EGI: advanced computing for research

Yannick Legré

Managing Director
yannick.legre@egi.eu

CREMLIN Workshop, Moscow, 15th Feb. 2017

www.egi.eu

EGI: advanced computing for research

EGI is a federation of over 300 computing and data centres spread across 56 countries in Europe and worldwide

EGI delivers services to support scientists, multinational projects and research infrastructures

www.egi.eu

EGI's mission is to create and deliver

open solutions for science and research

by federating digital capabilities,

resources and expertise

between communities and

across national boundaries.

EGI Foundation

- **22** national e-Infrastructure providers
- Two EIROs: **CERN** and **EMBL**
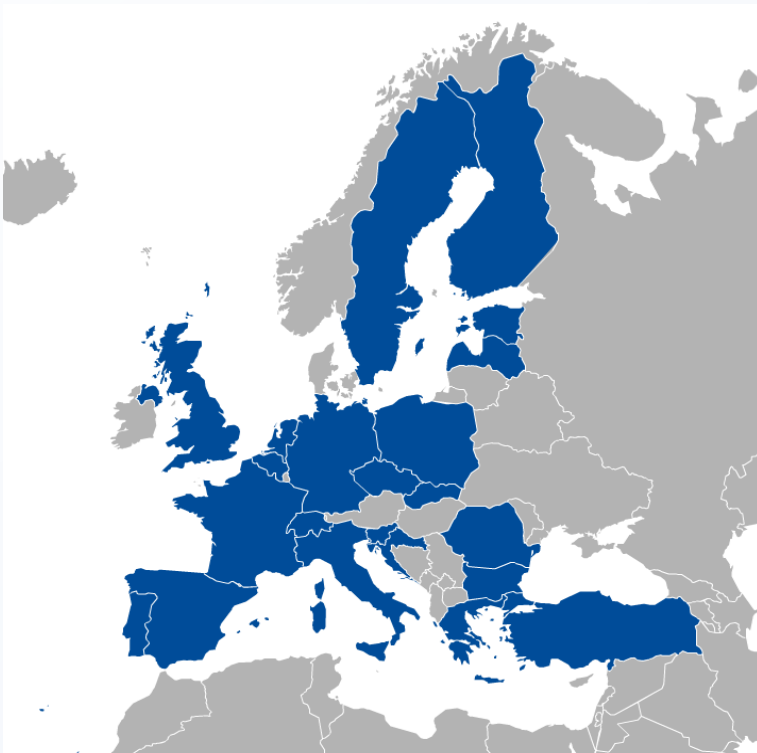
www.egi.eu/about/egi-foundation/

EGI Federation, 2016 QR4

The largest distributed compute e-Infra worldwide

~750.000 CPU cores

~565PB of storage

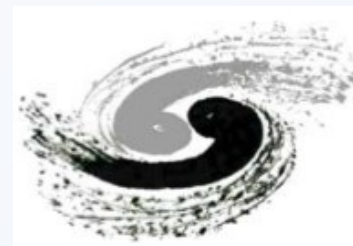1.7 Million jobs/day

2.6 billion CPU hours/year (+26%)

~50K users (+25%)

A global system of e-Infrastructures

Africa-Arabia, Asia and Pacific region, China, Europe, India, North and South America, Russian Federation

