KER 3: Research data spaces and processing tools for EOSC

Period 1 Review meeting, 24/05/2022

Giuseppe La Rocca
EGI-ACE WP5 Leader
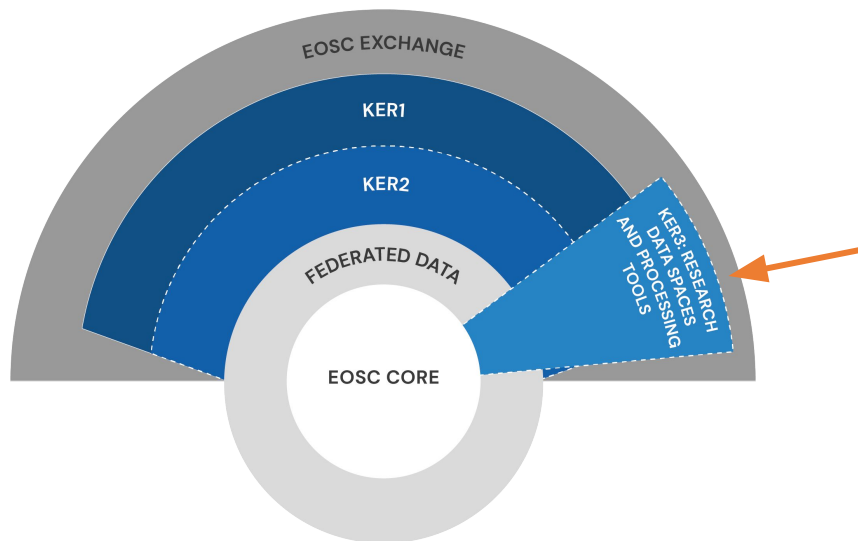Community Support Team Lead. at EGI Foundation

Dissemination level: Public

Disclosing Party: Project Consortium

Recipient Party: European Commission

KER OVERVIEW

KER3: RESEARCH DATA SPACES AND PROCESSING TOOLS FOR EOSC

- Health and Medicine
- Social Sciences and Humanities
- Climate Research
- Energy and Physical Sciences
- Environmental Sciences
- New Thematic Services

- **Provided by:** WP5, T2.3

KER3 OVERVIEW

ABOUT THE KER
- Thematic services (Data Spaces and Processing Platforms) from EGI-ACE to the EOSC Exchange
- Leveraging the EOSC Compute Platform (KER1) to deliver scalable environments for data analytics

USER GROUPS
- Researchers
- Scientific communities
- Research Infrastructures

KEY VALUE PROPOSITIONS
- Reduce development time for digital science, and contribute to the cross-fertilisation of knowledge across disciplines
- Facilitating cross-domain sharing and exploitation of research datasets

DISSEMINATION & COMMUNICATION
- **28** training events
- **2** webinars

EXPLOITATION
- **22** Thematic Services from Health Common, Green Deal, Fundamental research and Social Science and Humanities domains
- **49,000** users served
- **20,109,077** CPU/h consumed

IPR APPROACH
- Open-source licenses SW
- Documentation is licensed under **CC-BY**

Motivations

User Group: Researchers and scientific communities, and RIs

BEFORE

- Fragmentation to access datasets, tools and resources to support domain-specific research

- No inter-disciplinary research

- Lack of computing resources to support European researchers

AFTER

- Data Spaces facilitate research and development activities

- Reduce development time for digital science, and contribute to the cross-fertilisation of knowledge across disciplines.

