Handling time–critical service applications with EGI e-Infrastructure

Joao Pina, EGI Foundation
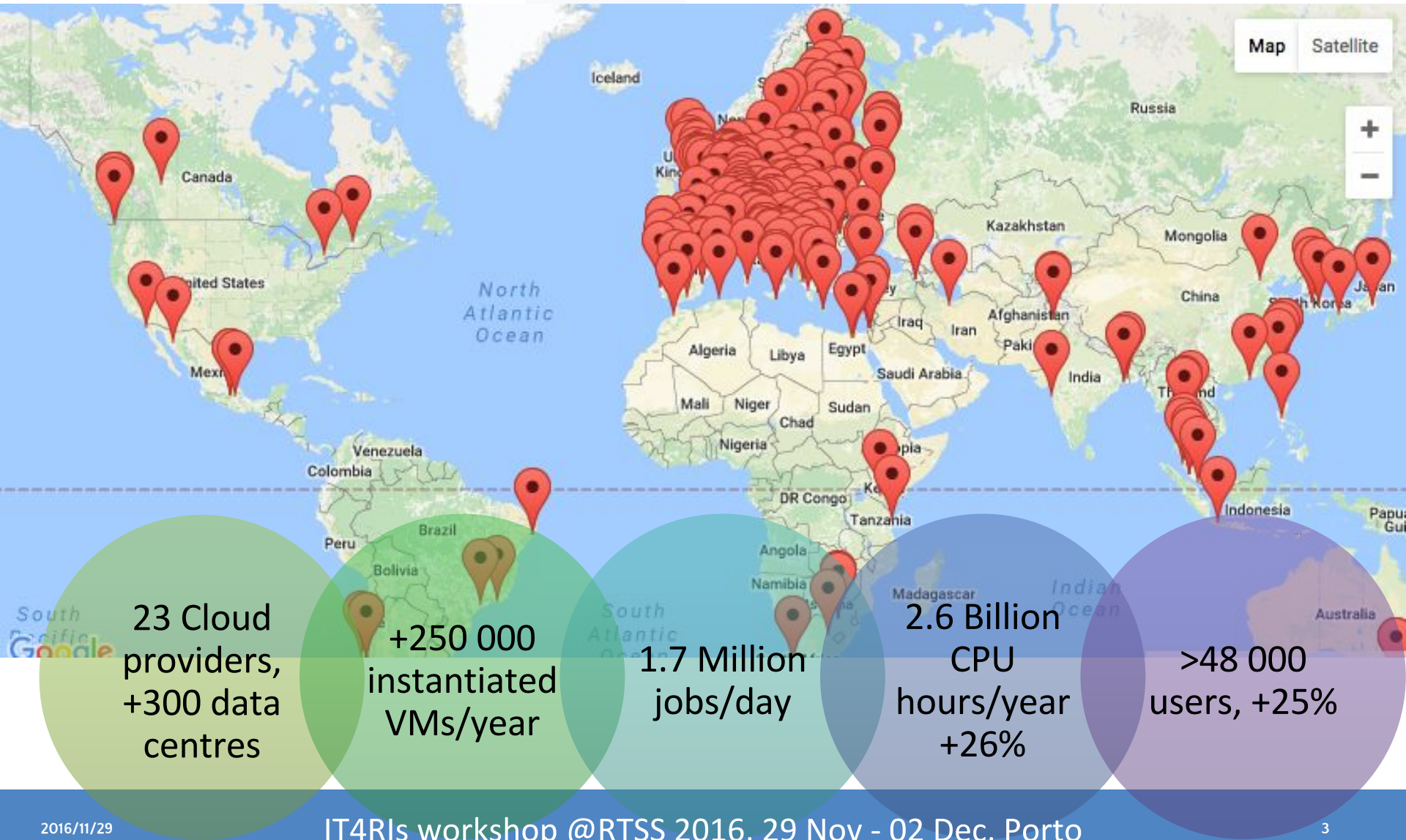
Portuguese NGI representative
INCD / LIP

