

EGI for Earth Science

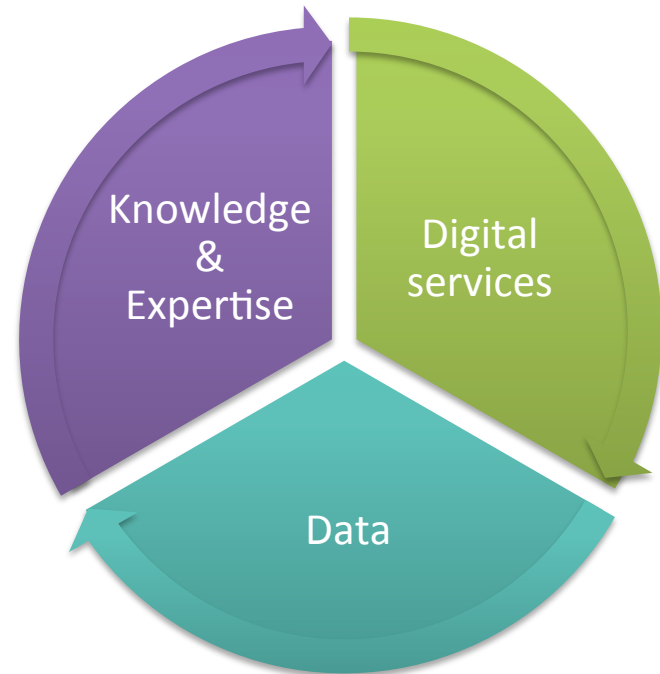
Tiziana Ferrari

Technical Director, EGI.eu

e-Infrastructures for Earth Sciences - workshop
Amsterdam, 22-23 January 2015



*Researchers from **all disciplines** have easy and open access to the innovative **digital services**, **data**, **knowledge and expertise** they need for performing **collaborative excellent research***



<http://go.egi.eu/osc>



Individual
Researchers
& Teams



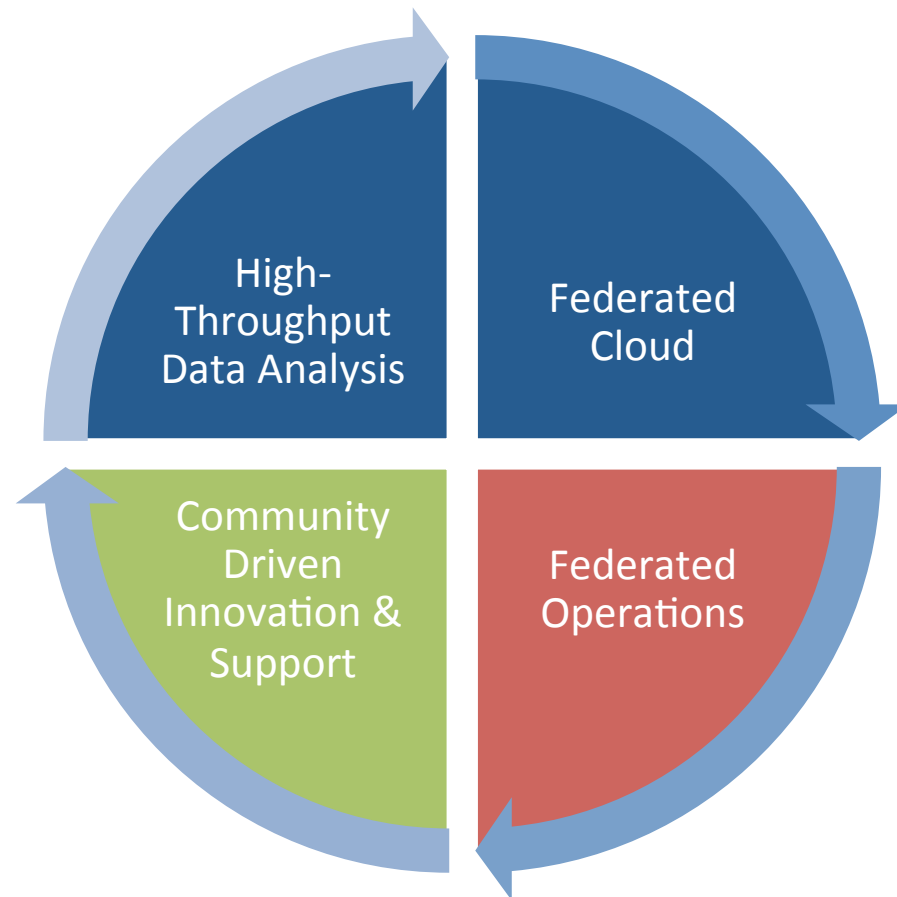
International
Research
Communities
& Institutions



