activities
  - FTS as Data Transfer pilot for EOSC-Future
  - Boost MasterPortal usage via DIRAC integration (soon in production)
  - PaaS Orchestrator usage expected to increase with SeaDataNet Data Space becoming operational

Final take-away from KER1

17 services on-boarded in EOSC

Hybrid: 27 cloud, +200 HTC, and 4 (pilot) HPC providers

Supported +7,000 users from 84 communities

~42M CPU hours delivered:
20M with VA, 20M with local and 2M with external project funds

Federated: 14 services to manage compute and data in a distributed infrastructure

Integrated: common identity and interoperable service offer

219M CPU hours allocated to 67 communities

All installations receiving new usage from EOSC (65% installations beyond/within target)

Thank you!

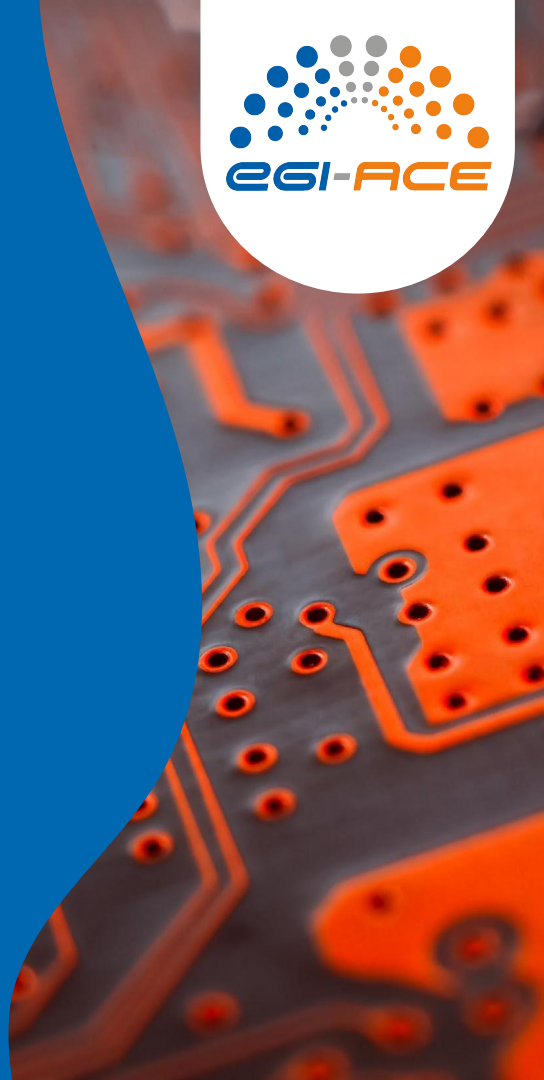
Contact: **egi-ace-po@mailman.egi.eu**
Website: **www.egi.eu/projects/egi-ace**

EGI Foundation

@EGI_eInfra

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

Backup slide



Value Proposition

Providers

BEFORE

- No integration with EOSC
- Support restricted to local communities
- Lack of relevant applications or data for national users
- Limited toolset for exploiting resources in novel ways

AFTER

- Become part of EOSC via a single entry point
- Enable support of international communities and users from EOSC
- Facilitate sharing relevant applications and data among providers
- Ecosystem of higher level tools to bring new usage modes to existing computing and storage resources.

Impact metrics

Impact indicator	Value (M15)	Target (M30)
# of notebook/binder sessions # of new Galaxy users # of new VIP users	520 8,000 22	500 30,000 300
# Countries integrated	2 (CN, ZA)	7
# Cross-regional federated cloud pilots	1 (EISCAT_3D)	3
# Users of non EU providers	12 (EISCAT_3D)	-

An integrated platform - before EGI-ACE

	Check-in	Cloud Compute	HTC	Online Storage	AppDB	IM	Dynamic DNS	Notebooks	Workload Manager	DEEP	DODAS	Rucio	FTS	DataHub	CVMFS	openRDM	EC3	Paas
Check-in																		
Cloud Compute																		
HTC																		
Online Storage																		
AppDB																		
IM																		
Dynamic DNS																		
Notebooks																		
Workload Manager																		
DEEP																		
DODAS																		
Rucio																		
FTS																		
DataHub																		
CVMFS																		
openRDM																		
EC3																		
PaaS																		

