



 www.egi.eu
 @EGI_eInfra

The ascent of scientific computing: the EGI role and contribution towards EOSC

Tiziana Ferrari, EGI Foundation
HIFIS Conference, 2019




The work of the EGI Foundation
*is partly funded by the European Commission
under H2020 Framework Programme*



Outline

- 15 years of scientific computing in the EGI Federation – figures and trends
- EGI for open science
- EGI and the European Open Science Cloud



EGI is a federation of > 250 computing and data centres spread across Europe and the rest of the world.

47 Countries

> 71,500 users

3,000 Open Access

Publications in 2018

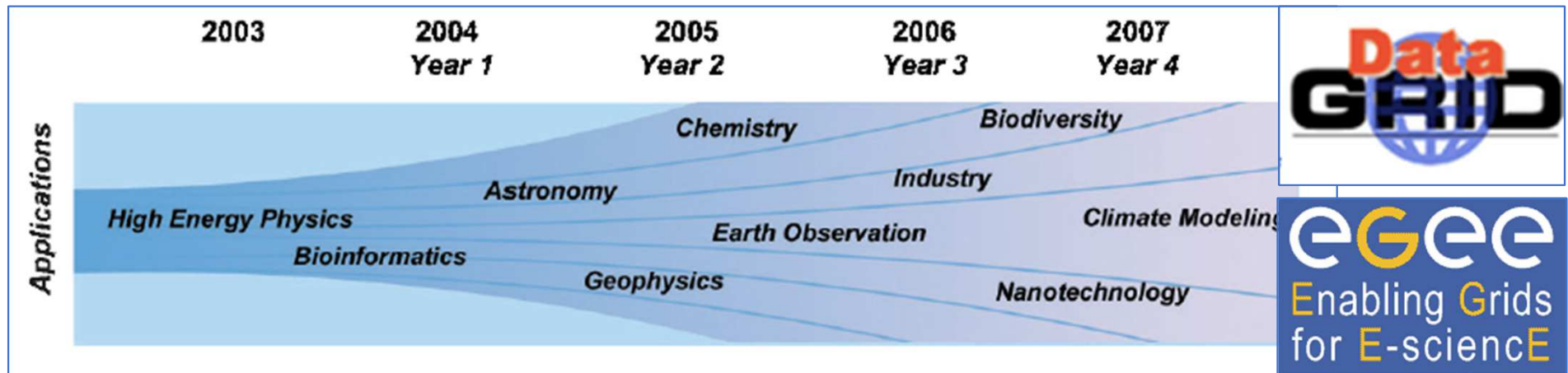
12 Integrated
e-Infrastructures

31 large-scale
research
collaborations

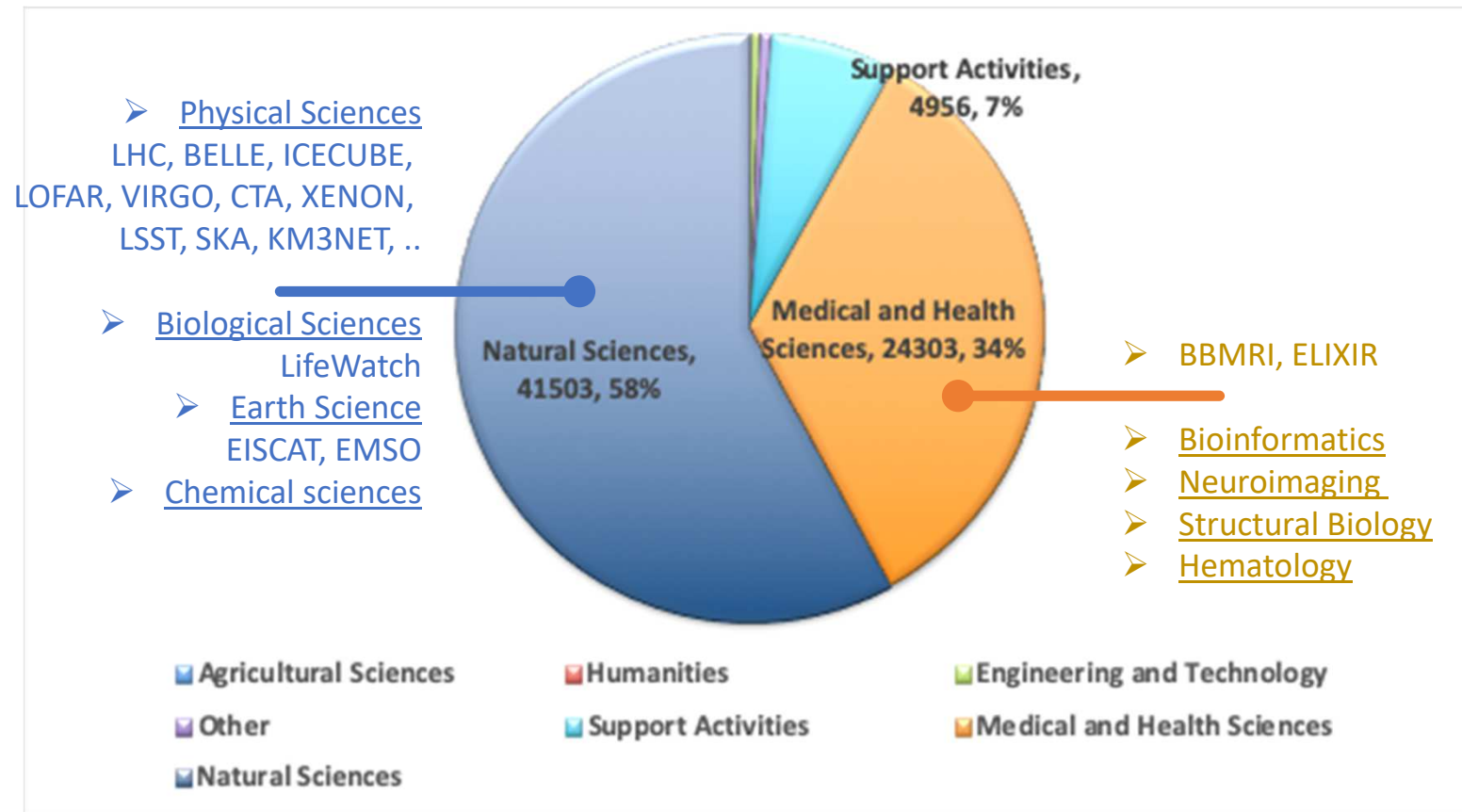
2001: From Design to Implementation

In 2000 the next generation High Energy Physics experiments at CERN established the design of the first large-scale scientific computing infrastructure in the Europe.

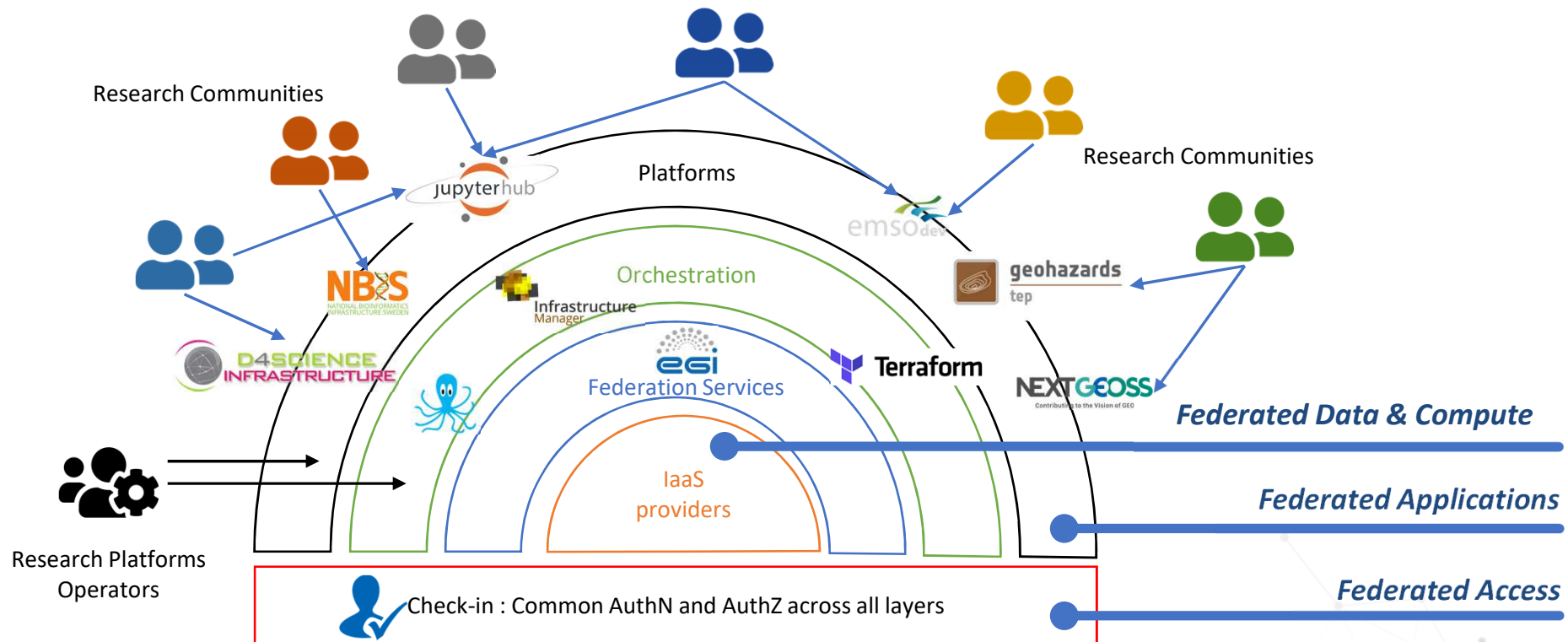
The EC-funded “DataGrid” project proposal was launched.