Europe

China:
Institute of HEP,
Chinese Academy of Sciences

Africa and Arabia:
Council for Scientific and
Industrial Research, South Africa

Latin America:
Universida de Federal do
Rio de Janeiro

India:
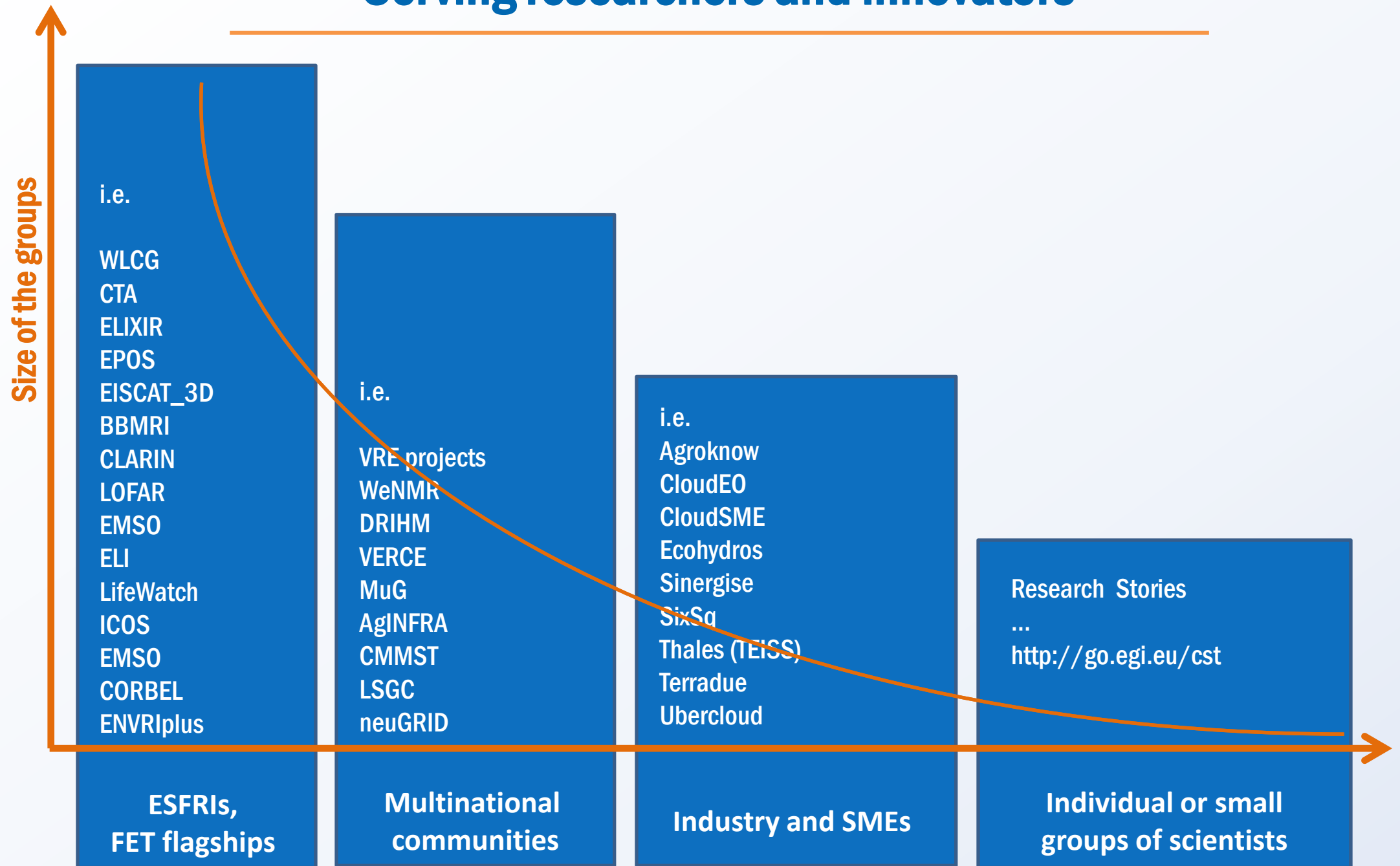Centre for Development of
Advanced Computing

Canada

USA

Russian Federation: NRC KI, JINR, …

Serving researchers and innovators

Size of the groups

i.e.

WLCG
CTA
ELIXIR
EPOS
EISCAT_3D
BBMRI
CLARIN
LOFAR
EMSO
ELI
LifeWatch
ICOS
EMSO
CORBEL
ENVRIplus

i.e.