Integrated In progress As a client

An integrated platform - at M15

	Check-in	Cloud Compute	HTC	Online Storage	AppDB	IM	Dynamic DNS	Notebooks	Workload Manager	DEEP	DODAS	Rucio	FTS	DataHub	CVMFS	openRDM	EC3	Paas
Check-in		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Cloud Compute	■			■	■	■	■	■	■	■	■			■	■	■	■	■
HTC	■			■					■						■			
Online Storage	■	■	■			■		■	■	■	■	■	■	■		■	■	■
AppDB	■	■				■											■	
IM	■	■		■	■		■	■		■	■						■	■
Dynamic DNS	■	■				■											■	■
Notebooks	■	■		■		■			■	■	■			■	■		■	■
Workload Manager	■	■	■	■				■							■			
DEEP	■	■		■		■		■										■
DODAS	■	■		■		■		■										■
Rucio	■			■					■				■					■
FTS	■		■	■								■						
DataHub	■	■						■										■
CVMFS	■	■	■					■	■									
openRDM	■	■		■														
EC3	■	■		■	■	■	■	■										
PaaS	■	■		■		■	■	■		■	■	■		■				

■ Integrated ■ In progress ■ As a client

KER1 The EOSC Compute Platform

BRIEF DESCRIPTION	**The EOSC Compute Platform is a distributed computing environment built on a hybrid infrastructure** ● Federated Cloud, High-throughput computing (HTC) High-Performance Computing providers + TRL8-9 higher level services to ease the deployment and operation of complex research workflows, applications, data spaces on the infrastructure ● User support and training ● Integrated with EOSC ● Free at the point of use services
USER GROUP	● Researcher communities and individual researchers ● Providers

VA CPU

Installation	VA at M15	% VA consumption	Target at M30	Use cases requests
CESGA-CPU	1,319,781	29.33%	4,500,000	9,171,808
MetaCentrumCloud - CPU	1,570,581	17.93%	8,760,000	14,884,089
CLOUDIFIN-CPU	2,451,907	49.04%	5,000,000	5,513,760
IFCA-LCG2-CPU	4,314,449	172.58%	2,500,000	6,917,280
CYFRONET-CLOUD-CPU	207,304	2.44%	8,500,000	1,093,200
DESY-FedCloud	20,302	1.02%	2,000,000	843,648
EGI - GSIOS	72,264	3.61%	2,000,000	1,093,200
IICT-BAS-CPU	0	0.00%	7,708,800	0
EGI-IISAS-CPU	901,816	14.71%	6,132,000	3,245,760
IN2P3-IRES-CPU	737,500	12.03%	6,132,000	1,913,568
INCD-Lisbon (NCG)-CPU	1,888,486	61.59%	3,066,000	1,788,480
INFN-BARI-CPU	3,415,879	77.99%	4,380,000	9,014,760
INFN-CNAF-CPU	520,426	11.88%	4,380,000	919,296
SCAI FedCloud v2	0	0.00%	2,000,000	5,947,008
Data Processing Compute	22,505	0.41%	5,500,000	4,083,330
TR-FC1-ULAKBIM - CP	2,264,105	36.92%	6,132,000	11,114,496
TOTALS	**19,707,305**	**25.04%**	**78,690,800**	**77,543,683**

- Requested capacity from use cases already near the M30 target for the whole set of installations
- Still their consumption rate needs to increase over the next period
- Use cases need to be better distributed over the available providers
- Extra +20 M CPU hours delivered with local and external project funds for supporting use cases

VA Storage

Installation	M15	% VA consumption	Target at M30	Use cases requests
CESGA-Storage	281	29.60%	950	109,376
MetaCentrumCloud - Storage	25	0.39%	6,500	714
CLOUDIFIN-Storage	1	0.06%	12,000	1
IFCA-LCG2-Storage	25	1.47%	1,700	33
CYFRONET-CLOUD-Storage	1	0.03%	4,500	0
IN2P3-IRES-Storage	24	0.57%	4,200	1,728
INCD-Lisbon (NCG)-Storage	200	8.15%	2,450	46
INFN-BARI-Storage	24	1.45%	1,650	906
INFN-CNAF-Storage	217	4.39%	4,950	60
TR-FC1-ULAKBIM-Storage	2,481	59.07%	4,200	8,750
SURF	1,500	59.52%	2,520	2,500
SURF	750	21.55%	3,480	1,500
TOTALS	**29,213**	**11.28%**	**248,200**	**122,363**