EGI: advanced computing for research

EGI…

…delivers advanced computing services to support scientists, multinational projects, research infrastructures and industry

www.egi.eu

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%

Service Catalogue

www.egi.eu

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 >
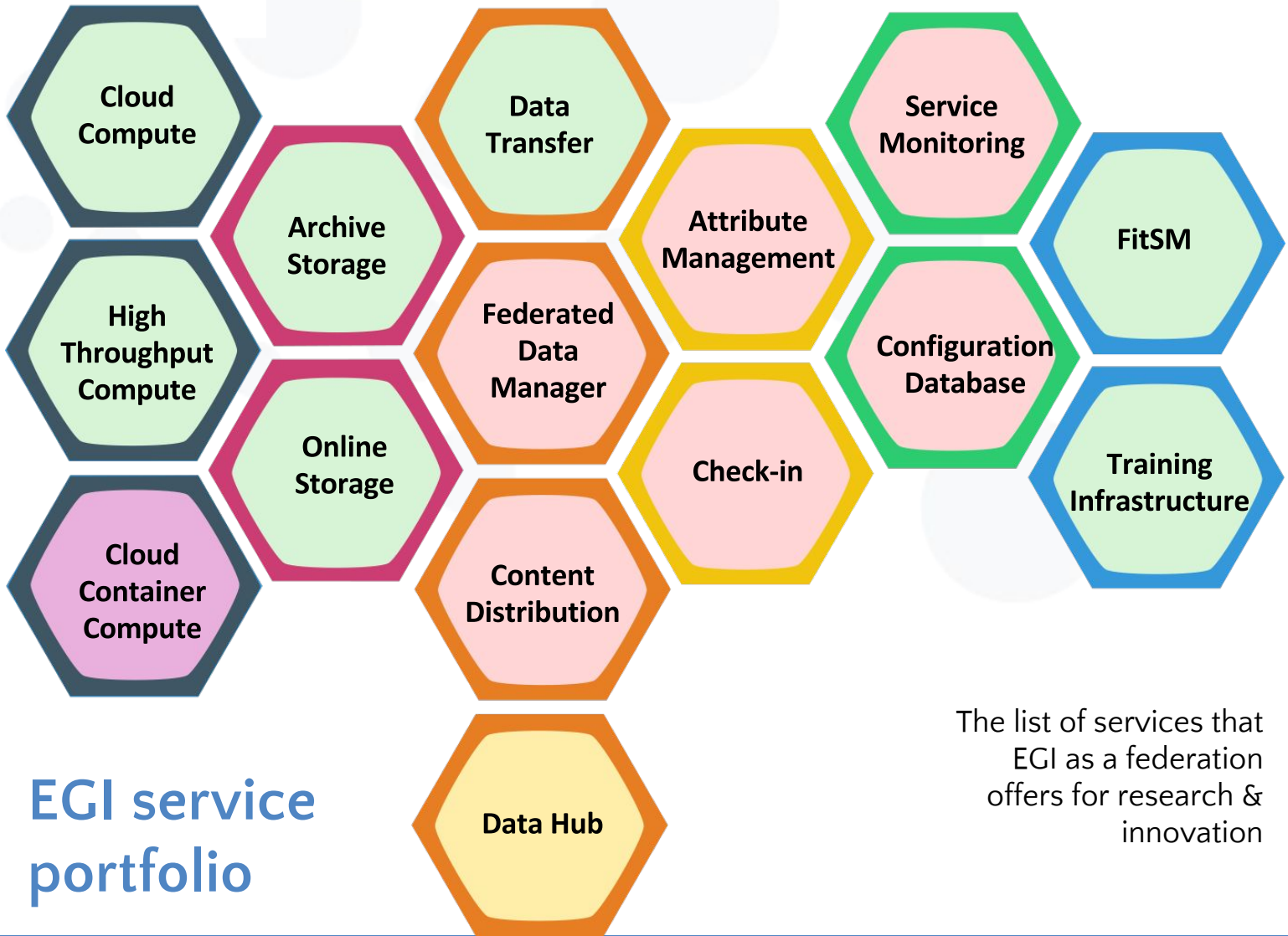Transfer large sets of data from one place to another