VRE projects
WeNMR
DRIHM
VERCE
MuG
AgINFRA
CMMST
LSGC
neuGRID

i.e.
Agroknow
CloudEO
CloudSME
Ecohydros
Sinergise
SixSq
Thales (TEiSS)
Terradue
Ubercloud

Research Stories
...
http://go.egi.eu/cst

**ESFRIs,
FET flagships**

**Multinational
communities**

Industry and SMEs

**Individual or small
groups of scientists**

A wide offer of services for Big Data

EGI Live Services

Compute

Cloud Compute : Run virtual machines on demand with complete control over resources

Cloud Container Compute : Run Docker containers in a lightweight virtualised environment

High-Throughput Compute : Execute thousands of 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-ups for 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

EGI Services in the pipeline

Operations

Service Monitoring: Monitor a wide range of services across your infrastructures

Configuration Database: Manage the configuration information of federated e-infrastructure assets and their functional relations

Data Management

Data Hub: Selected big datasets of public interest made exploitable – https://datahub.egi.eu

Federated Data Manager: Share, discover, and process data federated from different sources

Content Distribution: Deliver data in the most efficient way

Security

Attribute Management: Manage community membership and expose trusted information

Check-In: Handle transparent Single Sign-On from multiple heterogeneous identity providers

EGI Services for Participants

Operations

Accounting Track and report the usage of your services

Helpdesk Handle EGI service requests and incidents for distributed support teams

Operations Tools Integrate resources and operations

Collaboration and Community Management Tools Coordinate activities to ensure seamless operations

Configuration Database Manage the configuration information of federated e-infrastructure

Service Monitoring Monitor EGI services and provide operational and business insights

Validated Software and Repository Manage high-quality software releases

Coordination

Communications Share your successes at a larger scale

ITSM Coordination Ensures professional service management for EGI IT services

Community Coordination A joint approach to user engagement

Operations Coordination and Support Coordinate activities to ensure seamless operations

Security Coordination Enhance local security for a safer global infrastructure
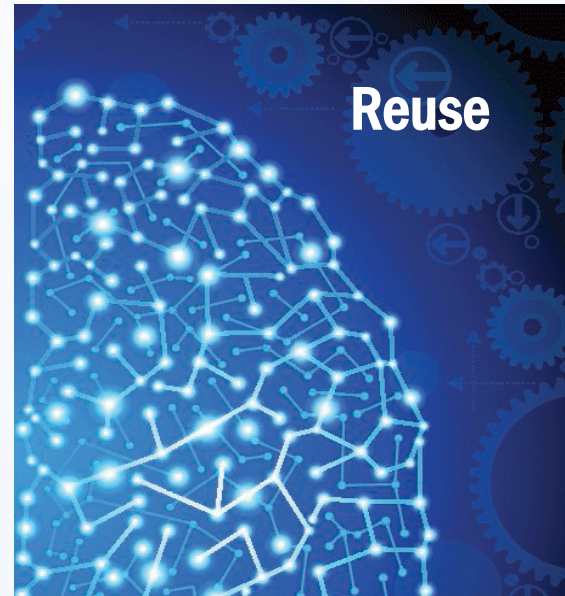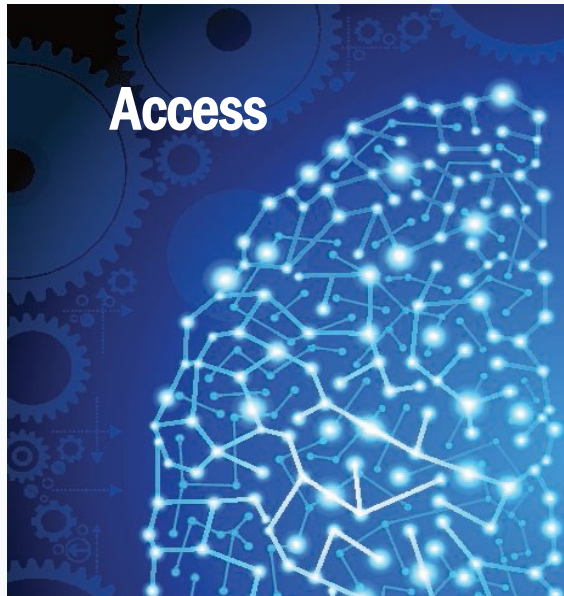
Project Management and Planning A joint approach to planning and management