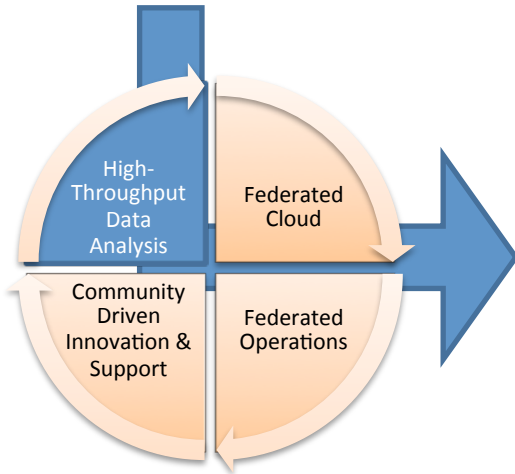
Individual
resource
Centres &
national e-
Infrastructures

EGI Solutions

<http://www.egi.eu/solutions/>



Do you need to manage and analyse large datasets, or to execute thousands of computational tasks?



European federation of publicly-funded clusters for High-Throughput Computing (HTC)

- **Grid Compute**: run large-scale computational jobs
- **Grid Storage**: store/access/retrieve files
- **File Transfer**: file data movement service
- **File Metadata Catalogue**: metadata catalogue service

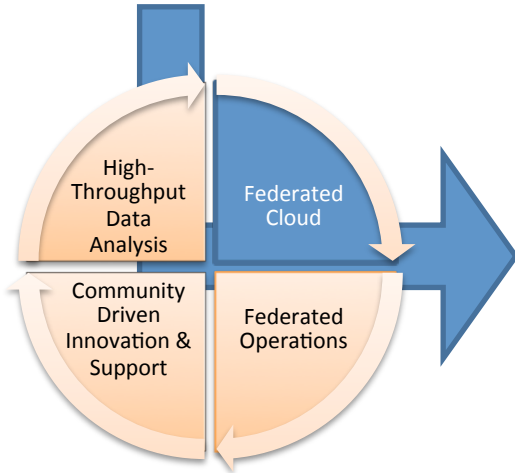
Based on **open standards** and **open source software**

Integrate heterogeneous infrastructure/technologies

Added Value

- **Uniform access** to distributed computing capabilities to run large-scale computational jobs processing big data and preventing single vendor lock-in
- Possibility to **federate your own resources**
- Facilitate **collaboration** across communities and borders by **sharing compute and data**

Do you need an infrastructure to deploy on-demand IT services for managing and processing your research data?



European federation of publicly-funded community clouds

- **Cloud Compute**: deploy/manage virtual machines (VM)
- **Cloud Storage**: store/access/retrieve digital objects
- **Virtual Appliance Marketplace**: marketplace for application developers and users

Based on **open standards** and **open source software**

Integrate heterogeneous cloud technologies

- OpenStack, OpenNebula, CloudStack,...