- Facilitating cross-domain sharing and exploitation of research datasets

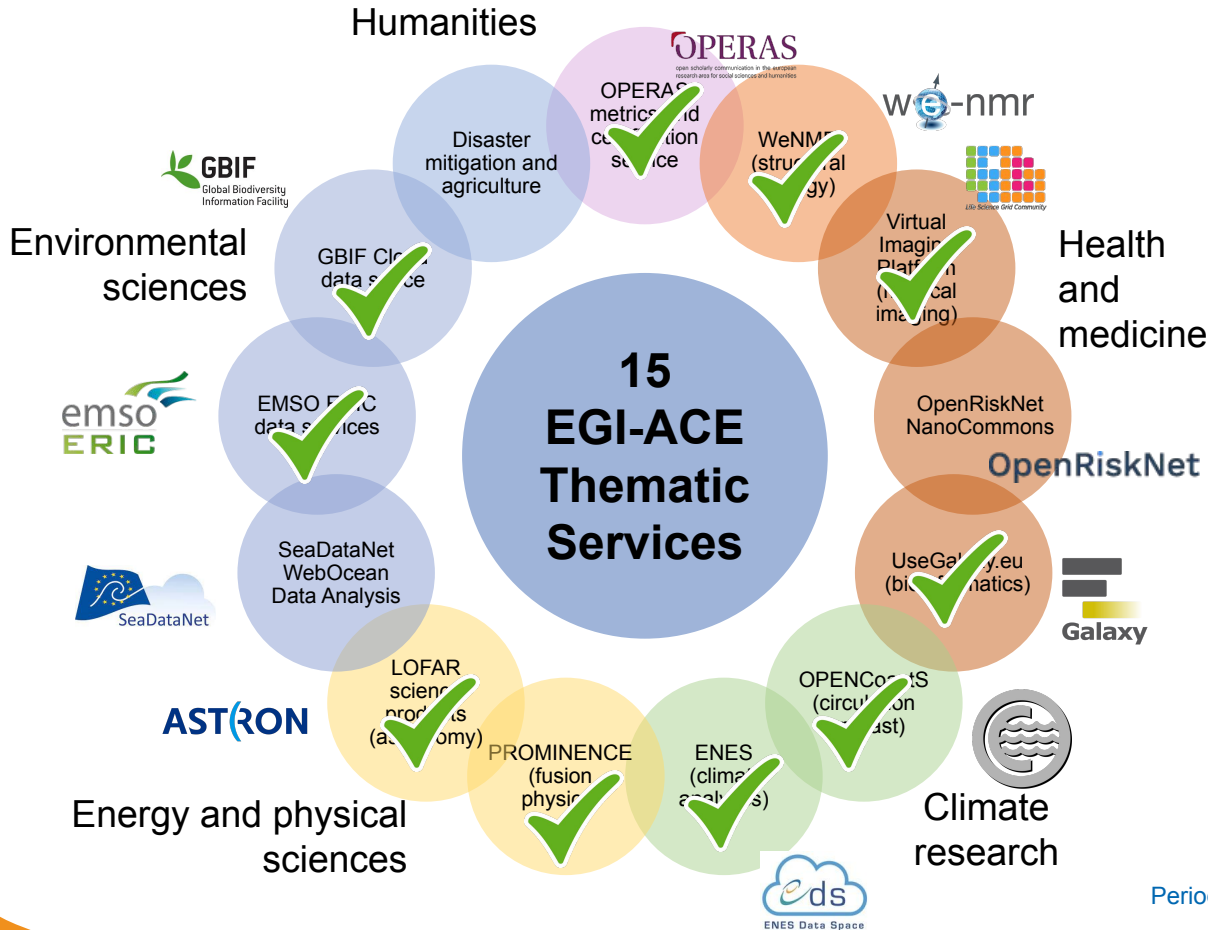# Where the Thematic Services come from?

From the project → This talk
- **WP5:** **8** Data Spaces + **7** Processing Tools
- **T2.3:** **7** Early Adopters

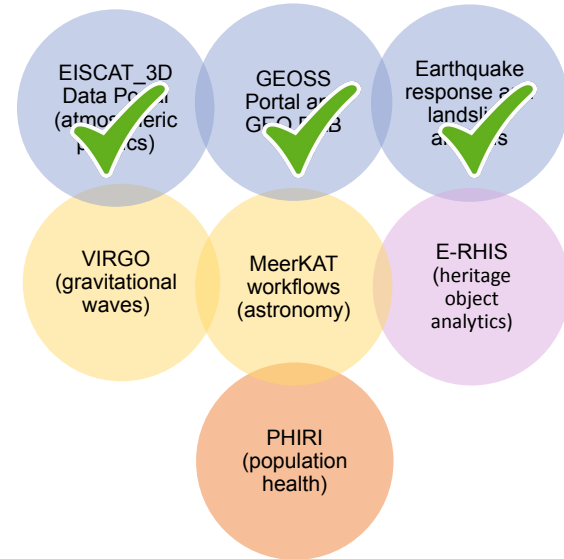***External*** → See next talk
- New communities engaged by EGI-ACE (T2.3) via the EGI-ACE Open Call
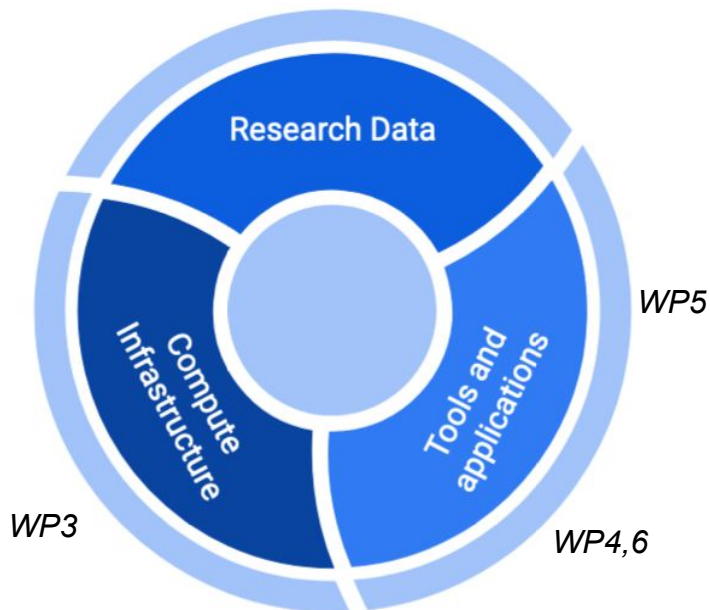
Thematic Services from EGI-ACE