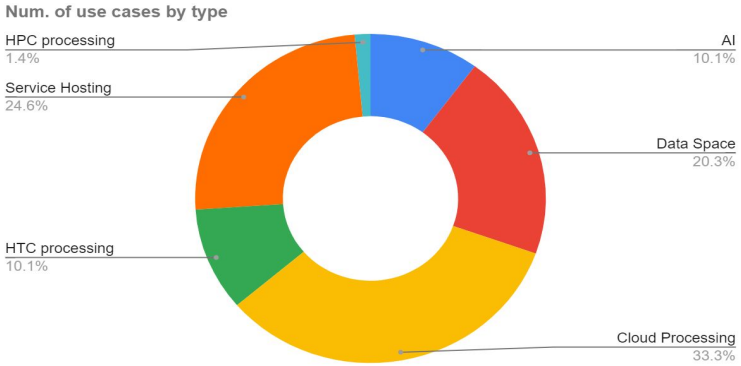
- Several Data Spaces becoming operational during Y2, requests and usage is expected to increase over the next period
- Ongoing use cases revision to ensure service uptake and adjustment of capacity to be delivered

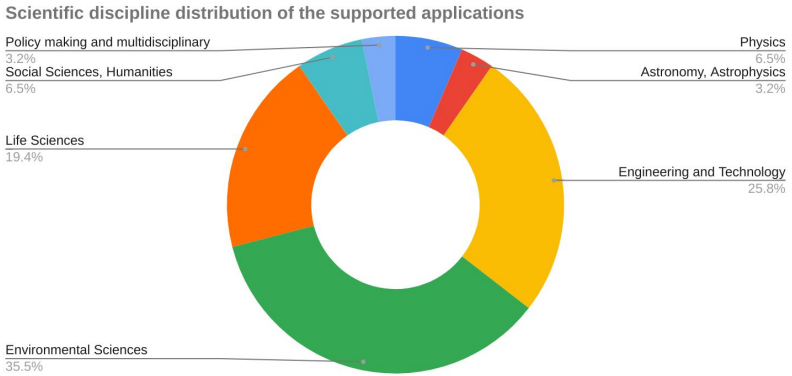
VA GPU

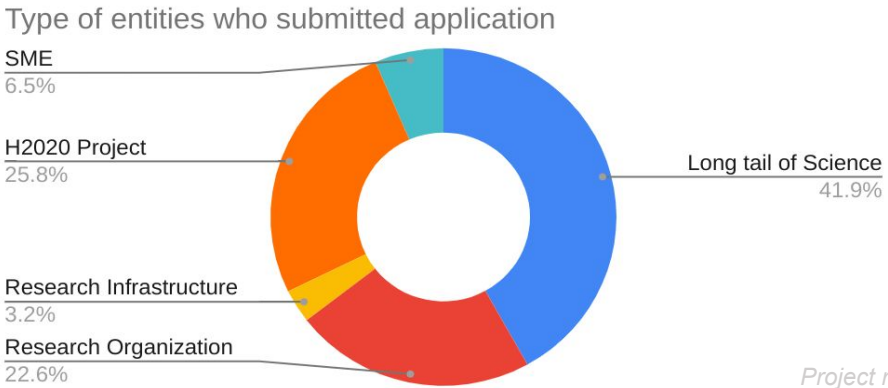
Installation	VA at M15	% VA consumption	Target at M30	Use cases requests
MetaCentrumCloud - GPU	20,448	10.00%	204,400	43,782
INFN-CNAF-GPU	8,765	20.01%	43,800	25,536
TOTALS	**29,213**	**11.77%**	**248,200**	**69,318**

- Promotion campaign started recently to increase usage
- Several use cases in the pipeline with GPU needs that will increase the requested capacity
- CNAF joined in Q3 2021, delaying the consumption

Exploitation: Communities







IMPACT: WP3 enabling services

Highlights

EGI Notebooks have been used by 273 users (245% growth over baseline in previous 15 months) belonging to the vo.notebook.egi.eu and the vo.access.egi.eu communities, open to individual users, and to four discipline/project specific communities: biomed, auger, eiscat_3d and vo.reliance-project.eu.