EGI User Statistics (Sep 2019)

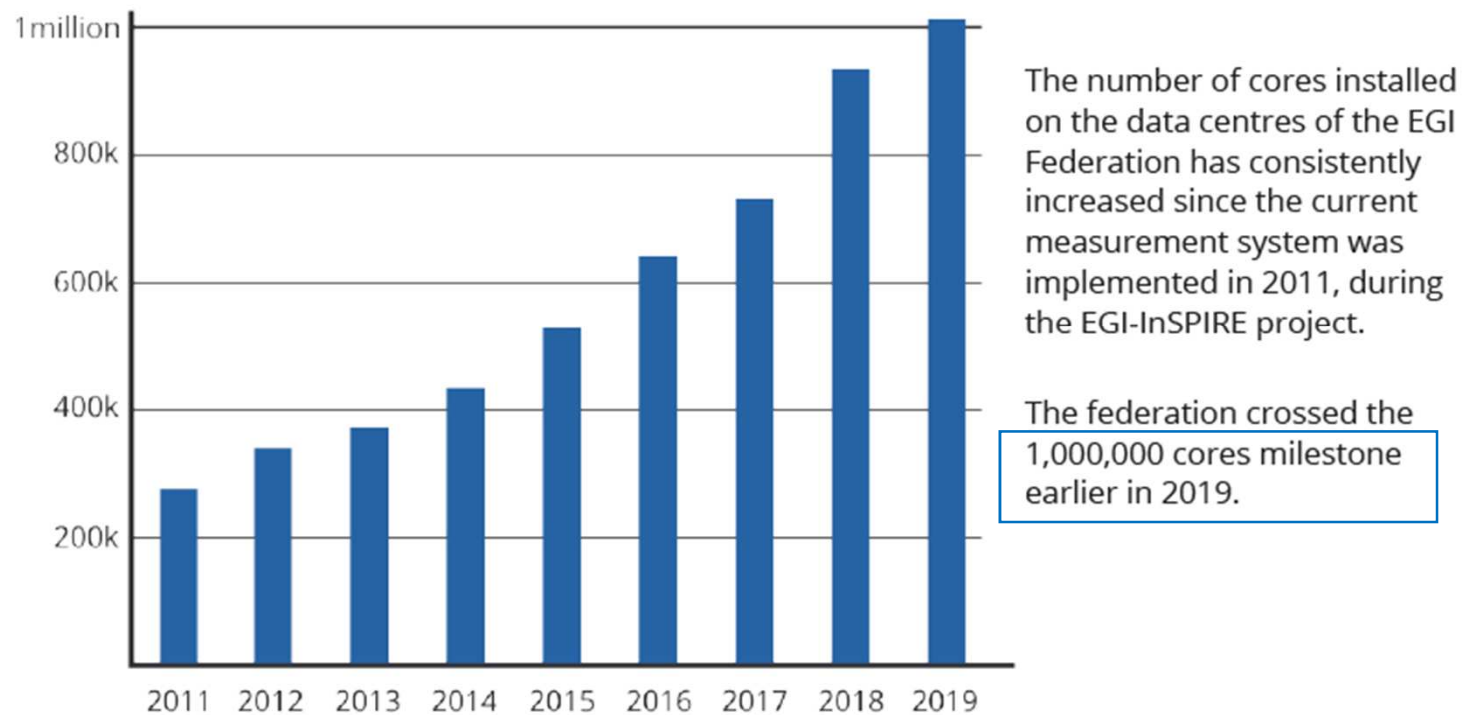


Data-centric Scientific Computing

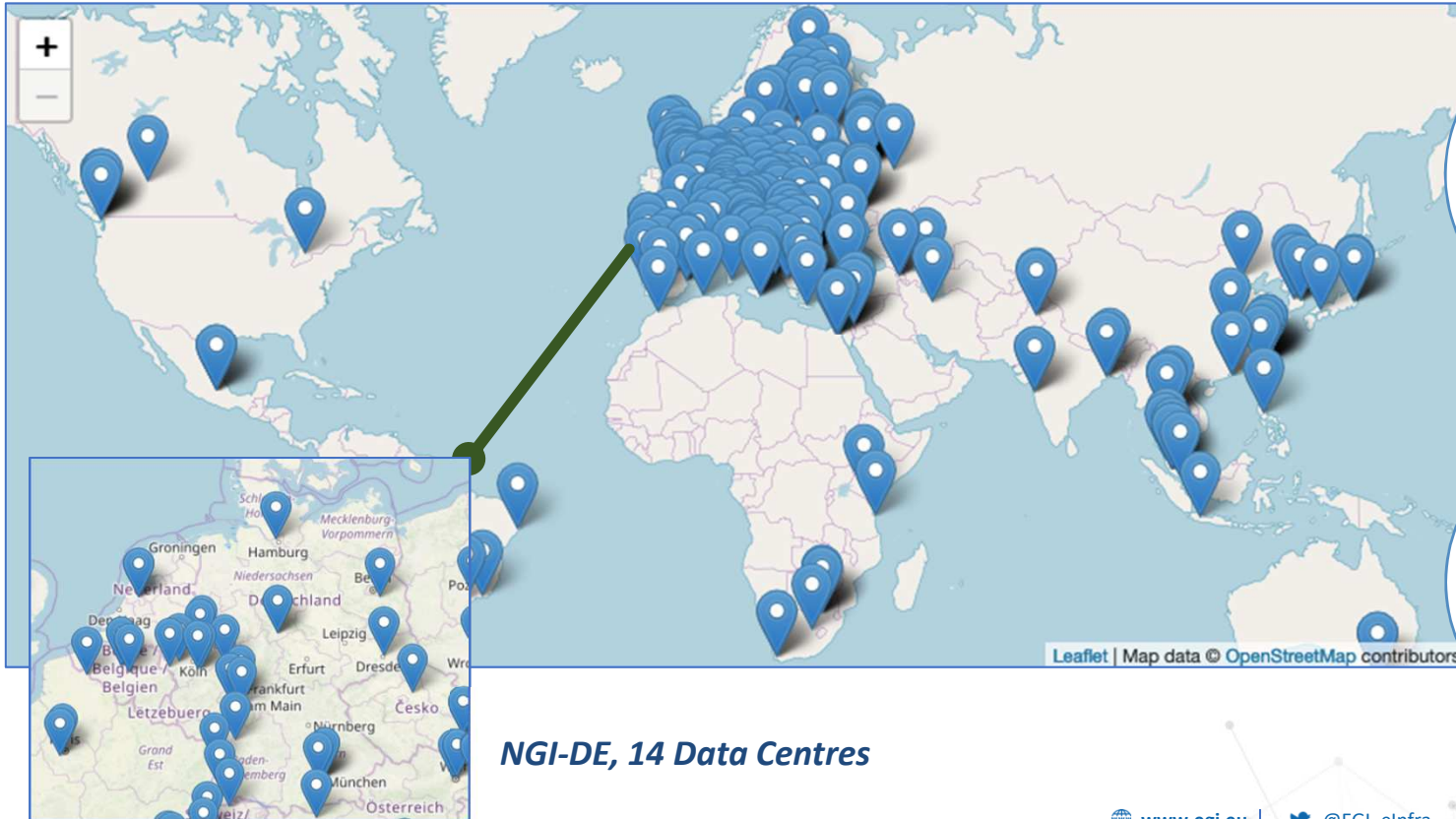


Installed Compute Capacity 2011-2019

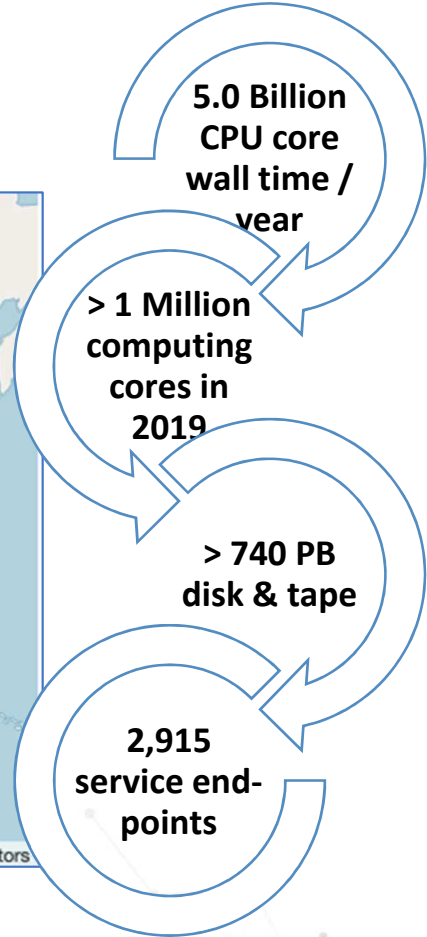
Installed compute capacity (number of cores), 2011-2019



The EGI Federation (Sep 2019)