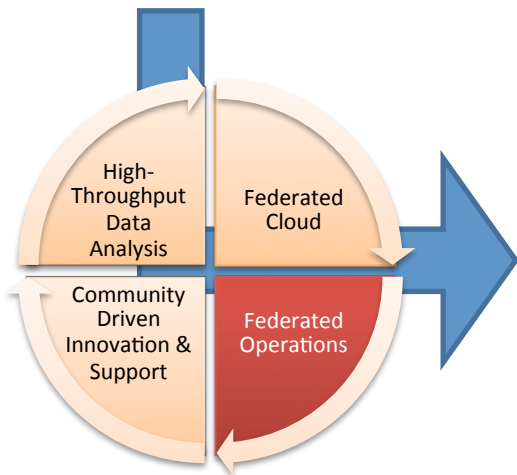
Integrate with commercial providers*

- Within EGI or through the Helix Nebula Marketplace

Added Value

- Promotion of open cloud standards for uniform access, **no lock-in** and **on-demand scale out** capabilities
- Easy deployment of **own/customised services clouds**
- **Efficiency** by **co-locating big data + cloud computing**

Do you need operational services and tools to run a distributed IT infrastructure for research?



Technologies, processes and expertise:

- To manage operations of heterogeneous distributed infrastructures
- To integrate resources from multiple independent providers with lightweight central coordination

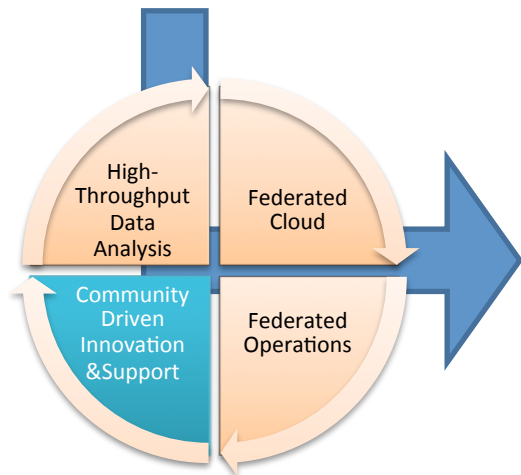
Relies on the best practices for IT service management (FitSM)

- **Operations, Technology, and Security Coordination**
- **Helpdesk support**
- **Governance & Project specialised consultancy services**
- **Technical consultancy and support**
- **Distributed service registry, accounting, monitoring,**

Added Value

- A cost-efficient framework to manage operations within a federated environment, while retaining responsibility of local infrastructure.
- Allows efficient implementation of Best Management Practices for IT services

Do you want to discover and implement the best approach to your data-intensive compute needs?



Expert **assistance** for easy and seamless integration and access to the computing and data services

- **European Distributed Competence Centres**
- Leveraging expertise of **user communities** and **technology providers**
- Partnership with **National Grid Initiatives** and **external competence centres**