Humanities

Environmental sciences

GBIF Cloud data service

EMSO ERIC data services

SeaDataNet WebOcean Data Analysis

Disaster mitigation and agriculture

OPERAS metrics and certification service

WeNMR (structural biology)

Virtual Imaging Platform (medical imaging)

OpenRiskNet NanoCommons

UseGalaxy.eu (bioinformatics)

15 EGI-ACE Thematic Services

Health and medicine

LOFAR science products (astronomy)

PROMINENCE (fusion physics)

ENES (climate analysis)

OPENCoastS (circulation forecast)

Climate research

Energy and physical sciences

7 Early Adopters

EISCAT_3D Data Portal (atmospheric physics)

GEOSS Portal and GEO DAB

Earthquake response and landslide analysis

VIRGO (gravitational waves)

MeerKAT workflows (astronomy)

E-RHIS (heritage object analytics)

PHIRI (population health)

Already on-boarded in the EOSC Portal

Methodology of supporting Thematic Services

WP5

WP5

WP3

WP4,6

Research Data

Compute Infrastructure

Tools and applications

Supported by federated resources, access and platform services with SLAs

EGI Document 2751-v32

[DocDB Home] [Search] [Last 20 Days] [List Authors] [List Topics] [List Events]
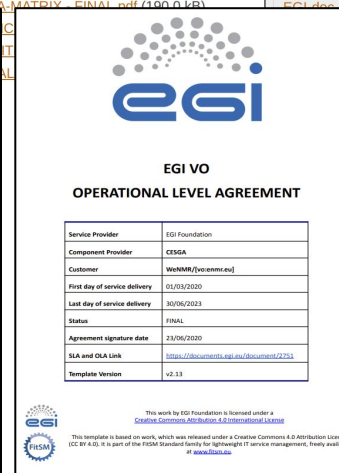
VO SLA WeNMR and VO OLAs WeNMR (Former Mobrain)

Permalink: https://documents.egi.eu/document/2751	**Abstract:** The main objective of WeNMR is to lower barriers for scientists to access modern e-Science solutions from micro to macro scales. The Customer is a consortium represented by the Faculty of Science – Chemistry, Utrecht University. Agreement from 01/01/2016
Document #: EGI-doc-2751-v32	
Document type: Internal document	
Submitted by: Unknown	**Files in Document:**
Updated by: Giuseppe La Rocca	EGI VO WeNMR OLA CESGA - FINAL.pdf (185.3 kB)
Document Created: 16 Feb 2016, 16:43	EGI VO WeNMR OLA CESNET-MCC - FINAL.pdf (198.3 kB)
Contents Revised: 11 Jan 2022, 17:18	EGI VO WeNMR OLA IFCA-LCG2 - FINAL.pdf (189.2 kB)
Metadata Revised: 11 Jan 2022	EGI VO WeNMR OLA INFN-LNL-2 - FINAL.pdf (195.0 kB)
	EGI VO WeNMR OLA INFN-PADOVA-STACK - FINAL.pdf (190.5 kB)
	EGI VO WeNMR OLA NCG-INGRID-PT - FINAL.pdf (185.4 kB)
	EGI VO WeNMR OLA NIKHEF-ELPROD - FINAL.pdf (190.4 kB)
	EGI VO WeNMR OLA SARA-MATRIX - FINAL.pdf (190.0 kB)

Viewable by:
Public document

Modifiable by:
office

Quick Links:
Latest Version

Other Versions:
EGI-doc-2751-v31
06 Jul 2021, 16:47

EGI-doc-2751-v30
18 Jan 2021, 17:03

EGI-doc-2751-v29
17:03

2751-v28
16:56

EGI VO

SERVICE LEVEL AGREEMENT

Customer	WeNMR/VO:enmr.eu
Service Provider	EGI Foundation
First day of service delivery	01/01/2016
Last day of service delivery	30/06/2023
Status	FINAL
Agreement signature date	18/02/2016
Agreement link	https://documents.egi.eu/document/2751
Template Version	v3.18

This work by EGI Foundation is licensed under a Creative Commons Attribution 4.0 International License

This template is based on work, which was released under a Creative Commons 4.0 Attribution License (CC BY 4.0). It is part of the FitSM Standard family for lightweight IT service management, freely available at www.fitsm.eu.

EGI VO

OPERATIONAL LEVEL AGREEMENT

Service Provider	EGI Foundation
Component Provider	CESGA
Customer	WeNMR/[vo:enmr.eu]
First day of service delivery	01/03/2020
Last day of service delivery	30/06/2023
Status	FINAL
Agreement signature date	23/06/2020
SLA and OLA Link	https://documents.egi.eu/document/2751
Template Version	v2.13

This work by EGI Foundation is licensed under a Creative Commons Attribution 4.0 International License

This template is based on work, which was released under a Creative Commons 4.0 Attribution License (CC BY 4.0). It is part of the FitSM Standard family for lightweight IT service management, freely available at www.fitsm.eu.