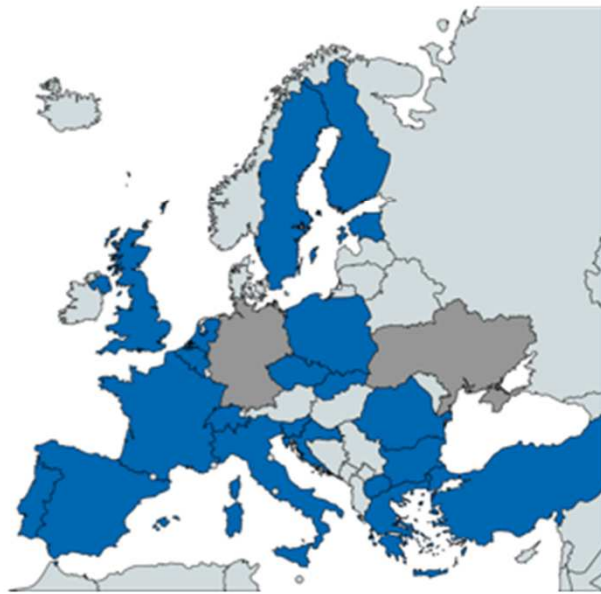


NGI-DE, 14 Data Centres



23 Countries and CERN

<https://www.egi.eu/about/egi-foundation/>

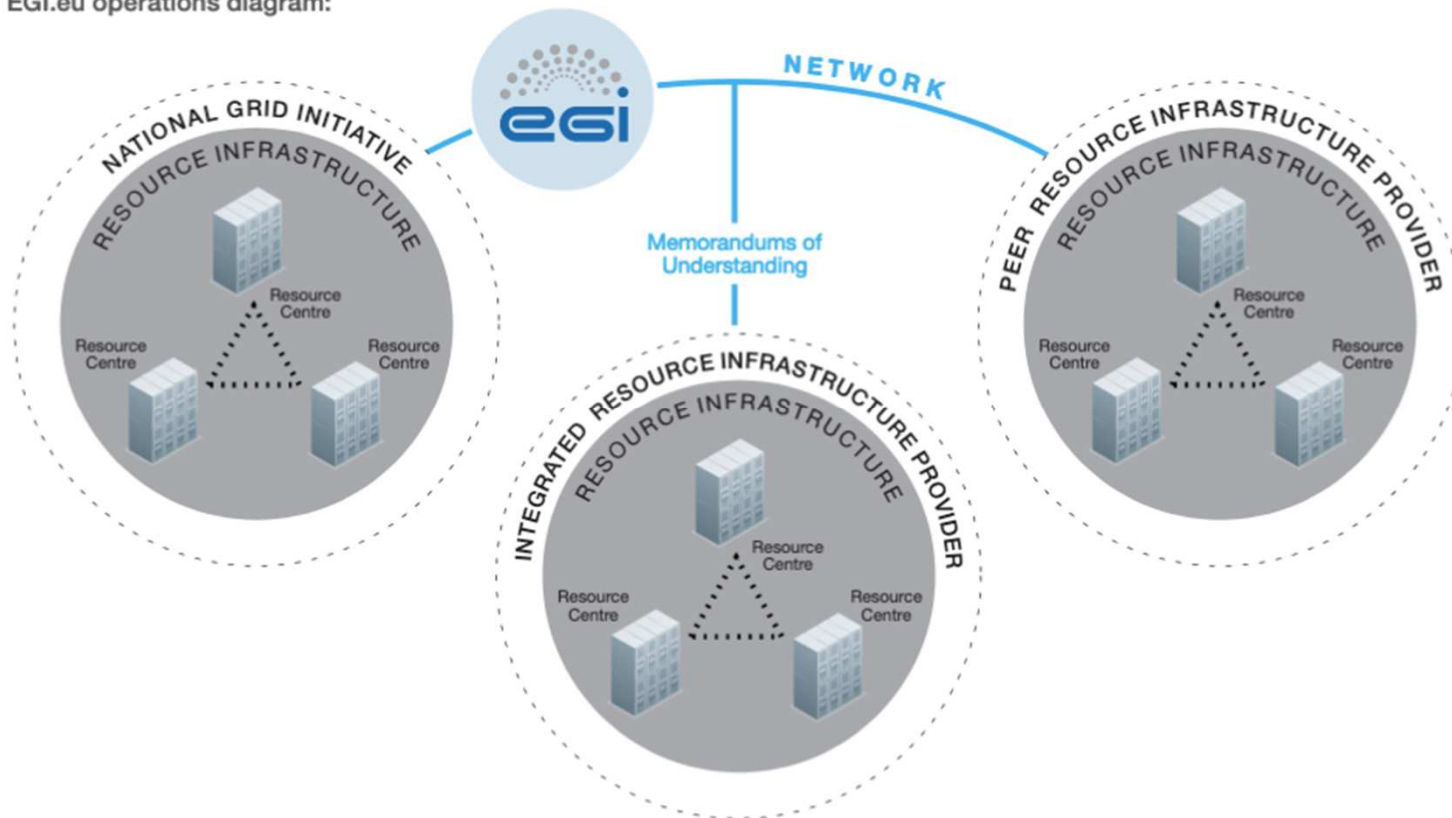


International Partners

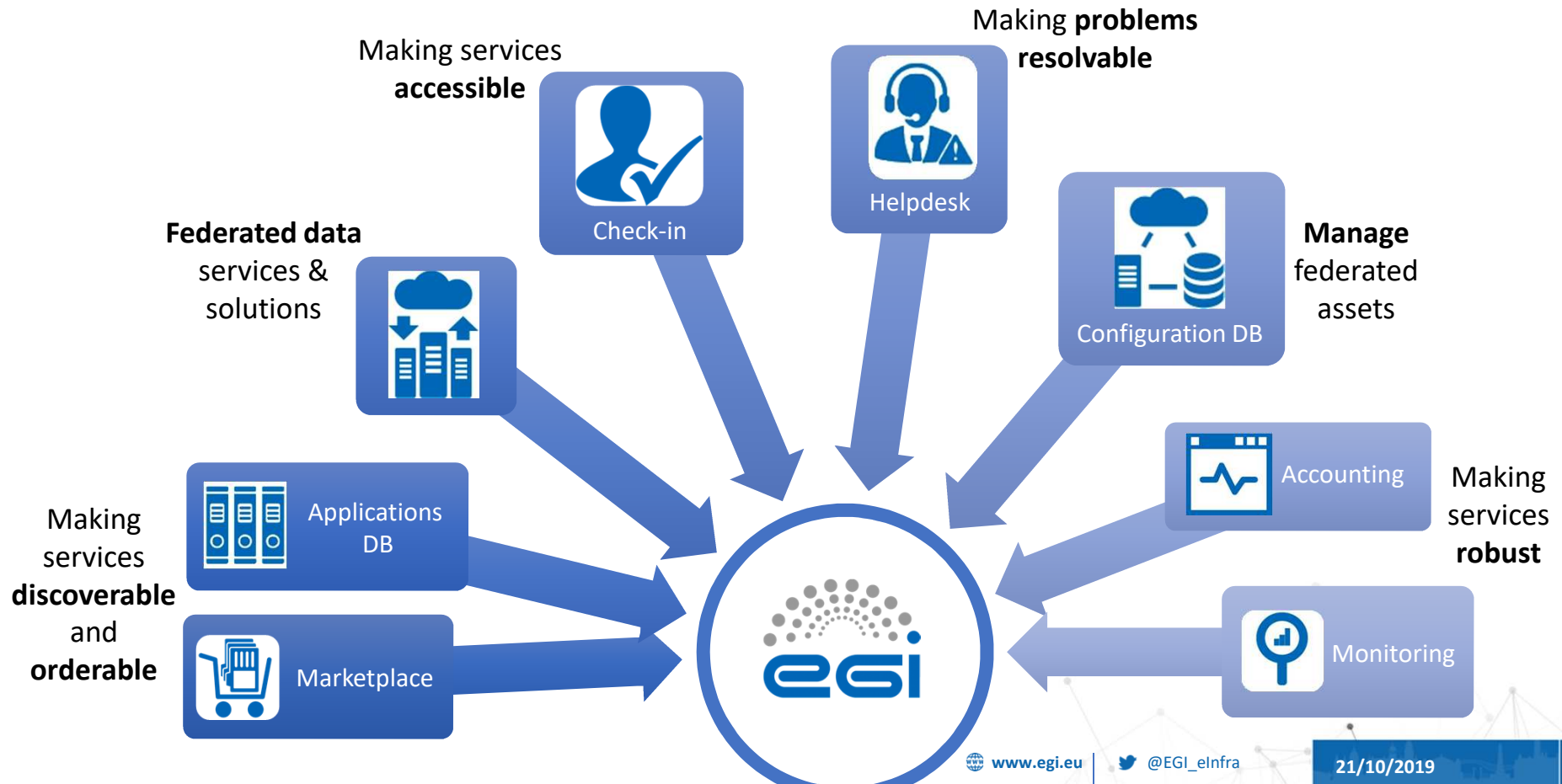


Federated Operations

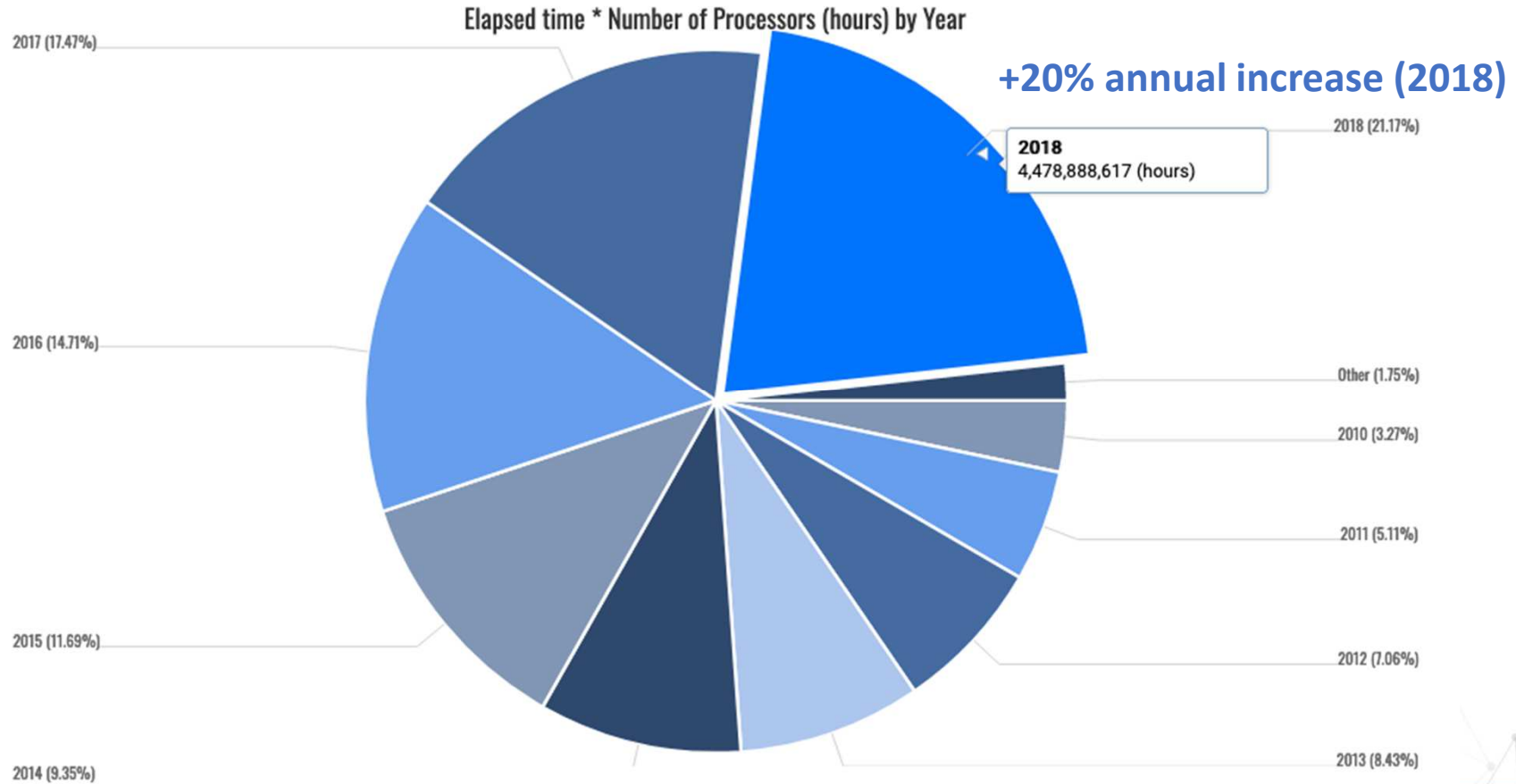
EGI.eu operations diagram:



EGI Federation Services



CPU wall time consumed 2009-2018



May 21, 2014: EGI Federated Cloud is launched

- Multi-cloud IaaS with Single Sign-On
- Federation features:
 - Common VM image catalogue
 - Discovery, accounting, SLO monitoring
 - Unified GUI dashboard



HPC Wire
Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them

- Home
- Technologies

EGI Launches Federated Cloud for European Researchers
May 21, 2014

HELSINKI, Finland, May 21 — The European Grid Infrastructure (EGI) today launched the Federated Cloud – a cloud service tailored for European researchers. The announcement was made at the annual EGI Community Forum, in Helsinki.