Support at both **national** and **EU level**

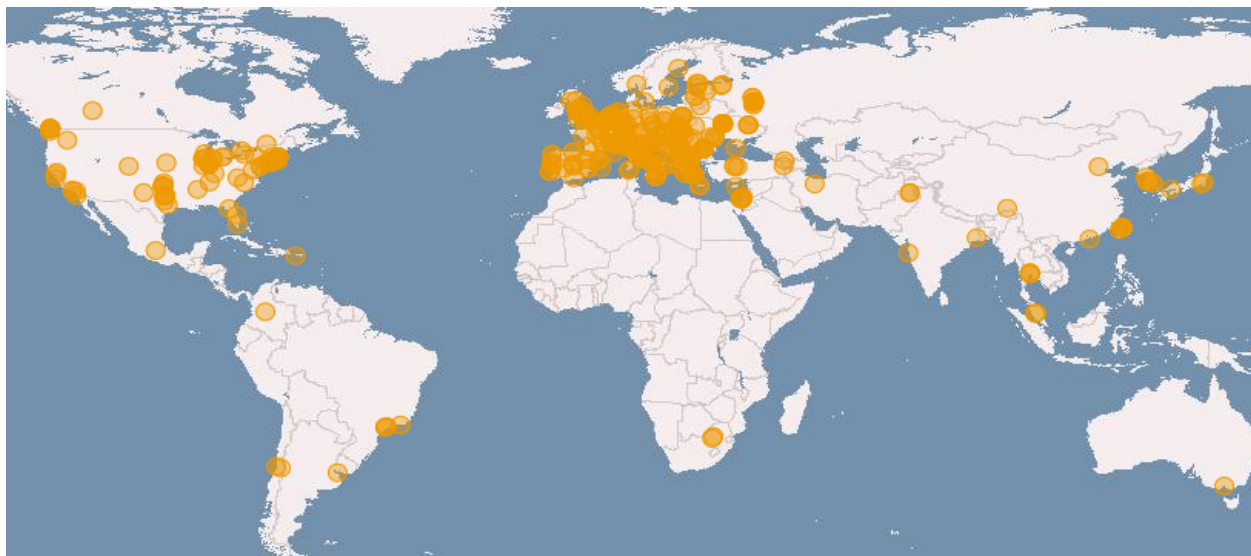
Added Value

- **Avoid duplication** of solutions
- **Leverage** expertise from the community
- Facilitate **collaboration** across communities and borders
- **User-driven innovation**

