EGI: advanced computing for research in Europe… and beyond!

Yannick LEGRÉ
Managing Director
EGI Foundation

RO-LCG 2016 Conference – Bucharest (Romania)
26-28 November 2016

www.egi.eu

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program under grant number 654142

EGI: Advanced Computing for Research

The EGI Services are provided by the EGI Federation

EGI Council participants: national e-infrastructure providers and international research organisations (CERN and EMBL)

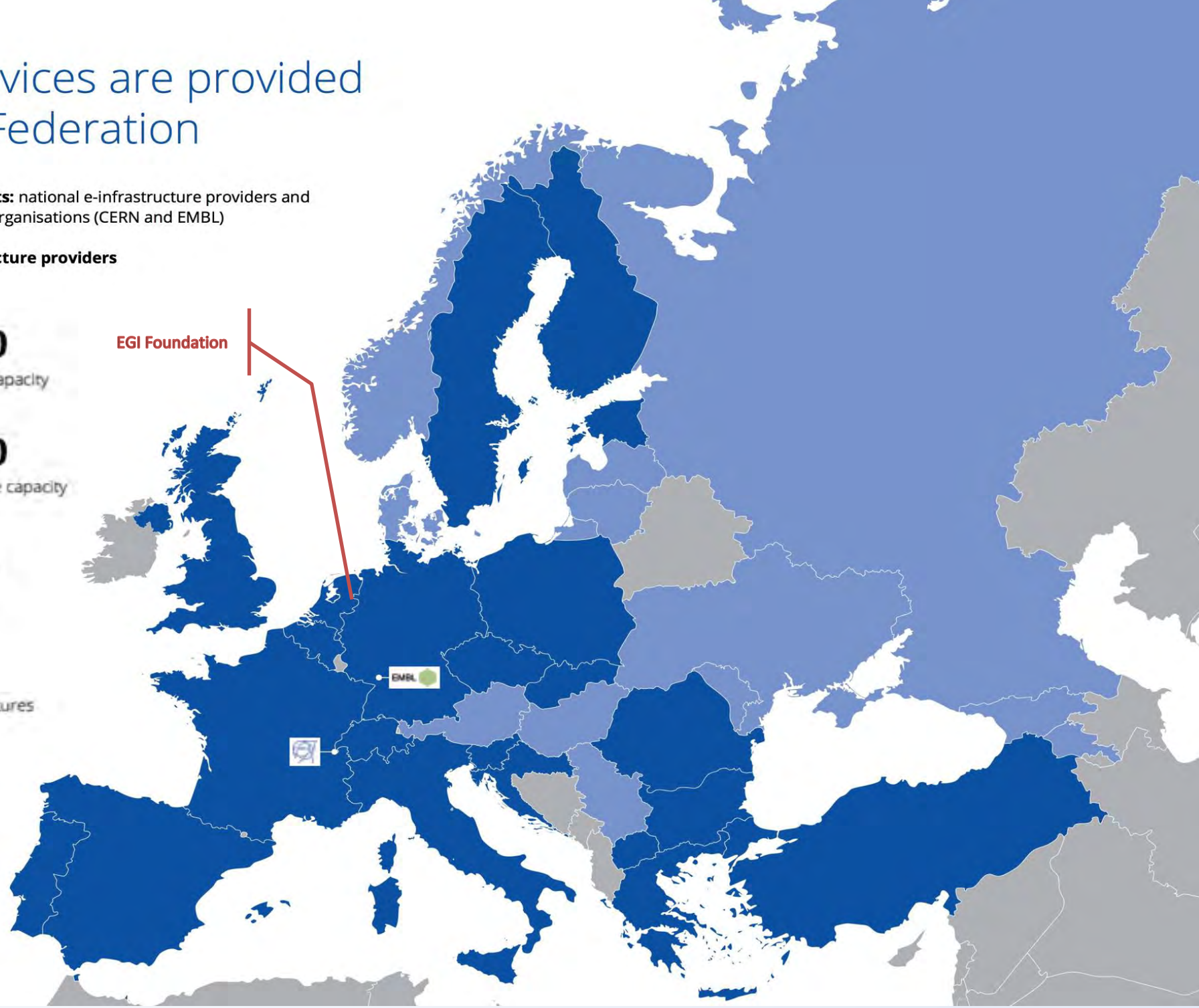
Integrated e-infrastructure providers

826,000
Cores of compute capacity

560,000
Terabytes of storage capacity

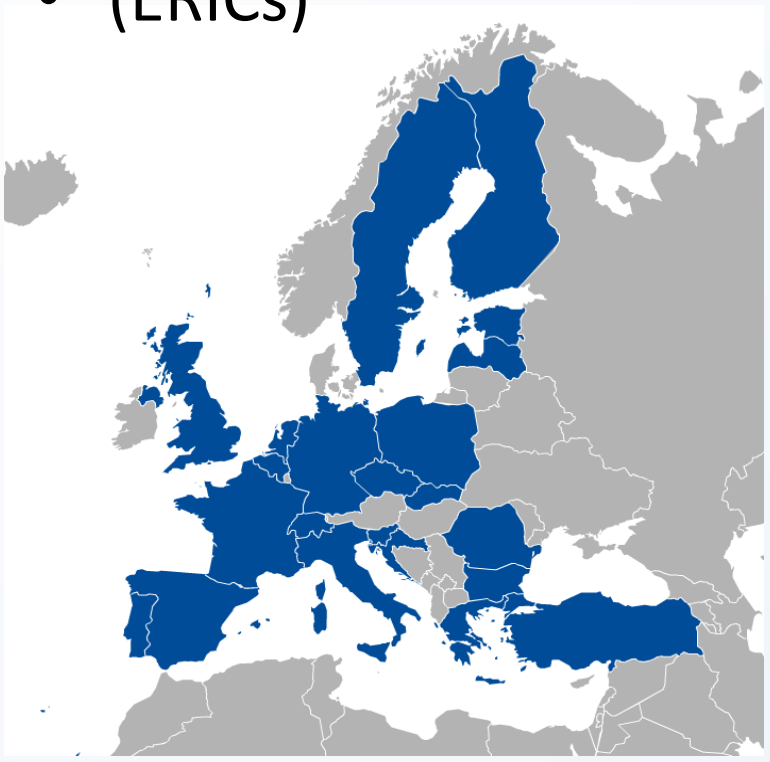
48,000
Users

15
Research Infrastructures integrated with EGI



EGI Membership

- Major national e-Infrastructures: 22 NGIs
- EIROs: CERN and EMBL-EBI
- EGI Foundation
- (ERICs)



belspo
IICT
Institute of Information and Communication Technologies
SwiNG
Supporting Swiss Scientific Computing
CESNET
EENet
Eesti Hariduse ja Teaduse Andmesidevõrk
CSIC
Consejo Superior de Investigaciones Científicas
CSC
cnrs