EGI Services powered by the Cloud Federation



Cloud Compute



Online Storage



Applications on Demand BETA



Cloud Container Compute BETA



Training Infrastructure



Notebooks BETA

Scientific Discipline Cloud — Elapsed time * Number of Processors (hours) by Scientific Discipline and Year

Scientific Discipline	2015	2016	2017	2018	Total ^	Percent
Other	0	0	0	2,771	2,771	0%
Medical and Health Sciences	1,305	1,295,248	1,190,800	2,206,120	4,693,473	5.34%
Engineering and Technology	4,059	863,467	2,256,929	2,277,126	5,401,581	6.15%
Humanities	2,632,434	3,201,538	1,313,777	1,248,065	8,395,814	9.56%
Support Activities	3,409,107	8,470,420	11,173,822	9,212,083	32,265,431	36.73%
Natural Sciences	9,669,539	7,403,863	7,977,551	12,025,093	37,076,047	42.21%
Total	15,716,444	21,234,536	23,912,880	26,971,257	87,835,117	
Percent	17.89%	24.18%	27.22%	30.71%		

1 - 6 of 6 results Number of rows per page 30

Humanities: CLARIN (2018), DARIAH (+28%, 2017),
Life Sciences: ELIXIR (+104%, 2017)
Environment: EMSO (2018), LifeWatch (+45%, 2015)
BIOISI (+45%, 2016), **BIOMED** (+575%, 2016), **CHIPSTER** (+256%, 2016),
GEOHAZARDS (+19%, 2016), **OpenCOAST** (2018), **EXTRAS** (2018),

685
 cloud virtual appliances

requested through AppDB in 2018

Scientific Computing for Open Science

Open data, open software, open applications

3000 open access publications / year

The Nobel Prize in Physics 2013



© Nobel Media AB. Photo: A. Mahmoud
François Englert
 Prize share: 1/2

© Nobel Media AB. Photo: A. Mahmoud
Peter W. Higgs
 Prize share: 1/2

LHC Collaboration

The Nobel Prize in Physics 2017



© Nobel Media AB. Photo: A. Mahmoud
Rainer Weiss
 Prize share: 1/2

© Nobel Media AB. Photo: A. Mahmoud
Barry C. Barish
 Prize share: 1/4

© Nobel Media AB. Photo: A. Mahmoud
Kip S. Thorne
 Prize share: 1/4

LIGO-VIRGO Collaboration

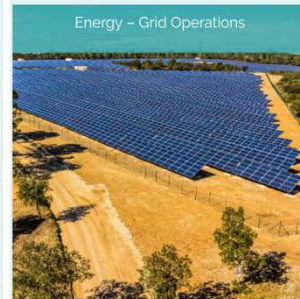




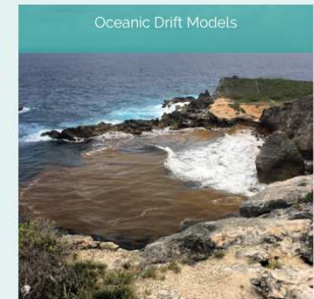
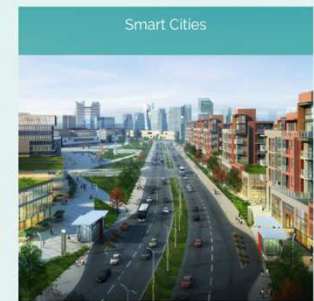
Open Data for the UN Sustainable Development Goals

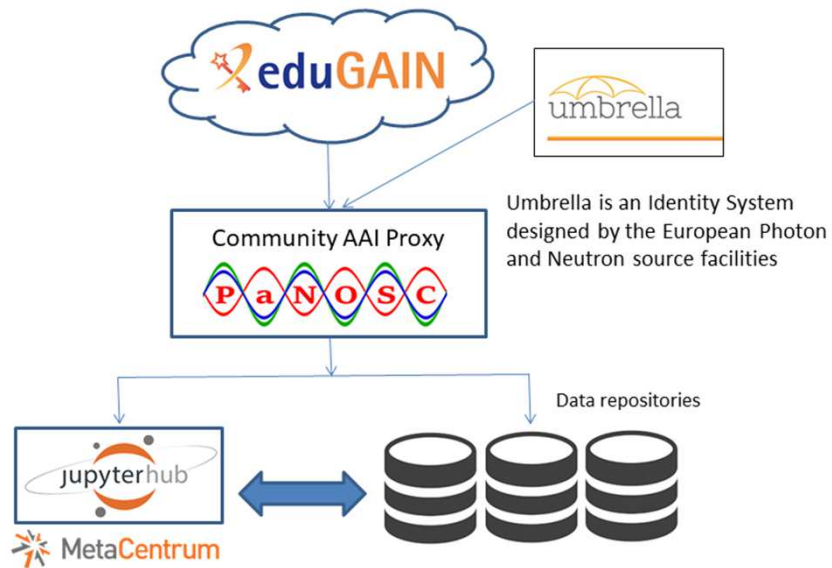
NEXTGEOSS

Innovative Pilots



Business Pilots





- ESRF (France)
- ILL (France)
- ESS (Sweden)
- ELI-DC (Belgium)
- XFEL (Germany)
- CERIC-ERIC (Italy)
- EGI Foundation (Netherlands)

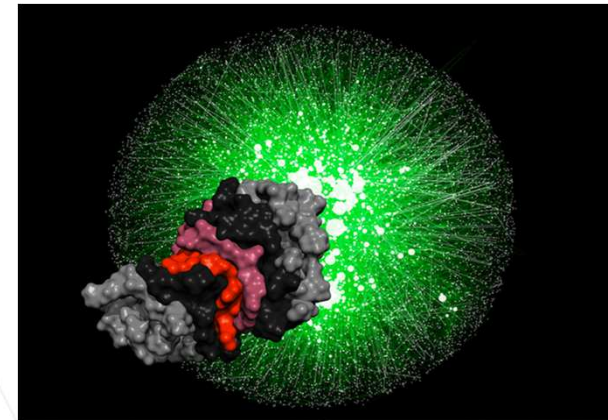
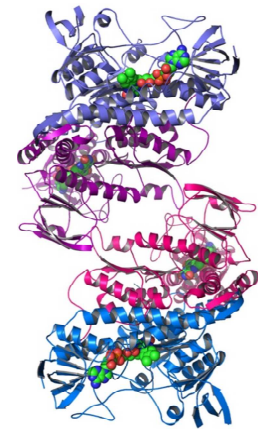


Notebooks
Federated Cloud

Sharing of scientific applications

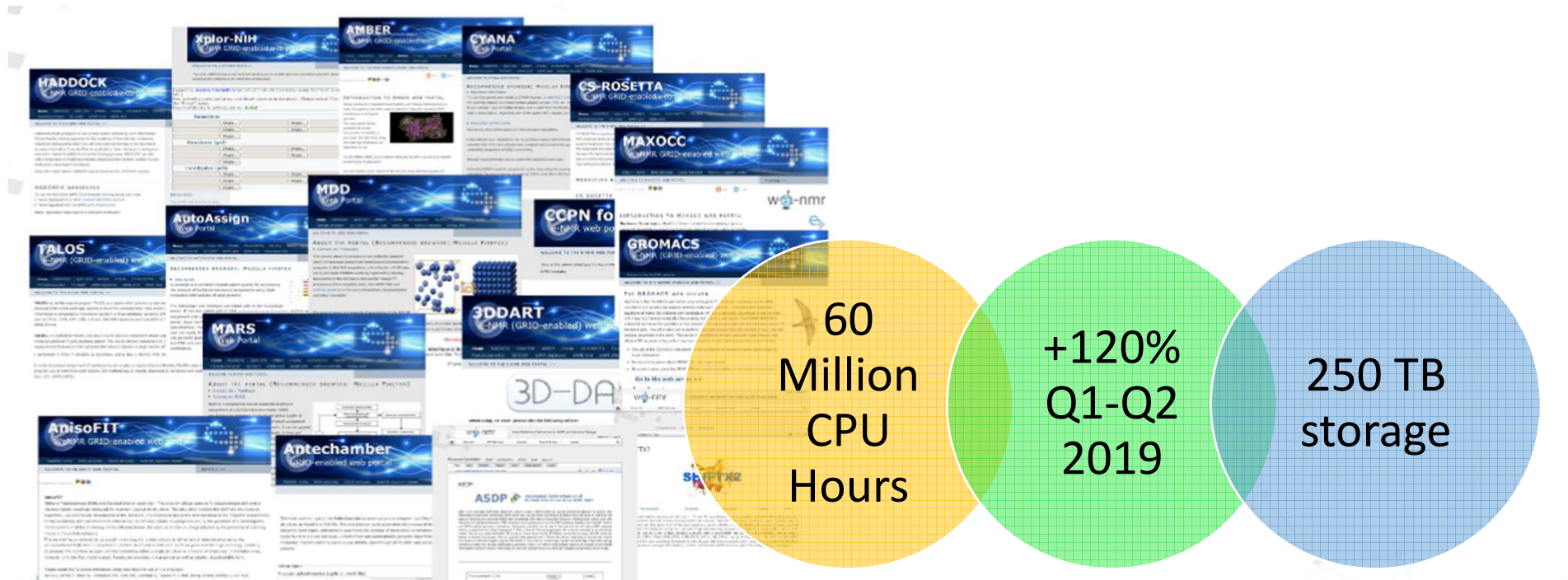
*Understanding the pathways to disease:
the WeNMR community of practice*