Methodology of supporting Thematic Services

- **Integrate with Infrastructure Services (WP3)**
  - The OPENCoastS Processing Tool

- **Integrate with Platform Services (WP4)**
  - The EISCAT_3D Early Adopter

- **Integrate with Federated Services (WP6)** DEMO
  - ENES Data Space

OPENCoastS - Coastal Dynamics on-demand forecast

OPENCoastS(+) service assembles on-demand forecast systems for selected coastal areas and generates daily forecasts of the region of interest for 48h
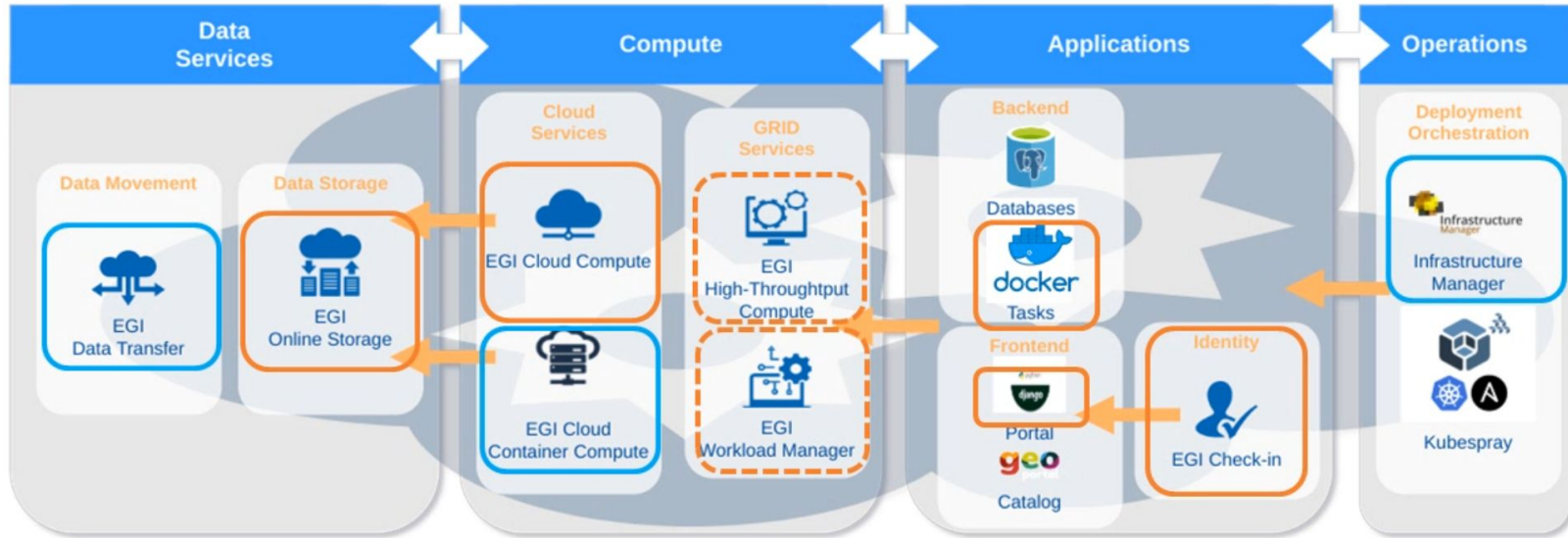
- Full orange - **Integrated** in EGI-ACE
- Dashed orange - Integrated before EGI-ACE
- Full blue - **Planned** for the next part of the project

OPENCoastS - Coastal Dynamics on-demand forecast

Ambition in EGI-ACE:

- Leverage on the EOSC Compute Platform to create a one-stop-shop and robust service for coastal hydrodynamics and faecal contamination modelling and forecasting
- Applicable to any coastline in the world
- Available via a Web portal: Configuration Wizard, Forecast Manager, and Outputs viewer

Impact:

- Total resource consumed : **1,690,033** Cloud CPU/h
- #. of users: **+54%**
- # of country reached: **58**
- # of use cases: **168**
- # of international deployment: **200**
- # of training events: **4**
- Amount of data processed: **4TB**

The EISCAT data federation case

Incoherent scatter radar data fusion and computation

EISCAT-3D radar, next generation incoherent scatter radar system, EISCAT association.
Sanya Incoherent Scatter radar (SYISR), next generation incoherent scatter radar, IGGCAS.
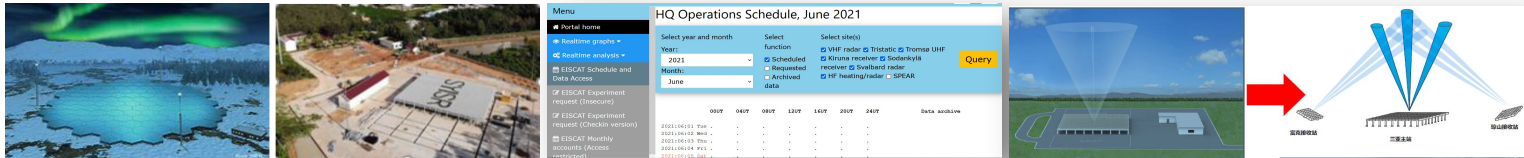
Use Case Scenarios