Strategy and Policy Development Define common strategies and policies in Europe and worldwide

Technical Coordination Progress and innovation through collaboration

Security

Attribute Management: Manage community membership and expose trusted information

Services offered to SMEs and Industry

✓ Pilots
✓ Proof of concepts

Access

Reuse

✓ Datasets
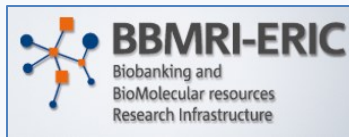✓ Tools/Software
✓ Applications

✓ EU visibility
✓ Networking

Market

Co-design

✓ Products/Services
✓ Integrate
✓ Share expertise

https://www.egi.eu/business/

(Big Data)

Use Cases

Cooperating with Research Infrastructures

- Joint collaborations with key EU Research Infrastructures and communities
- Co-development of customised EGI services
- E.g.:

WLCG: Online Storage and Data Transfer at unprecedented scales

ATLAS control room at CERN

In 2016, the Worldwide LHC Computing Grid (WLCG) transferred on average **80 Petabytes of data per month.**

This corresponds to more than **1 billion files per month** transferred to **thousands of particle physicists** working across the world.

The WLCG is the largest resource provider and service consumer of the EGI Federation

Cherenkov Telescope Array: the world's leading gamma-ray observatory

Credit: Akihiro Ikeshita Mero-TSK

1350 scientists from 32 countries

CTA EGI usage (2013-2016):

- 360 million HS06 CPU hours

- 11 PB of data transferred

- 2 PB currently in storage

- 11 million compute jobs

CTA used EGI's High-Throughput Compute and Online Storage services to manage its computational challenges during the array preparatory phase.
http://go.egi.eu/cta

Supporting the Geohazard Exploitation Platform

Very fast access to large volumes of data

- Earth Observation data

- In-situ data

Distributed processing

- Hybrid cloud infrastructure

- Access to HTC resources

Making sense of ocean salinity & temperature data

http://www.euro-argo.eu/

Euro-Argo

- Operates an array of ca. 800 floating sensors

- Enhances coverage in the Europe's regional seas

- Provides access to quality controlled data and data products

Hadoop clusters for data filtering are being deployed and evaluated in EGI Federated Cloud

Understanding the genetics of *Salmonella* infections

Published in *Nature*
(doi:10.1038/nature16547)

When human cell meets *Salmonella*

K. Förstner (Univ. Würzburg) used Cloud Compute to run READemption - 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.

Discovering new viruses implicated in a fatal snake disease

http://chipster.csc.fi/

Chipster

A user-friendly analysis software for high-throughput data with over 350 analysis tools and a large collection of reference genomes

Users can save and share automatic analysis workflows, and visualize data interactively

Predicting biomolecular interactions at the molecular level

http://haddock.science.uu.nl/

HADDOCK

A web portal offering tools for structural biologists

Used to model the structure of proteins and other molecules.

So far, HADDOCK processed over 130,000 submissions from more than 7,500 scientists.

We are open for collaborations!

EGI is

An open ecosystem for data-/compute-intensive applications

Bridging research and innovation

An enabler for stimulating the data value chain

Follow us:
@EGI_eInfra
@EGIeInfrastructure

Discover more

Visit www.egi.eu/

Services www.egi.eu/services/

Publications www.egi.eu/about/publications/

Contact us:

- For public institutions: contact@egi.eu
- For private organisations: business@egi.eu