Principles of the federation

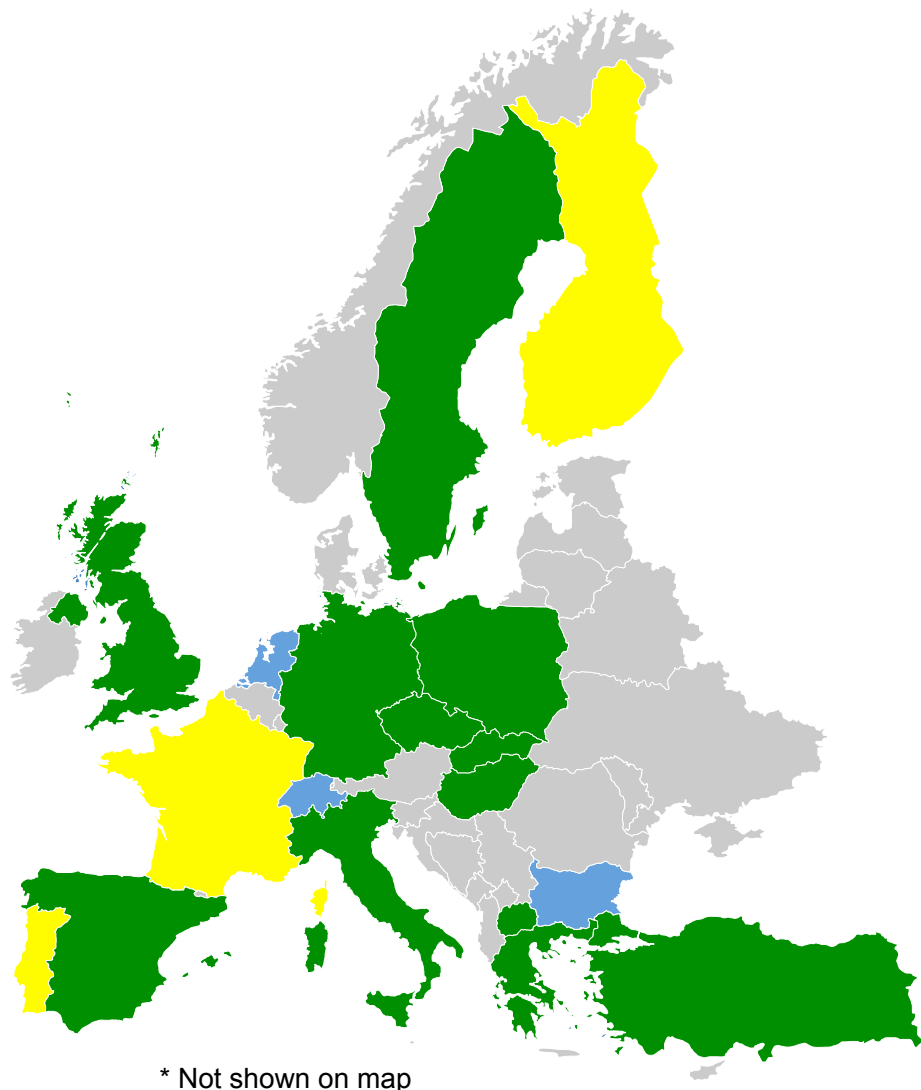


- Uniform, secure access to multi-disciplinary heterogeneous data management and compute services and capacity provisioning
- Federation services of e-Infrastructure services from
 - Publicly-funded infrastructures
 - Commercial providers
 - Partnership with Helix Nebula
 - Free at point of delivery or pay per use
- Federation of community-specific services

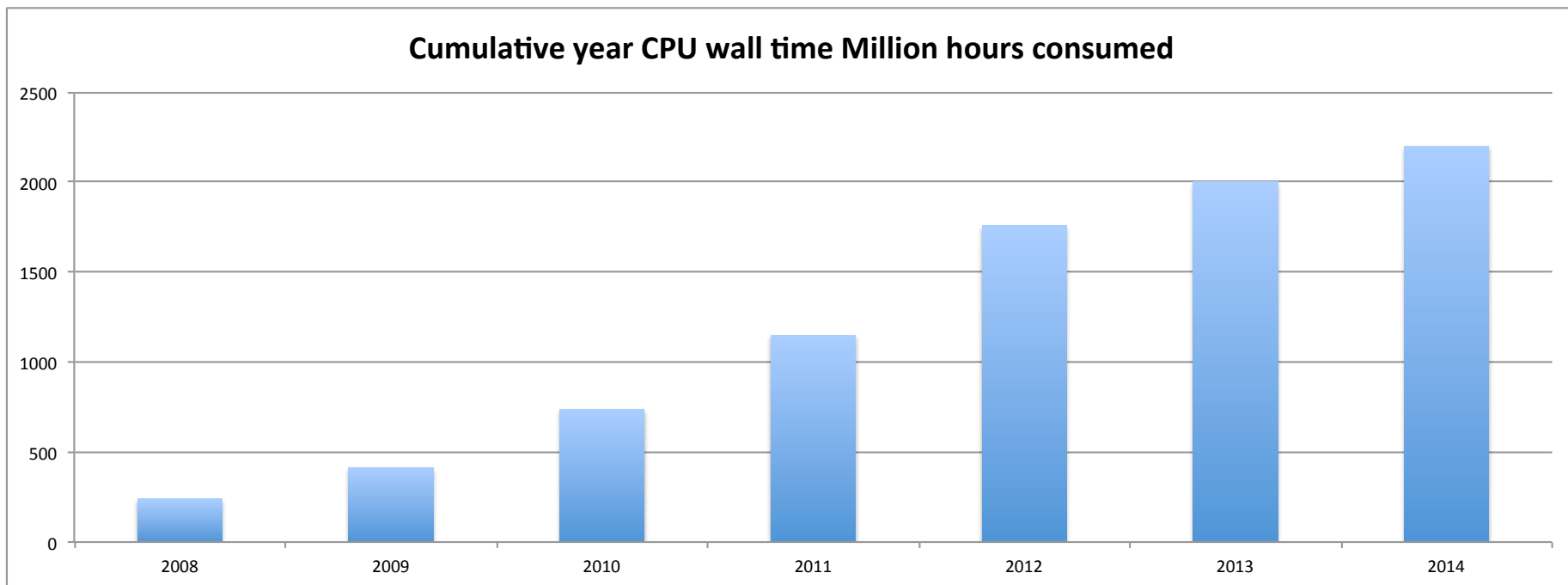


- Distributed and federated data and computing facilities
- HTC and Cloud compute platform
- 340 data centres in 34 National Grid Initiatives/EIROs
- Partnerships with non-EU infrastructures
- > 200 International Research Communities
- Domains-specific Software & Tools
- The HTC federation in numbers:
 - 525,000 cores, 1.5 M jobs/day
 - 300 PB disk capacity
 - 200 PB tape capacity