- EISCAT & SYISR (Meta)data federation
- Federated processing
- On-demand data movement

Key deliverables

- Cross-continental data migration and federated computing
- High-speed and high-capacity experimental data processing
- Data sharing policies and mechanisms

Credits: Häggström and X. Yue (2020)

The EISCAT data federation case

Ambition in EGI-ACE:
- Leverage on the EOSC Compute Platform to:
  - Federate open data collected from different scatter radars in different geographic regions in Europe and China
  - Set-up a federate processing infrastructure where open data can be processed across the different data centres
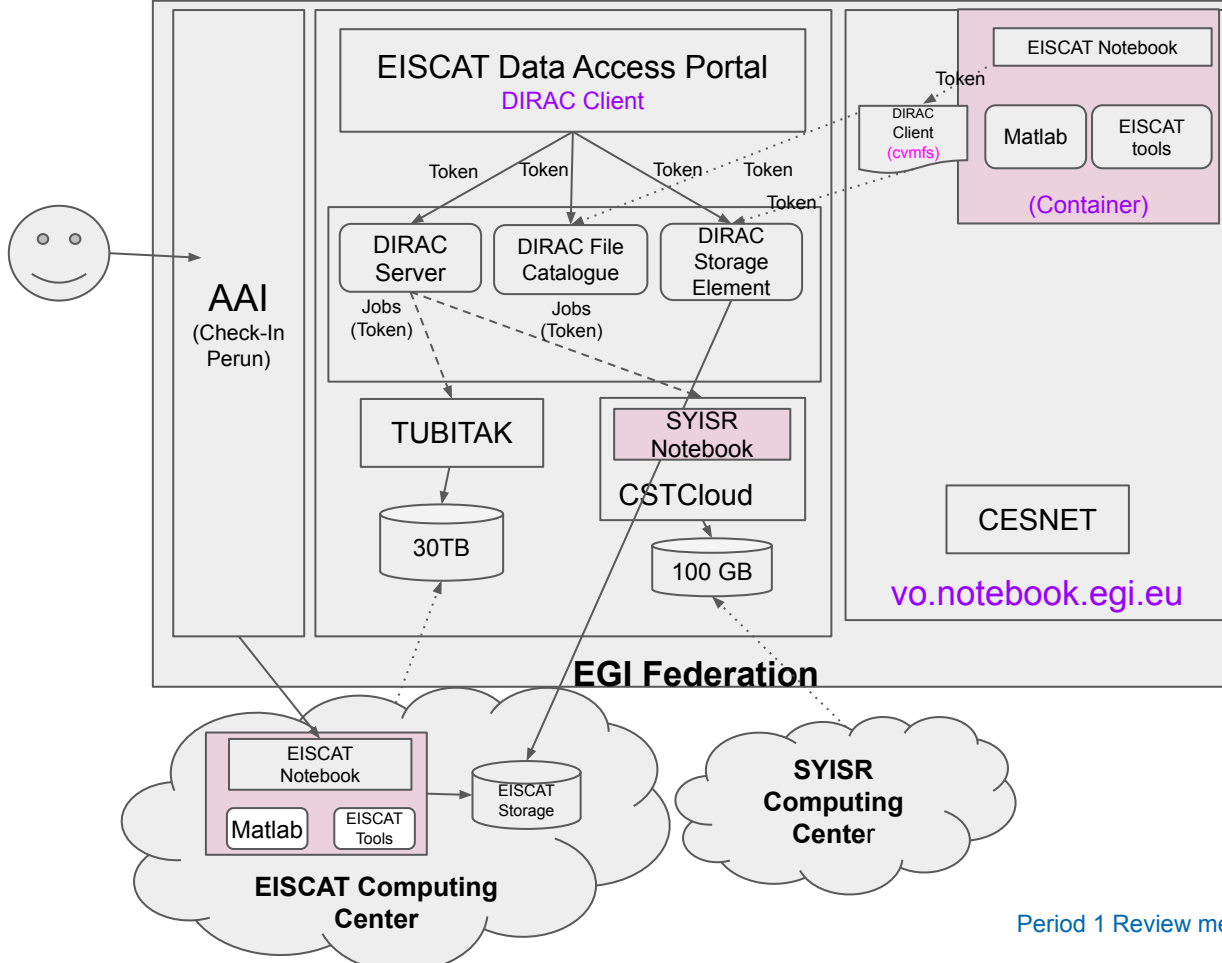
Impact:
- Total resource consumed : **30,781** Cloud CPU/h
  - Starting from Dec. 2021 using CNIC resources
- EGI-ACE Services integrated:
  - EGI Check-in, EGI Cloud Compute,
  - EGI Workload Manager, EGI Notebooks
- Open Data: **30 TB**
- EISCAT user access workshops (use Check-in + Perun)
- # training events: **7**

Integration with the Platform Services (WP4)

Open Research Data served by WP5 Data Spaces

ENES Data Space CMIP6 subset
- Repository: https://esgf-node.ipsl.upmc.fr/search/cmip6-ipsl/
- Datasets: **20-25PB**
- Resources: 172 vCPU cores, 172GB of RAM and 300TB of block storage