EGI DIRAC has been used by 18 different communities (106% growth over baseline in previous 15 months), with 2 new communities incorporated during the project. DIRAC has supported 735 users belonging to these communities (118% over baseline).

The **CSIC DEEP training facility** has received usage from one of the use cases from the 4th Open Call communities. For the **LIP DEEP training facility**, there are initial discussions with one use from the last Open Call to start its support.

DODAS was used by one community: fermi-lat.infn.it with a total of 3 users (8% over baseline).

IMPACT: WP4

Highlights

EGI Notebooks have been used by 273 users (245% growth over baseline in previous 15 months) belonging to the vo.notebook.egi.eu and the vo.access.egi.eu communities, open to individual users, and to four discipline/project specific communities: biomed, auger, eiscat_3d and vo.reliance-project.eu.

EGI DIRAC has been used by 18 different communities (106% growth over baseline in previous 15 months), with 2 new communities incorporated during the project. DIRAC has supported 735 users belonging to these communities (118% over baseline).

The **CSIC DEEP training facility** has received usage from one of the use cases from the 4th Open Call communities. For the **LIP DEEP training facility**, there are initial discussions with one use from the last Open Call to start its support.

DODAS was used by one community: fermi-lat.infn.it with a total of 3 users (8% over baseline).

WP6 VA Metrics (I)

Highlights

EGI Check-in has been used by 148 service providers and 7,748 users, a 420% and 387% increase in 15 months.

EGI FTS didn't see so far an increase of users communities during the project, (4 VOs using the service in the M06-M10 period). There is a new use case in the pipeline that is expected to improve this in the second period.

EGI CVFMS has been used by 2 new communities, with the number of hosted files and storage occupied increased by 26% and 23 % respectively.

EGI Rucio onboarded 1 new community using the service since the project starting date.

WP6 VA Metrics (II)

Highlights

EGI Onedata has been integrated into 5 new community use cases, with new providers installed in Turkey, France, Poland and Czech Republic.

Master Portal has been integrated by 2 new community portals.

Orchestrator has been integrated by 1 community representing an EGI Data Space (SeaDataNet WebODV Data Analysis), reaching 47 deployments .

PERUN is now used by 577 new users.

EC3 has been used both by WP5 Data Spaces installations and use case applications, with an average of 22 deployments orchestrated per year.

openRDM has been deployed in EGI cloud and support to 9 research institutes has been given for their on-premise installations.

DISSEMINATION AND COMMUNICATION

WP3, WP4 and WP6 Installation Webinars

Name	Date	Stats
openRDM	12/01/2022	Num. of Registrations: 29
		Num. of Participants: 20
		Num. of Countries: 9
Managing Singularity, Docker and udocker containers, and Kubernetes clusters in the EGI Cloud	28/04/2021	Num. of Registrations: 54
		Num. of Participants: 39
		Num. of Countries: 14
Providing controlled access to distributed resources and services with EGI Check-in: the provider perspective	05/05/2021	Num. of Registrations: 43
		Num. of Participants: 35
		Num. of Countries: 13
Access and analyze data from the	Analyze your data using DODAS generated cluster	22/09/2021	Num. of Registrations: 15
			Num. of Participants: 11
			Num. of Countries: 8
Monitoring services with ARGO	Using EGI Cloud infrastructure with fedcloudclient	29/09/2021	Num. of Registrations: 38
			Num. of Participants: 29
			Num. of Countries: 13
Deploying virtual infrastructures v	Data Management in EGI with Rucio and FTS	06/10/2021	Num. of Registrations: 77
			Num. of Participants: 50
Using Dynamic DNS service in EG			Num. of Countries: 17
	How to orchestrate services in the EOSC Compute Platform with the INDIGO PaaS	27/10/2021	Num. of Registrations: 57
			Num. of Participants: 33
Running containers in user space v			Num. of Countries: 8
	The Role of Research in Data Spaces and Data Ecosystems	24/11/2021	Num. of Registrations: 50
			Num. of Participants: 32
			Num. of Countries: 15
	How to train your AI model in EOSC	01/12/2021	Num. of Registrations: 47
			Num. of Participants: 32
			Num. of Countries: 9