GA Gauß-Allianz
grnet
srce
INFN
MARGI
SURF
CYFRONET

IFIN-HH
arnes
SNIC
TÜBITAK ULAKBİM
JISC
EMBL
CERN

www.egi.eu/about/egi-foundation/

International Partnerships

EGI Federation, 2016 QR3

The largest distributed compute e-Infra worldwide

23 Cloud providers, +300 data centres

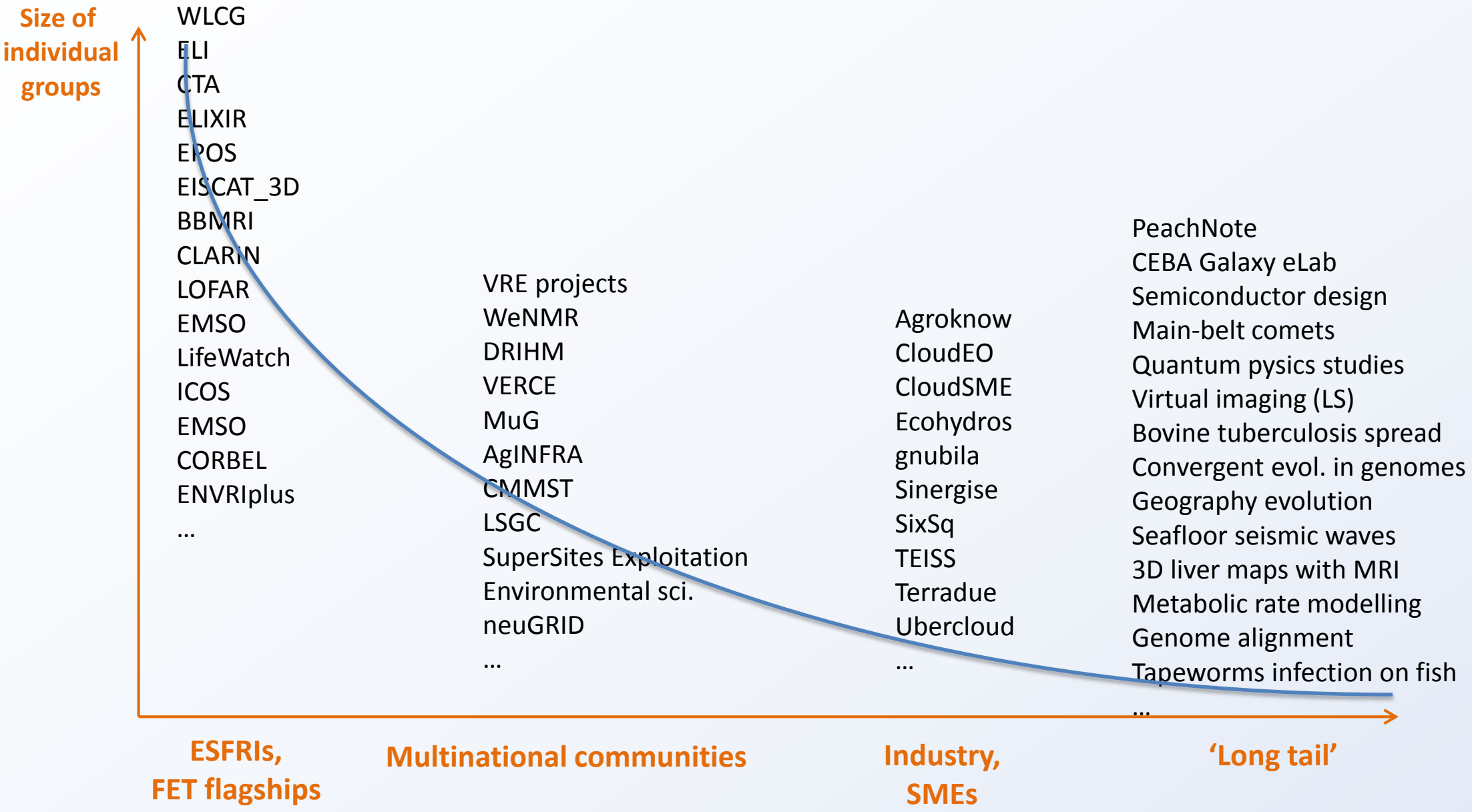
+250 000 instantiated VMs/year

1.7 Million jobs/day

2.6 Billion CPU hours/year +26%

>48 000 users, +25%

Serving researchers and innovators



EGI Service Catalogue

Compute

Cloud Compute

Run virtual machines on demand with complete control over computing resources

Cloud Container Compute

Run Docker containers in a lightweight virtualised environment

High-Throughput Compute

Execute thousands of computational tasks to analyse large datasets

Storage and Data

Online Storage

Store, share and access your files and their metadata on a global scale

Archive Storage

Back-up your data for the long term and future use in a secure environment

Data Transfer