GBIF Biodiversity Data Portal
- Repository: https://datos.gbif.es/?lang=en,
- Datasets: **386** datasets (from the GBIF Spanish node), **178.132** species, **10TB**
- Datasets: **286** datasets (from the GBIF Portugal node), **3TB** of datasets

EMSO-ERIC Data Spaces
- Repository: https://erddap.emso.eu,
- Datasets: **10TB** of hydrophone data available
- Resources: 492 vCPU cores, 1.7TB of RAM and 10.6TB of block storage
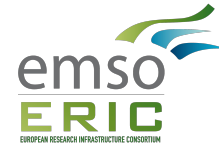
SeaDataNet (require authentication)
- Repository: https://files.seadatanet.org/aggregated_datasets/
- Repository: https://data-argo.ifremer.fr/
- Datasets: **38 GB** of data available
- Resources: 585 vCPU cores, 1.3TB of RAM and 60.8TB of block storage

LOFAR
- Repository: http://spaceweather.astron.nl/SolarKSP/data/dynspec/
- Datasets: **9TB** of data available
- Resources: 4.9M HTC hours and 150TB of block storage

DISSEMINATION AND COMMUNICATION

- **28 training events** were organized by the WP5 partners, with focus on domain researchers (M15):
  - **9** by WeNMR,
  - **7** by EISCAT_3D,
  - **4** by OpenCoastS,
  - **3** by LOFAR,
  - **3** by Virtual Imaging Platform,
  - **1** by ENES, and
  - **1** by EMSO-ERIC.

- **2 webinars** were organized (M15):
  - Virtual Imaging Platform
  - The ENES Data Space

- **1,703** participants in total

EXPLOITATION: Users per country (total = 49,000)

- Austria — 1.7%
- Belgium — 2.0%
- Czech Republic — 0.8%
- France — 8.9%
- Total outside Europe — 33.1%
- Germany — 12.5%
- Greece — 0.4%
- Hungary — 0.5%
- Italy — 4.9%
- Macedonia — 0.0%
- Netherlands — 4.8%
- Poland — 1.3%
- Portugal — 3.1%
- United Kingdom — 12.1%
- Turkey — 1.7%
- Switzerland — 1.5%
- Spain — 9.2%
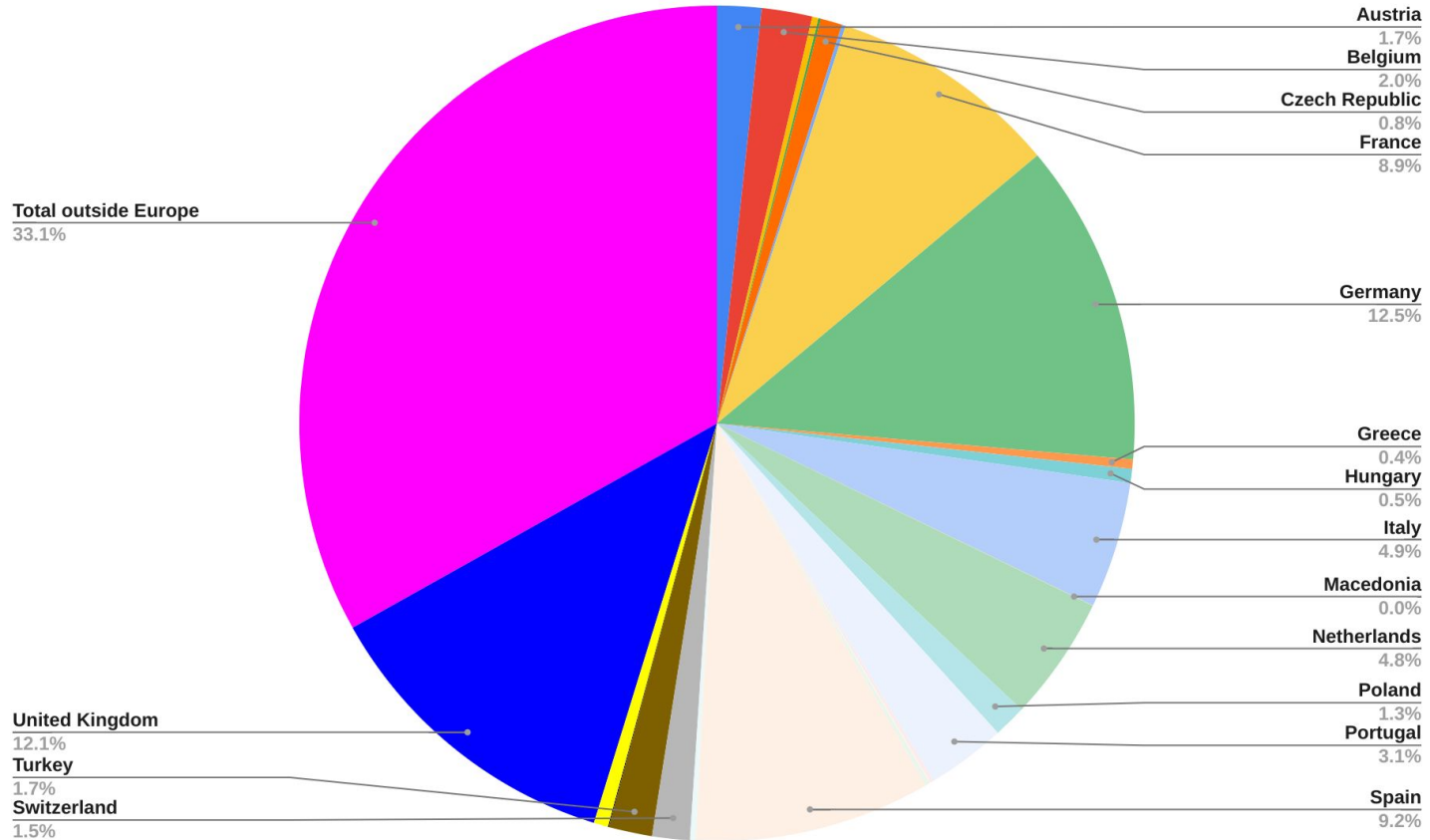
No. of new registered users (during M01-M15)