Impact:

WP3

Installation	Metric	Variation
AppDB	# of users	+111%
IM	# of users	+560%
DynamicDNS	# of users	+158%

WP4

Installation	Metric	Variation
Notebooks	# of users	+215%
DIRAC	# of communities	+108%
DEEP	# of communities	-94%
DODAS	# of communities	-90%

Impact: VA metrics – WP6

Installation	Metric	Variation
Check-in	# of users	+340%
	# of providers	+420%
MasterPortal	# of users	-86%
Perun	# of users	-85%
PaaS Orchestrator	# of communities	-66%
EC3	# of users	-40%

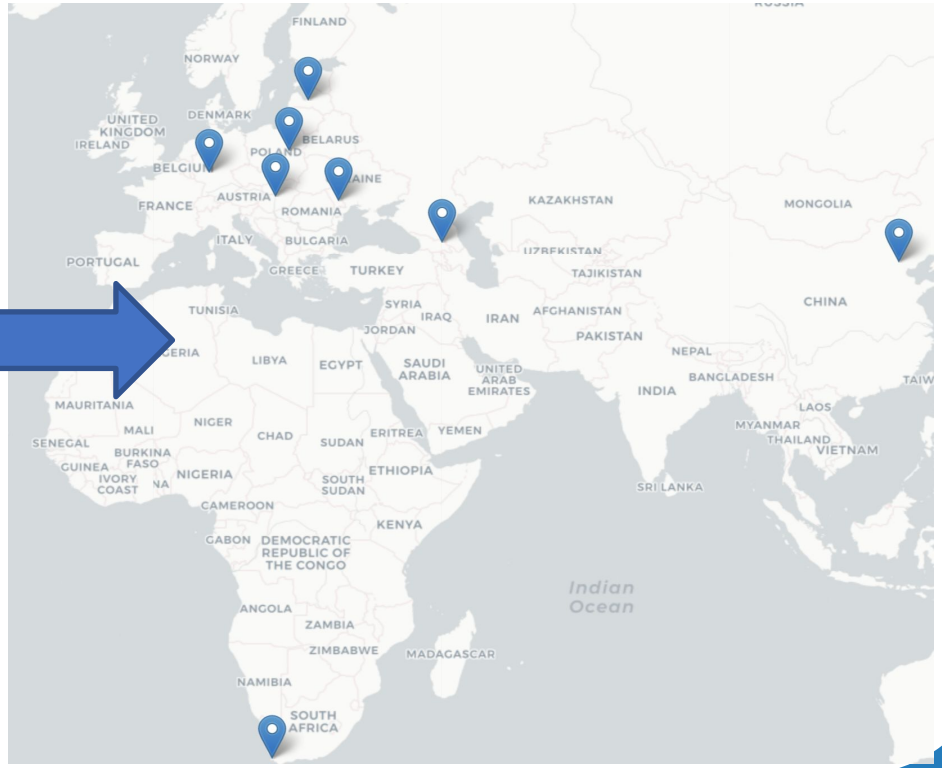
Installation	Metric	Variation
RUCIO	# of users	+150%
Onedata	# of communities	+166%
FTS	# of communities	-50%
openRDM	# of deployments	+900%
CVMFS	# of communities	+109%

Exploitation: Cloud integration

Check-in – single sign on across services and providers

DataHub – replication of scientific data to national clouds

AppDB – distribution of applications & services to national clouds



Federated Resources

Distributed Compute and storage facilities deliver CPU, GPU and Storage

- 27 (15 with VA funding in EGI-ACE) Cloud providers
- 200+ (1 with VA funding in EGI-ACE) HTC providers
- 4 pilot HPC centres