Transfer large sets of data from one place to another

Training

FitSM training

Learn how to manage IT services with a pragmatic and lightweight standard

Training infrastructure

Dedicated computing and storage for training and education

Cloud Compute

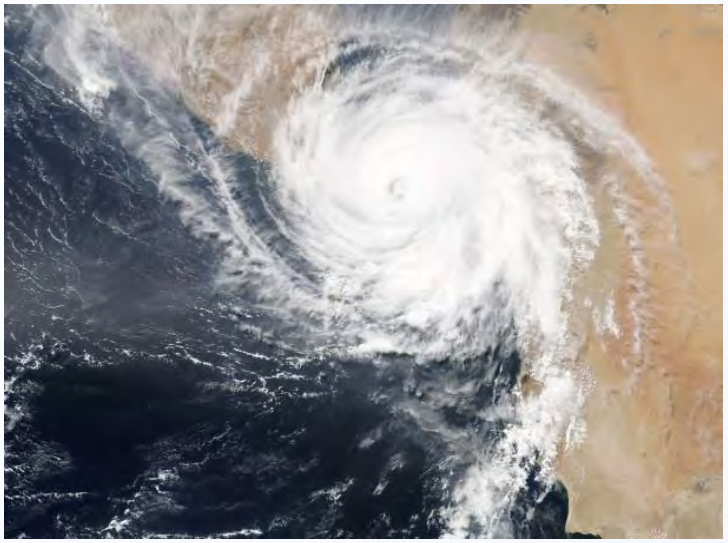
Run virtual machines on-demand with complete control over the computing resources

With Cloud Compute you can:

- Execute compute- and data-intensive workloads
- Host long-running services (e.g. web servers or databases)
- Create disposable testing and development environments
- Select virtual machine configurations to fit your requirements
- Manage your Cloud Compute resources in a flexible way with integrated monitoring and accounting capabilities

See Cloud Compute for service information and request

Powered by Cloud Compute



DRIHM Project

- prototype an e-infrastructure to simulate extreme weather events

EXTraS Project

- implement four software pipelines to harvest data collected on-board ESA's space observatory XMM-Newton.