Thematic Service	Baseline	Metrics at M15	%
WeNMR	4,500	8,056	+41%
VIP	1,000	1,388	+11%
OpenRiskNet/NanoCommons Virtual Environment	100	**In development stage**	
useGalaxy.eu	20,000	48,000	+92%
OPENCoastS	250	316	+54%
ENES Data Space	40	**Registered in EOSC recently**	
PROMINENCE	3	4	+33%
LOFAR Science Product	0	**Registered in EOSC recently**	
SeaDataNet WebOcean Data Analysis	100	**In development stage**	

No. of new registered users (during M01-M15)

Thematic Service	Baseline	Metrics at M15	%
EMSO-ERIC Data Services	200	1,977	+690%
GBIF Cloud data space	8,031	**Registered in EOSC recently**	
Disaster Mitigation & Agriculture	150	**EOSC registration in progress**	

IMPACT: Cloud CPU-hour use by Thematic Services

Legend: New communities · WP5 communities · Other EGI-ACE communities

New communities

WP5 communities

Strongest increase

~2.5x increase

Existing communities with new users (BioISI, BILS, …)

Minimal increase
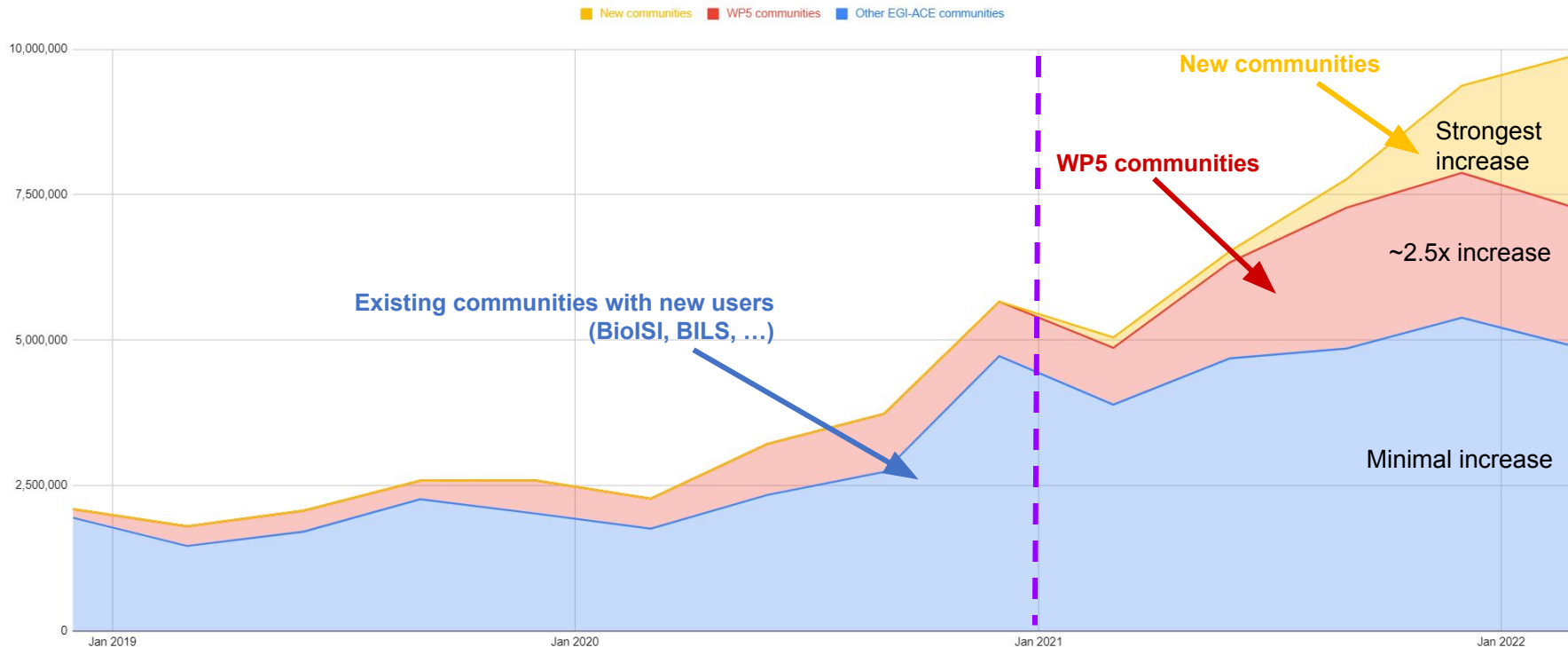
Start of EGI-ACE

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

FUTURE PLAN

Plan for the next period

- Complete the integration of the **SeaDataNet**, **OpenRiskNet** Data Spaces and **Disaster Mitigation and Agriculture** Thematic Service in the EOSC Compute Platform.
- Increase **outreach and dissemination activities** to promote the services uptake.
- Ensure long term support for user facing activities
  - EOSC procurement, Lot 4 as possible funding source

Final take-away from KER3

22 Thematic Services supported

18 Thematic Services in production (WP5+T2.3)

4 Thematic Services in development stage

VIP and **WeNMR** are operating production services

EMSO-ERIC moving in production

~49,000 users served

20M Cloud and HTC CPU/h allocated

+25 PB of Open Data shared by the Thematic Services

30 training events organized (including **2** webinars)

Thank you!

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

EGI Foundation

@EGI_eInfra

Backup slide

KER3 DESCRIPTION

BRIEF DESCRIPTION	• **Thematic Services = Data Spaces + Processing Platforms**
	• **Data Space** is *a public collection of FAIR, quality data and related resources consumed, produced and provided by identified participants, each respecting societal values and operating within an explicit framework of trust and governance*
	• **Processing Platforms** do not store and manage data for users ○ Users can upload (or define) input data, define/select data analysis/transformation steps, then download the results after the computation is finished.

FAIR maturity of the EGI-ACE Data Spaces

FAIR Data Maturity Model: specification and guidelines from the RDA FAIR Data Maturity Model WG

https://zenodo.org/record/3909563