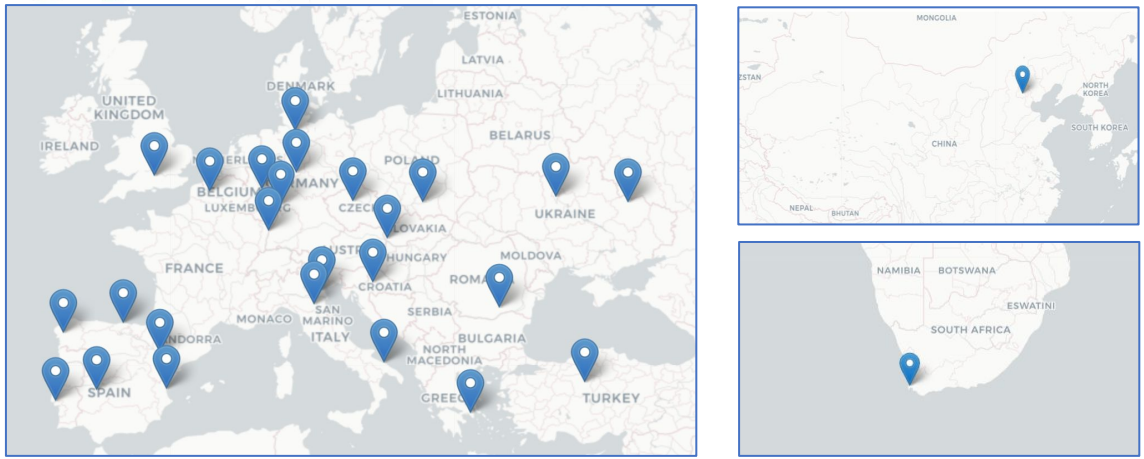
Cloud Compute

Cloud Container Compute

High-Throughput Compute

Online Storage

New in 2021: INFN CNAF, grnet, GSI, 中国科技云 China Science & Technology Cloud

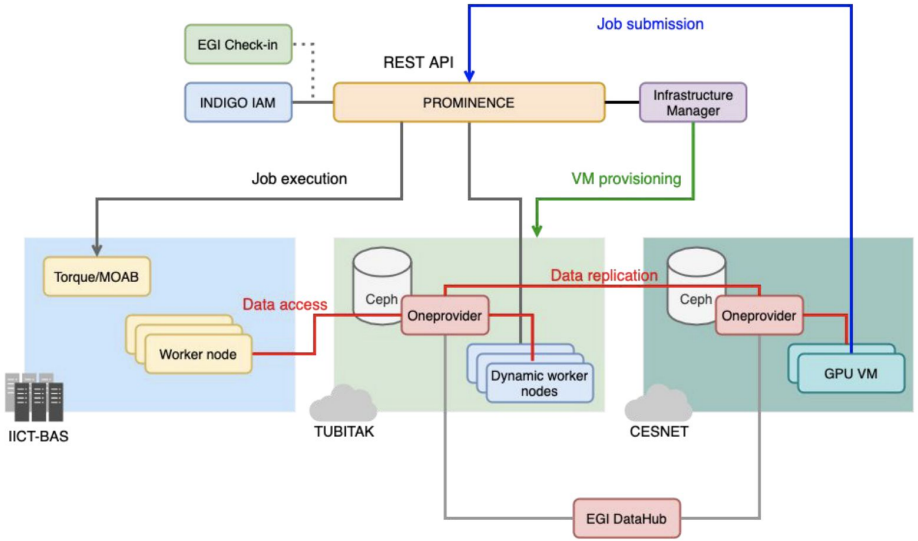


New in 2022: IDIA Inter-University Institute for Data Intensive Astronomy

SUCCESS STORIES EXAMPLES

Prominence Data Space

- Access to WP3 Clouds (CESNET and TUBITAK), HPC (IICT-BAS) and GPU (CESNET)
- Usage of WP4 Infrastructure Manager for cloud orchestration
- Integration with WP6 EGI Check-in almost completed
- Access to 60TB storage via WP6 EGI Onedata at 2 cloud sites (CESNET and TUBITAK)



Andrew Lahiff, UK Atomic Energy Authority
EGI Conference 2021