- Launch: May 2014
- 19 cloud providers from 12 countries
 - Czech Republic, Germany, Greece, Hungary, Italy, Macedonia, Poland, Slovakia, Spain, Sweden, Turkey, United Kingdom
 - Publicly funded and commercial
- 4 countries currently integrating
 - France, Finland, Portugal, South Korea
- Coming: Canada (CANFAR), D4SCIENCE
- Worldwide interest
 - South Africa (SAGrid)
 - United States* (NIST, NSF Centres)
 - Australia* (NeCTAR)



Cumulative year CPU wall time Million hours consumed



- of which **1,000 Million CPU hours** for Atmospheric sciences, Biodiversity conservation, Climate research, Earth Sciences, Ecology, Environmental engineering, Geophysics, Hydrology, Space Science

Making sense of seismic noise

How grid computing helps scientists to correlate data from millions of calculations to unveil the rock structure of an oil field under the North Sea.

Oil and mining companies use seismic waves to figure out the structure and type of rocks underground, so they can plan their work. Because they can't wait for earthquakes, they artificially induce seismic waves with explosions. This is however expensive and has severe environmental consequences.

Aurélien Mordret, a seismologist based in France, is working on an alternative that puts the background seismic noise to good use. Thanks to a mathematical method called cross-correlation and grid computing, it is possible to collect meaningful seismic data from the rumble of waves as they travel across the seafloor.

When applied to the Valhall oil field in the North Sea, the method revealed a structure crisscrossed with old riverbeds and the telltale signs of a depleted oil reservoir.



An oil and gas exploration platform in the North Sea.

Image: JanChr / wikicommons

Modelling earthquakes in Thessaloniki

Predicting earthquakes is impossible, but grid computing helps to anticipate their effects.

Earthquakes are amongst the world's most destructive natural disasters. Just in 2011, an earthquake destroyed the town of Christchurch in New Zealand, while another triggered the tsunami in Japan.

It's impossible to predict when a catastrophe is good plan

Protecting Portugal's Aveiro Lagoon

To better manage the delicate ecosystem of the Aveiro Lagoon in Portugal, we need to understand the diverse factors affecting its balance. This can be done using complex computer modelling techniques, which demand a lot of computing power. Using grid computing speeds things up, thereby helping to improve coastal resource management.