- Proteins and biomolecules interact in a complex network
- Glitches in this network can cause diseases like cancer
- Researchers use 3D models to study how protein and biomolecules interact
- This structural information is key to understand the origins of disease and to develop new drugs



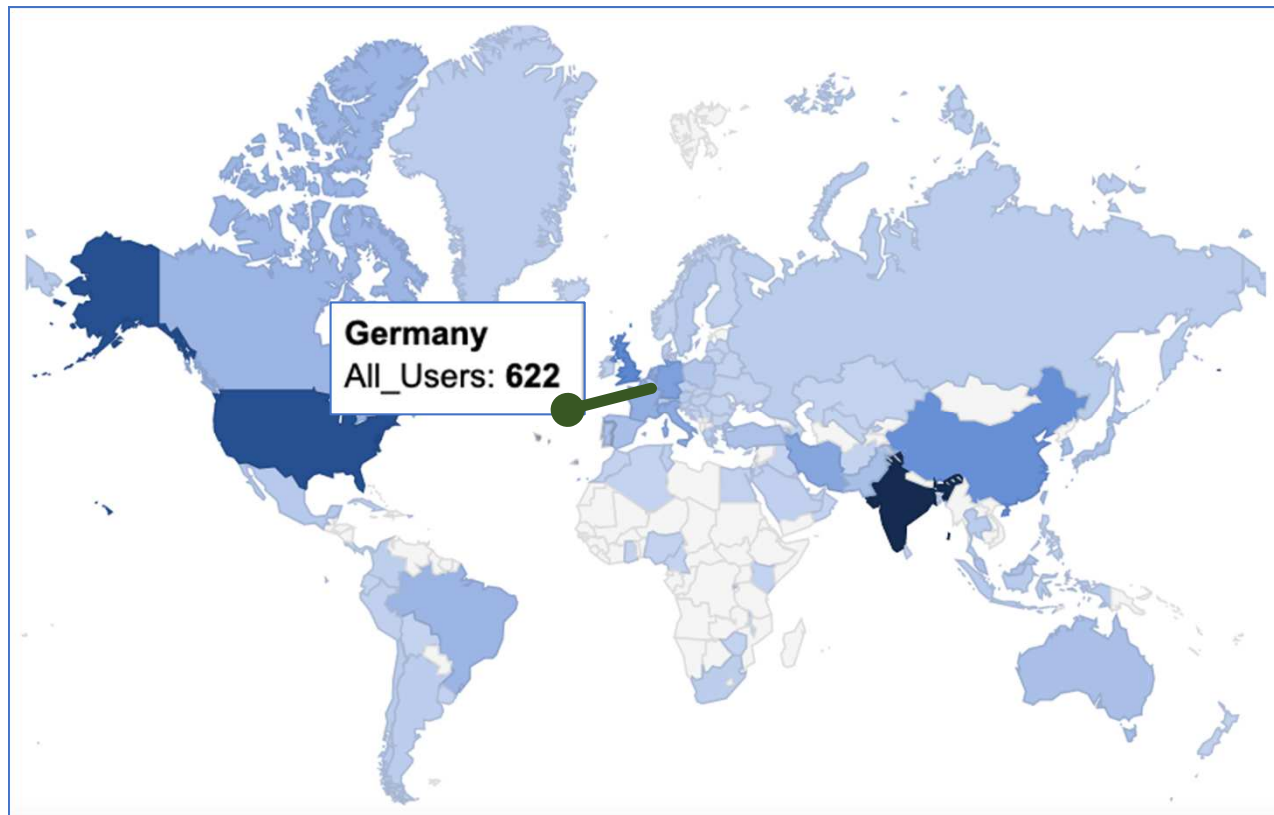


The WeNMR service portfolio



Geo Distribution of the WeNMR Users

https://wenmr.science.uu.nl/user_map



EGI and the European Open Science Cloud

Data, applications and computing infrastructure as integrated solution

Towards a data-centric compute infrastructure

- Growing demand of Cloud access to HTC, GPU and HPC resources
 - Machine learning adopted by many applications running in the EGI Federation
 - Hybrid provisioning is key
- End-users engaged at the SaaS level
 - Easy to use solutions bring together
 - Federated AAI
 - Federated data discovery and management
 - Federated data analytics capabilities
- Data-centric approach to Cloud
 - Zero download access to large research data holdings e.g Copernicus

Climate (big) data challenges and scientific workflow

Dealing with Data

PERSPECTIVE

Climate Data Challenges in the 21st Century

Jonathan T. Overpeck,^{1*} Gerald A. Meehl,² Sandrine Bony,³ David R. Easterling⁴

Climate data are dramatically increasing in volume and complexity, just as the users of these data in the scientific community and the public are rapidly increasing in number. A new paradigm of more open, user-friendly data access is needed to ensure that society can reduce vulnerability to climate variability and change, while at the same time exploiting opportunities that will occur.

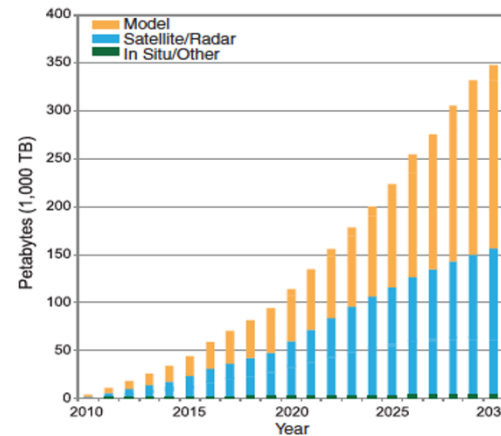
Climate variability and change, both natural and anthropogenic, exert considerable influences on human and natural systems. These influences drive the scientific quest for an understanding of how climate behaved in the past and will behave in the future. This understanding is critical for supporting the needs of an ever-broadening spectrum of society's decision-makers as they strive to deal with the influences of Earth's climate at global to local scales. Our understanding of how the climate system functions is built on a foundation of climate data, both observed and simulated (Fig. 1). Although research scientists have been the main users of these data, an increasing number of resource managers (working in fields such as water, public lands, health, and marine resources) need and are seeking access to climate data to inform their decisions, just as a growing range of policy-makers rely on climate data to develop climate change strategies. Quite literally, climate data provide the backbone for billion-dollar decisions. With this gravity comes the responsibility to curate climate data and share it more freely, usefully, and readily than ever before.

The Exploding Volume of Climate Data
Documenting the past behavior of the climate system, as well as detecting changes and their causes, requires the use of data from instrumental, paleoclimatic, satellite, and model-based sources. The earliest instrumental (thermometer and barometer) records stretch back to the mid-to-late 1800s.

evolution of climate. Inevitably, there are uncertainties in the observational records that need to be translated into the degree of confidence associated with our understanding of how the climate system behaves.

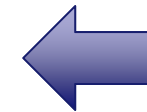
In addition to the already large body of digital instrumental data available in diverse holdings around the globe, a substantial number of critical observations, such as many early temperature observations, are not yet widely available as digital records. It is important to create and maintain central repositories of these data in a manner that firmly defines the origin and nature of the data and also ensures that they are freely available (1, 2). In addition, an increasing array of paleoclimatic proxy records from human and natural archives, such as historical documents, trees, sediments, corals, and ice cores, are being generated. These records are particularly helpful in understanding climate variability before the period of instrumental data.

Climate Data Challenges in the 21st Century
Jonathan T. Overpeck, Gerald A. Meehl, Sandrine Bony and David R. Easterling (February 10, 2011)
Science 331 (6018), 700–702. [doi: 10.1126/science.1197869]



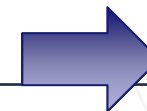
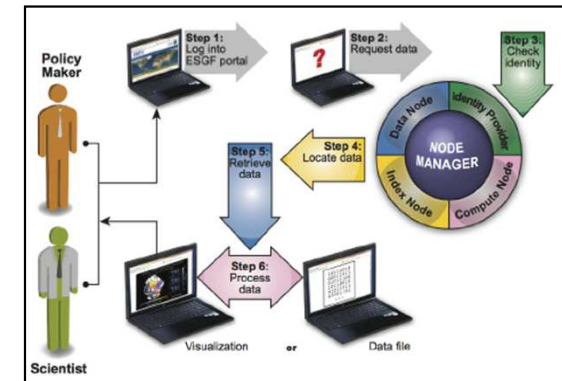
Big data from simulations

The volume of worldwide climate data is expanding rapidly, creating challenges for both physical archiving and sharing, as well as for ease of access and finding what's needed, particularly if you are not a climate scientist.