Powered by Cloud Compute

Nature (doi:10.1038/nature16547)

Read more...

When a human cell meets *Salmonella*

- K. Förstner, Univ. Würzburg, used Cloud Compute to run a pipeline for the analysis of sequencing data.
- Cloud Compute helped the team to handle demand peaks and that sped up the whole process significantly.

Powered by Cloud Compute

Read more...

Chipster
Open source platform for data analysis

Chipster

- Chipster is a user-friendly analysis software for high-throughput data
- Users can save and share automatic analysis workflows, and visualize data interactively
- Relies on Cloud Compute

Cloud Container Compute

Run Docker containers in a lightweight virtualised environment

Main features of Cloud Container Compute:

- On-demand provisioning
- Lightweight environment for maximised performance
- Standard interface to deploy on multiple service providers
- Interoperable and transparent
- Removes friction between development and operations environments.

See Cloud Container Compute for service information and request

High-Throughput Compute

Execute thousands of computational tasks to analyse large datasets

Main features of High-Throughput Compute:

- Access to high-quality computing resources
- Integrated monitoring and accounting tools to provide information about the availability and resource consumption
- Workload and data management tools to manage all computational tasks
- Large amounts of processing capacity over long periods of time

See High-Throughput Compute for service information and request

Powered by High-Throughput Compute

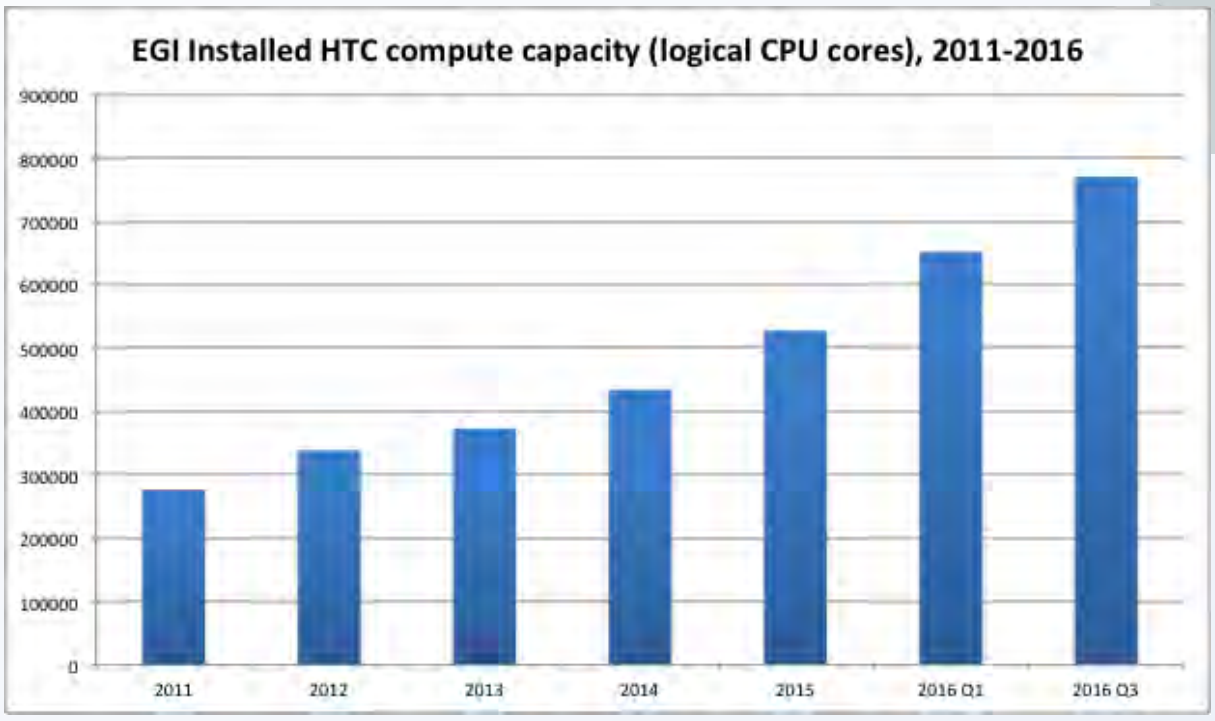
Read more...