The FAIR Data Maturity Model includes **47** indicators:

- **7** for the **F**indable, **12** for the **A**ccessible, **12** for the **I**nteroperable, and **16** for the **R**eusable aspect.
- **3** priorities for each Indicator: **Essential**, **Important** and **Useful**
- Each indicator can be evaluated as **TRUE**, or **FALSE** for the object that is in the focus of the assessment.

Data Space	Findability (F)	Accessible (A)	Interoperable (I)	Reusability (R)
useGalaxy.eu	100% Compliant	82% Compliant	86% Compliant	89% Compliant
GBIF	100% Compliant	100% Compliant	86% Compliant	100% Compliant
ENES	71% Compliant	91% Compliant	100% Compliant	100% Compliant
LOFAR Science Project	100% Compliant	100% Compliant	57% Compliant, but working on it	78% Compliant
EMSO ERIC	100% Compliant	91% Compliant	14% Compliant, but working on it	66% Compliant
SeaDataNet	100% Compliant	100% Compliant	100% Compliant	100% Compliant

Dependency [*] of the Thematic Services on the ECP

Color codes

Investigating	Technology/service is considered for adoption, but the integration and assessment work is yet to start
Planning	Integration is ongoing in the context of the EGI-ACE project
Integrated	Technology/service was integrated in the context of EGI-ACE
Adopted	Technology/service was already integrated before the EGI-ACE

Data Space	WP3 Infrastructure Services				WP4 Platform Services		WP6 Federated Services						WP7 Service Delivery and Planning
	EGI Cloud Compute	EGI High-Throughput Compute	EGI Online Storage	Infrastructure Manager	EGI Workload Manager	EGI Notebooks	CVMFS	EGI Check-in	EGI DataHub	EGI Data Transfer	PaaS Orchestrator	EC3	HPC
Disaster Mitigation & Agriculture	Planning		Planning			Investigating							
EMSO ERIC Data Space	Adopted		Adopted					Adopted					
ENES Data Space	Integrated		Integrated	Integrated				Integrated	Planning			Integrated	Integrated
GBIF Cloud Data Space	Planning		Planning			Investigating							
LOFAR Science Products		Integrated	Integrated						Investigating				
OpenCoastS	Adopted	Adopted	Adopted	Planning	Adopted			Integrated	Planning				
OpenRiskNet/NanoCommons Virtual Environment	Planning		Planning					Investigating					
OPERAS Metrics service and Certification service	Planning		Planning					Planning					
Prominence	Adopted		Adopted					Planning	Integrated				Integrated
SeaDataNet WebOcean Data Analysis	Integrated		Integrated				Integrated	Integrated		Integrated			
useGalaxy.eu				Investigating				Planning	Planning				
Virtual Imaging Platform	Adopted	Adopted	Adopted		Adopted	Investigating	Adopted	Integrated					
WeNMR services suite	Adopted	Adopted	Adopted		Adopted		Adopted	Adopted					Investigating

[*] Based on the current assessment