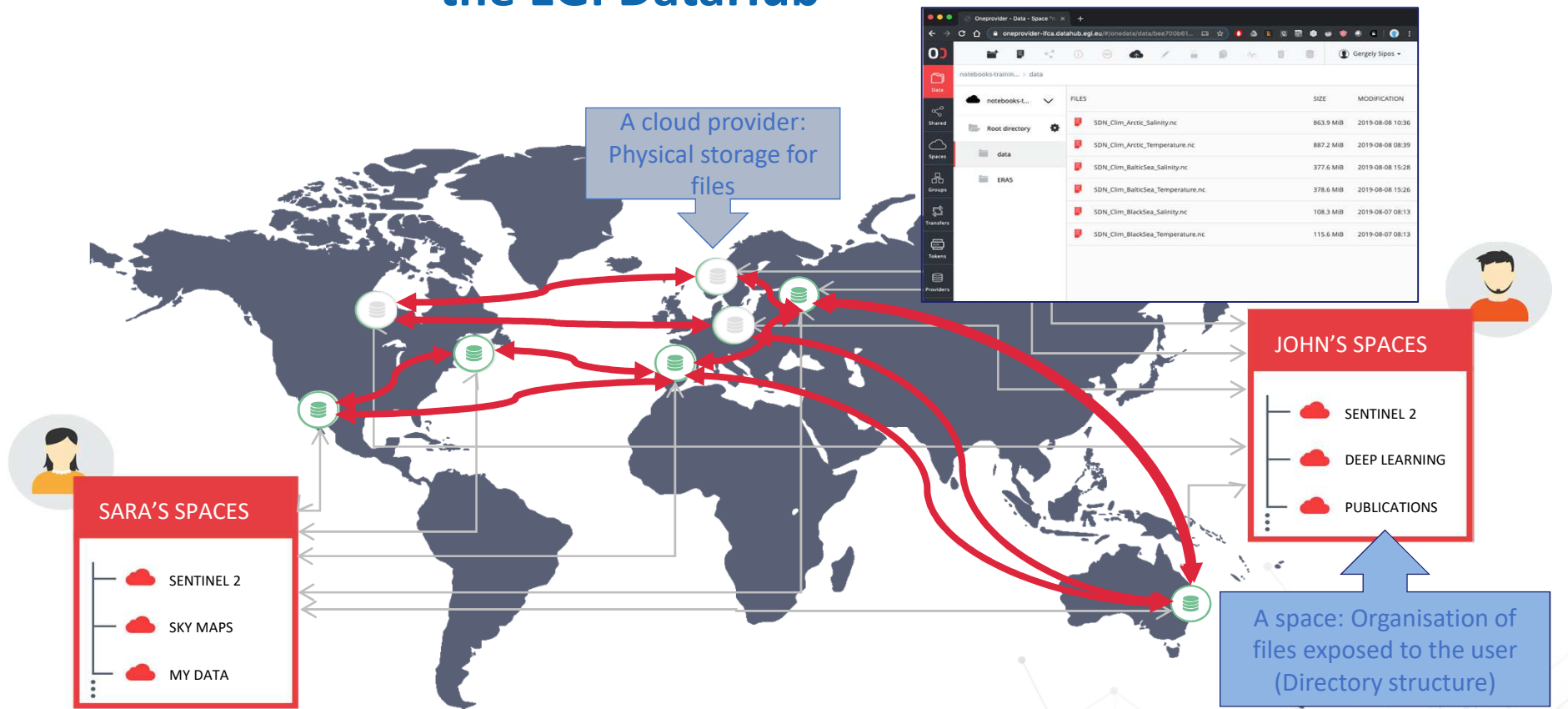


Scientific workflow

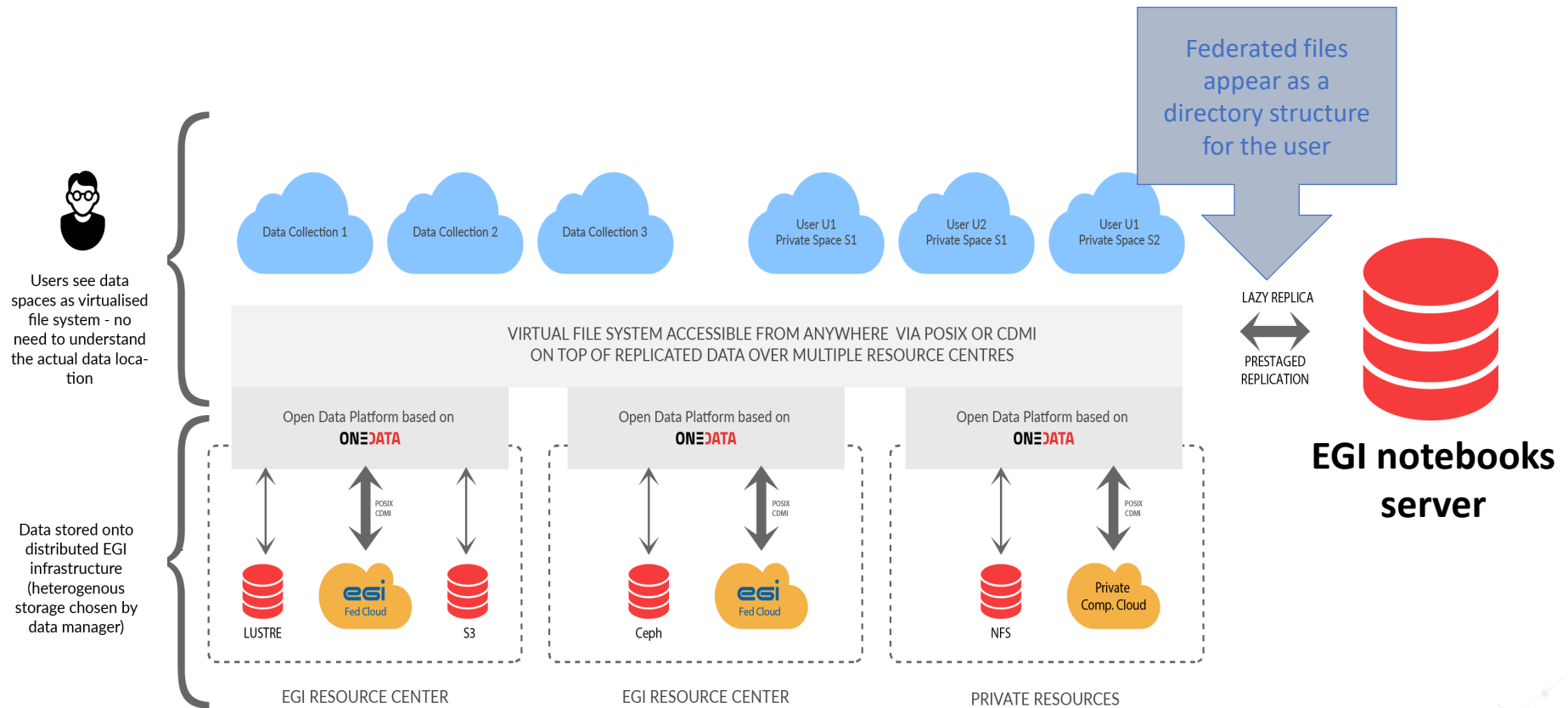
- Still mostly based on the *workflow: search, locate, download, analyze, visualize*
- *client-side and sequential nature of the current approach*
- *Strong need for server-side approaches with advanced parallel capabilities and no more data download*