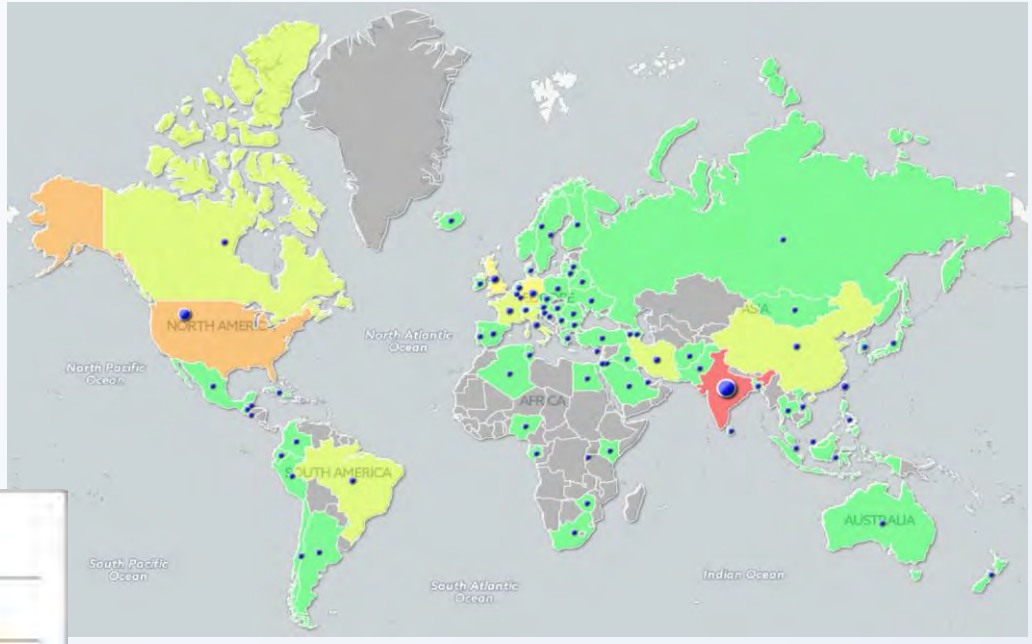
HADDOCK

- A web portal offering tools for structural biologists
- Used to model the structure of proteins and other molecules.
- So far, HADDOCK processed + 130,000 submissions from over 7,500 scientists.

Supporting international research communities and thematic services

Installed compute capacity trends 2011-2016





Distribution of Structural Biology user community

- 2700 users
- 81 countries

(credits: A. Bonvin, WeNMR)

High-Throughput Compute capacity

826,500 cores

See High-Throughput Compute for service information and request

Online Storage

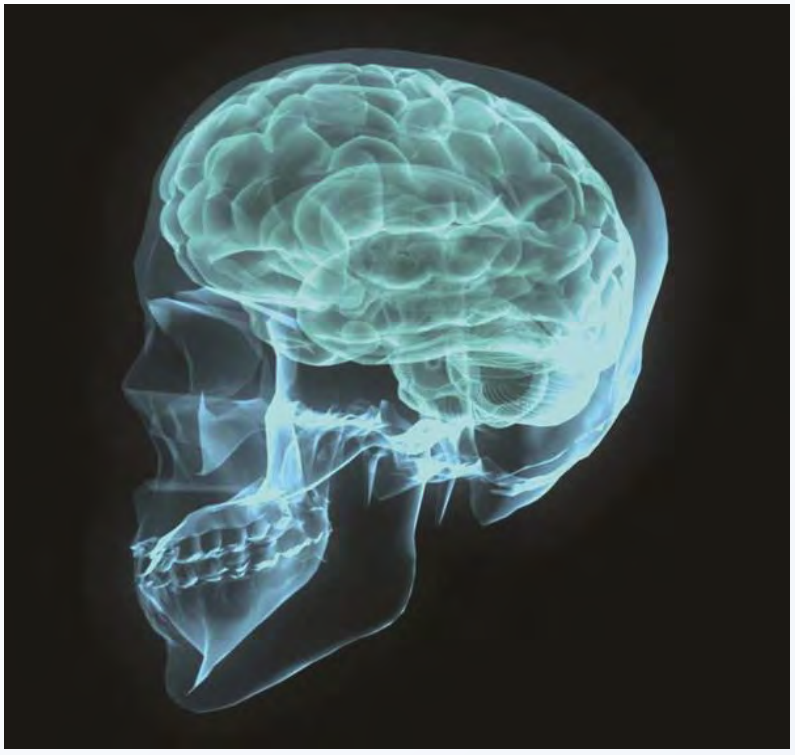
Store, share and access your files and their metadata on a global scale

Main features of Online Storage:

- Assign global identifiers to files
- Access highly-scalable storage from anywhere
- Control the data you share
- Organise your data using a flexible hierarchical structure

See Online Storage for service information and request

Powered by Online Storage and High-Throughput Compute



Read more...