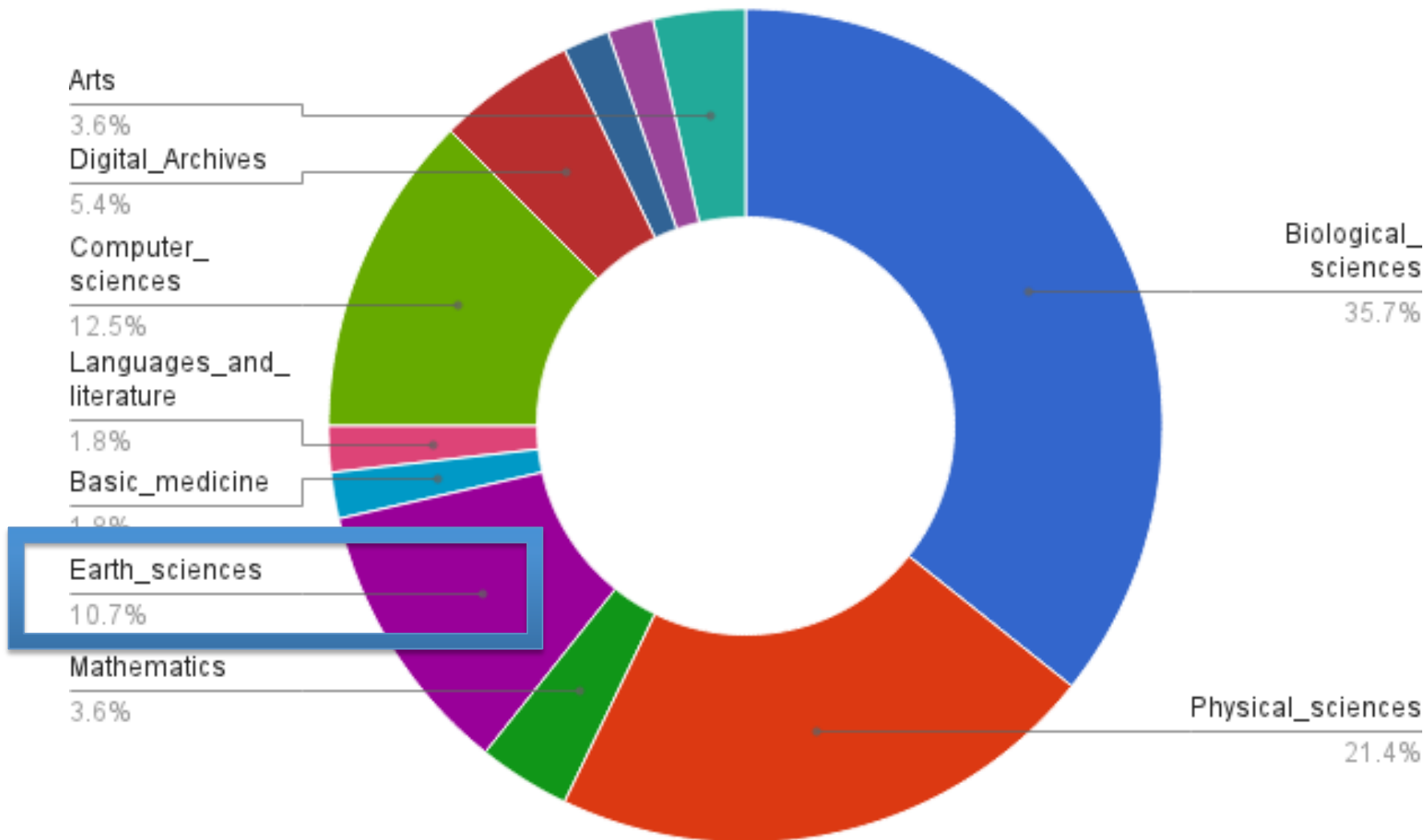


A gull perched on top of the mast of a 'moliceiro' (a traditional Portuguese

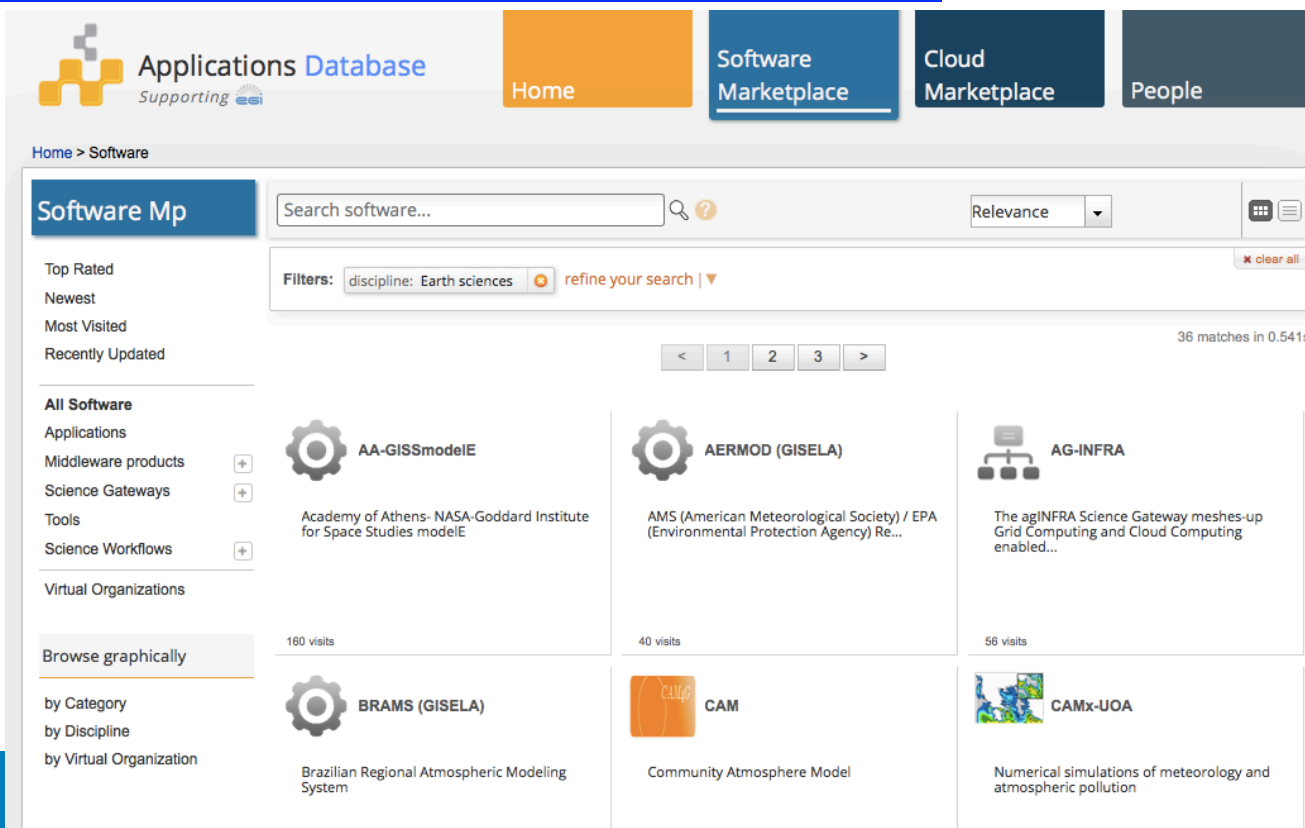
try.

<http://www.egi.eu/case-studies/index.html>

Cloud Use Cases Discipline Classification (2° tier)



- Build communities for scientific codes for sharing and reuse of applications and science workflows: 36 for earth science
<http://go.egi.eu/AppDB-earthscience>



The screenshot displays the 'Applications Database' web interface. At the top, there are navigation buttons for 'Home', 'Software Marketplace', 'Cloud Marketplace', and 'People'. The main content area is titled 'Software Mp' and features a search bar with the text 'Search software...'. Below the search bar, there are filters for 'discipline: Earth sciences' and a 'refine your search' option. The search results are displayed in a grid format, showing six software applications:

- AA-GISSmodelE**: Academy of Athens- NASA-Goddard Institute for Space Studies modelE (160 visits)
- AERMOD (GISELA)**: AMS (American Meteorological Society) / EPA (Environmental Protection Agency) Re... (40 visits)
- AG-INFRA**: The agINFRA Science Gateway meshes-up Grid Computing and Cloud Computing enabled... (56 visits)
- BRAMS (GISELA)**: Brazilian Regional Atmospheric Modeling System
- CAM**: Community Atmosphere Model
- CAMx-UOA**: Numerical simulations of meteorology and atmospheric pollution

The interface also includes a sidebar with navigation options like 'Top Rated', 'Newest', 'Most Visited', 'Recently Updated', and 'Browse graphically'.

- DRIHM and VERCE
 - E.g. Running various hydrological models in the EGI Federated Cloud
- EPOS Competence Centre
 - Design of the use of grid and cloud for the integrated solid Earth Sciences research as part of the European Plate Observing System
 - Identify and validate authentication and authorisation services for EPOS
 - Test cloud resources and usage models
 - compute/data resources for applications
 - hosting services
 - Science gateway (no specific platform yet)
 - [EGI/EUDAT service interoperability use cases](#)
- ESA: e-Collaboration for Earth Observation (e-CEO) platform
 - support online contexts where researchers can work collaboratively and compare and evaluation different problem-solving approaches
- Centre of Excellence for environmental sciences – HTC/HPC/Cloud (EINFRA-5 project proposal EnCompAS)