Federating data in a multi-cloud with the EGI DataHub



Example: EGI Notebooks

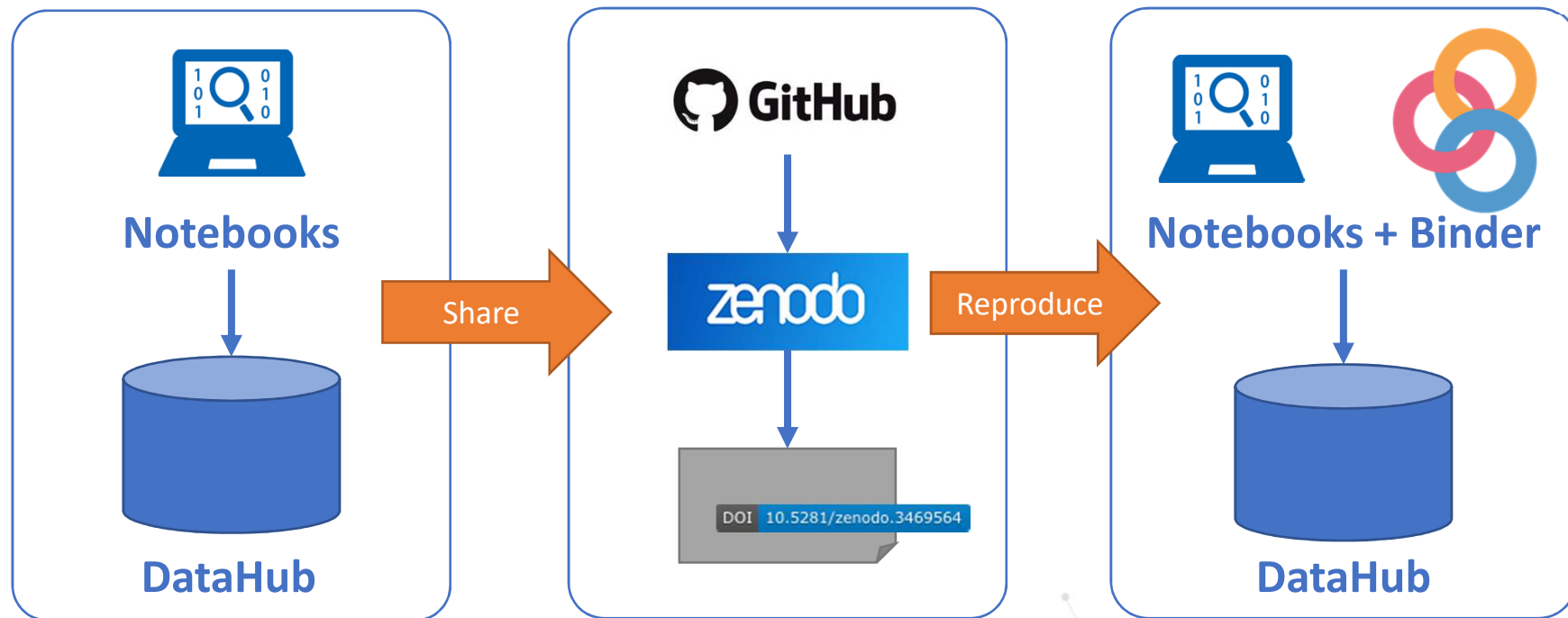


Users see data spaces as virtualised file system - no need to understand the actual data location

Data stored onto distributed EGI infrastructure (heterogenous storage chosen by data manager)

Open science with EGI and OpenAIRE Services

Reproducible and discoverable analysis



EOSC-hub **The European Open Science Cloud initiative**

Vision

“Give the European Union a global lead in research data management and ensure that European scientists reap the full benefits of data-driven science, by offering 1.7 million European researchers and 70 million professionals in science and technology a virtual environment with free at the point of use, open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines”



<https://eosc-portal.eu/>



EOOSC-hub



The EOOSC-hub project mobilises providers of pan-European relevance offering services, software and data for advanced data-driven research and innovation.

These resources are offered via the Hub – the integration and management system of the European Open Science Cloud, acting as a European-level entry point for all stakeholders.

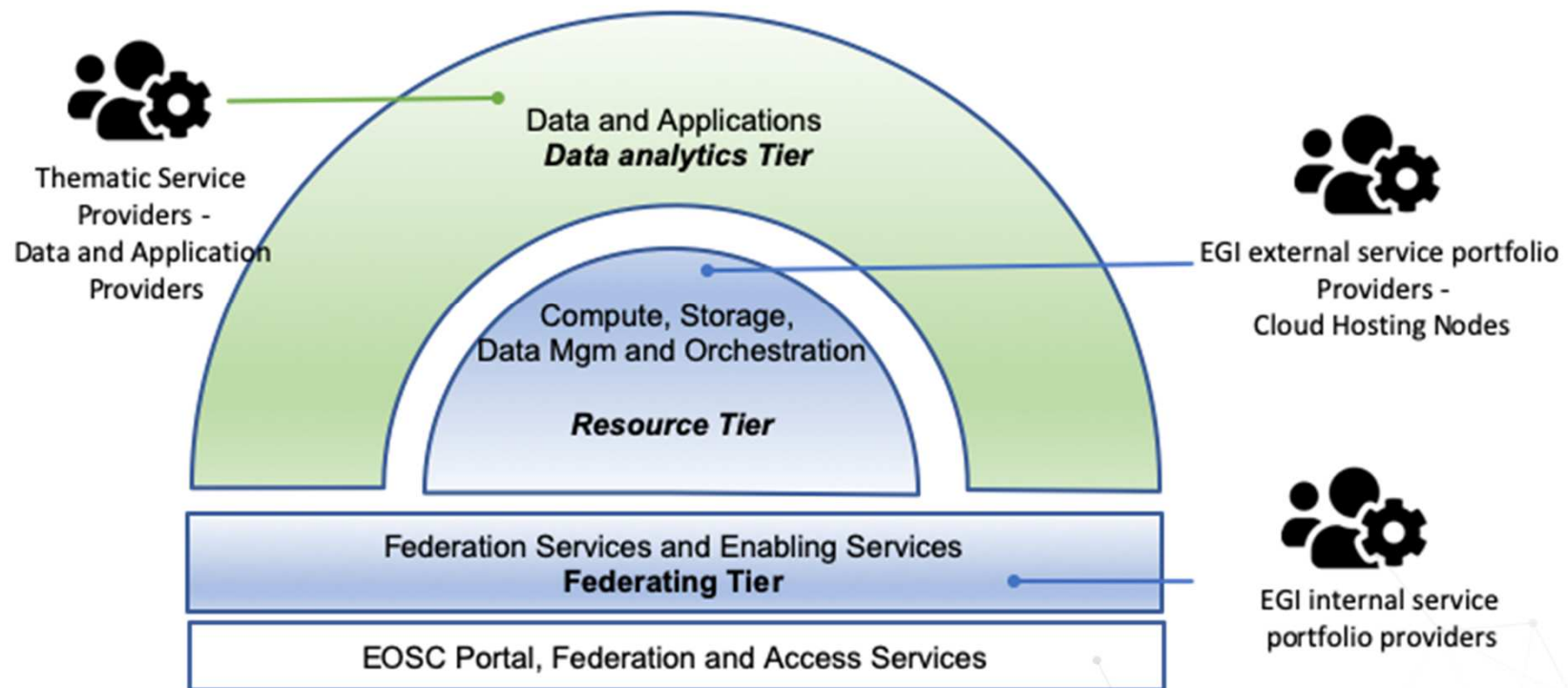
Vision

Computing, storage, data, software and a rich portfolio of research community-specific services for data analytics will be federated and made available for on-demand access in the European Open Science Cloud through the participation of EGI members and partners.

EGI Mission

*To implement the **data processing pillar** of the European Open Science Cloud, by **extending the on-demand capacity and capabilities** of the EGI external service catalogue,
- leveraging the EGI federation services and the funding, development and operational efforts of EGI participants and partners.*

Future EGI tiered architecture



- Scientific computing infrastructures are a necessary instrument of scientific excellence and open science in Europe
- EGI: from HTC infrastructure, to hybrid HTC & HPC infrastructure supporting cloud IaaS, PaaS, SaaS access
 - Federated identity provisioning and access
 - Federated data management towards exabyte scale data processing
- European Open Science Cloud
 - Supporting the entire research data lifecycle from production to zero download data exploitation

EGI: Advanced Computing for Research

 www.egi.eu
 @EGI_eInfra

Thank you
for your attention.

Questions?



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



The work of the EGI Foundation
is partly funded by the European Commission
under H2020 Framework Programme