Virtual Imaging Platform

- VIP is a web portal for medical image data analysis
- It is used by researchers worldwide
- VIP relies on High-Throughput Compute and Online Storage

Online Storage

285 Petabytes

285,000,000 GB

11,400,000 Blu-ray Disc

See Online Storage for service information and request

Archive Storage

Back-up your data for the long term and future use in a secure environment

Main features of Archive Storage:

- Store large amount of data
- Free up your online storage
- Store data for long-term retention

See Archive Storage for service information and request

Archive Storage

280 Petabytes

280,000,000 GB

11,200,000 Blu-ray Disc

See Archive Storage for service information and request

Data Transfer

Transfer large sets of data from one place to another

Main features of Archive Storage:

- Ideal for very large files
- Able to handle large amounts of files
- Transfer process with automatic retry

> "The most critical infrastructure and tools are the networks and the online storage and data transfer services (…). Without these the analysis of the data from the LHC would be almost impossible."
>
> **Ian Bird, WLCG project leader**

See Data Transfer for service information and request

Online Storage and Data Transfer at unprecedented scales

- *The Worldwide LHC Computing Grid (WLCG) is a global collaboration of more than 170 computing centres set up to provide the computing resources needed to store, distribute and analyse the data generated by the Large Hadron Collider (LHC) at CERN.*

- *During 2016, WLCG transferred on average 80 Petabytes of data per month, with peaks at 96 Petabytes during summer. This corresponds to more than 1 billion files per month transferred to thousands of particle physicists working across the world.*

"WLCG must manage hundreds of petabytes of data, with more than 10PB of new data being added each month to the LHC runs. Without these, the analysis of the data from the LHC would be almost impossible."

Ian Bird, WLCG project leader

FitSM Training

Learn how to manage IT services with a pragmatic and lightweight standard

With FitSM Training you can:

- Increase your expertise in managing IT services
- Raise your professional profile by a recognised certification

> "I learned how to implement FitSM in an IT organisation and gained from the benefits the framework provides for efficient service management. **Pavel Weber, KIT**

See FitSM Training for service information and request

Training Infrastructure

Dedicated computing and storage for training and education

Main features of the Training Infrastructure

- Target-specific courses and added value for scientific communities
- Easy-to-use, on-demand access and improvements in the training offer
- Allows easy deployment of courses and reuse

See Training Infrastructure for service information and request

E-Infrastructure services enable the Open Science Vision

The European Cloud Initiative

- *European Open Science Cloud (EOSC)*
  - Integration and consolidation of e-infrastructures
  - Federation of existing research infrastructures and scientific clouds
  - Development of cloud-based services for Open Science
  - Connection of ESFRIs to the EOSC

- *European Data Infrastructure (EDI)*
  - Development and deployment of large-scale European HPC, data and network infrastructure

- *Widening access*
  - SMEs, Industry at large, Government

Courtesy of the European Commission

E-Infrastructures are the foundation of the European Open Science Cloud

https://ec.europa.eu/futurium/en/content/e-infrastructures-making-europe-best-place-research-and-innovation

Courtesy of the European Commission

In short …

The European Open Science Cloud will encompass data, computing and networking services for the benefit of the whole scientific community and beyond

But…

- The European Union provide only 10 – 25% of the necessary funding. Member states provide the remaining 75 -90%
- The European Open Science Cloud can only be achieved by aggregating National Open Science Cloud
- National Open Science Clouds have to rely on 4 major pillars
  - A strong political push
  - Committed Scientific Communities
  - National pluridisciplinar (otherwise thematic) e-Infrastructures
  - Long-term and sustainable funding

It is amazing what you can accomplish if you do not care who gets the credit.

Harry S. Truman

Get in touch!

@EGI_eInfra

EGI Foundation • Science Park 140 • 1098 XG Amsterdam •
The Netherlands
+31 (0)20 89 32 007 • egi.eu