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

Service Catalogue
EGI: advanced computing for research
www.egi.eu

http://go.egi.eu/ServiceCatalogue

EGI service portfolio

Compute	Storage	Data	Security	Operations	Training

Compute
- Cloud Compute
- High Throughput Compute
- Cloud Container Compute

Storage
- Archive Storage
- Online Storage

Data
- Data Transfer
- Federated Data Manager
- Content Distribution
- Data Hub

Security
- Attribute Management
- Check-in

Operations
- Service Monitoring
- Configuration Database

Training
- FitSM
- Training Infrastructure

The list of services that EGI as a federation offers for research & innovation

Design Phases

- Discovery
- Alpha
- Beta
- Production
- Retired

www.egi.eu

- *EGI Federated Cloud*

EGI Federated Cloud
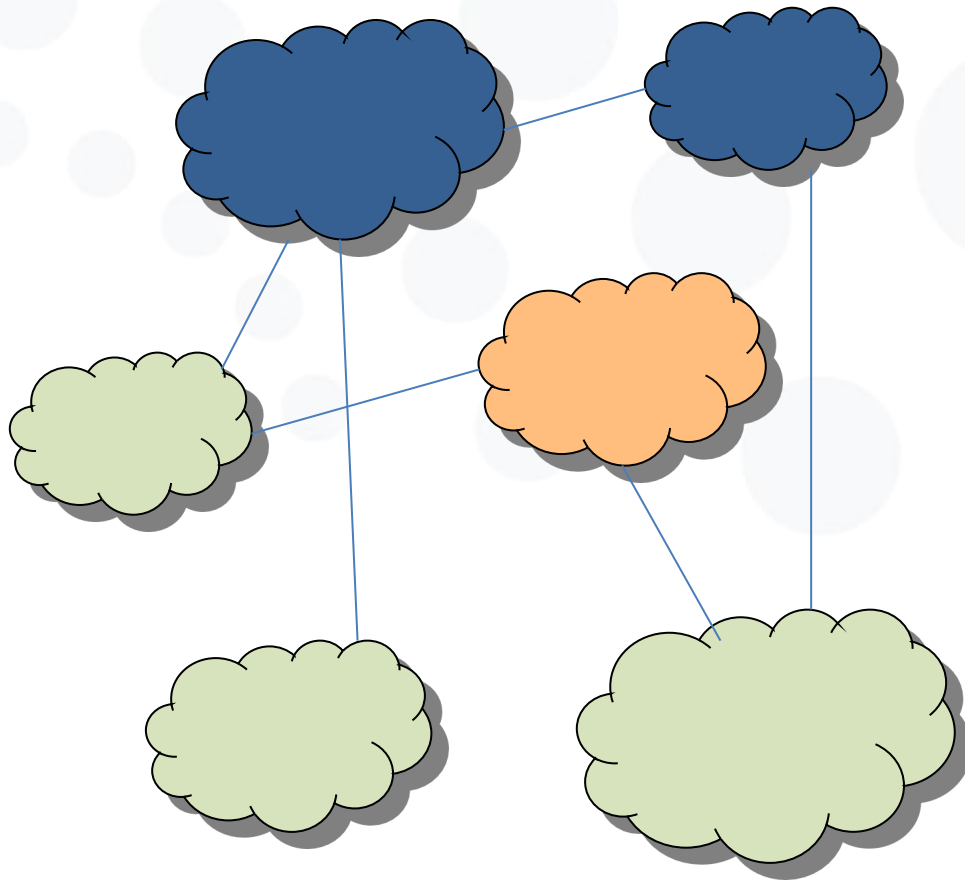
- Cloud of clouds
- Unified user interfaces
- Harmonised operational behaviour
- Clouds and their interconnections are based on open standards, open technologies
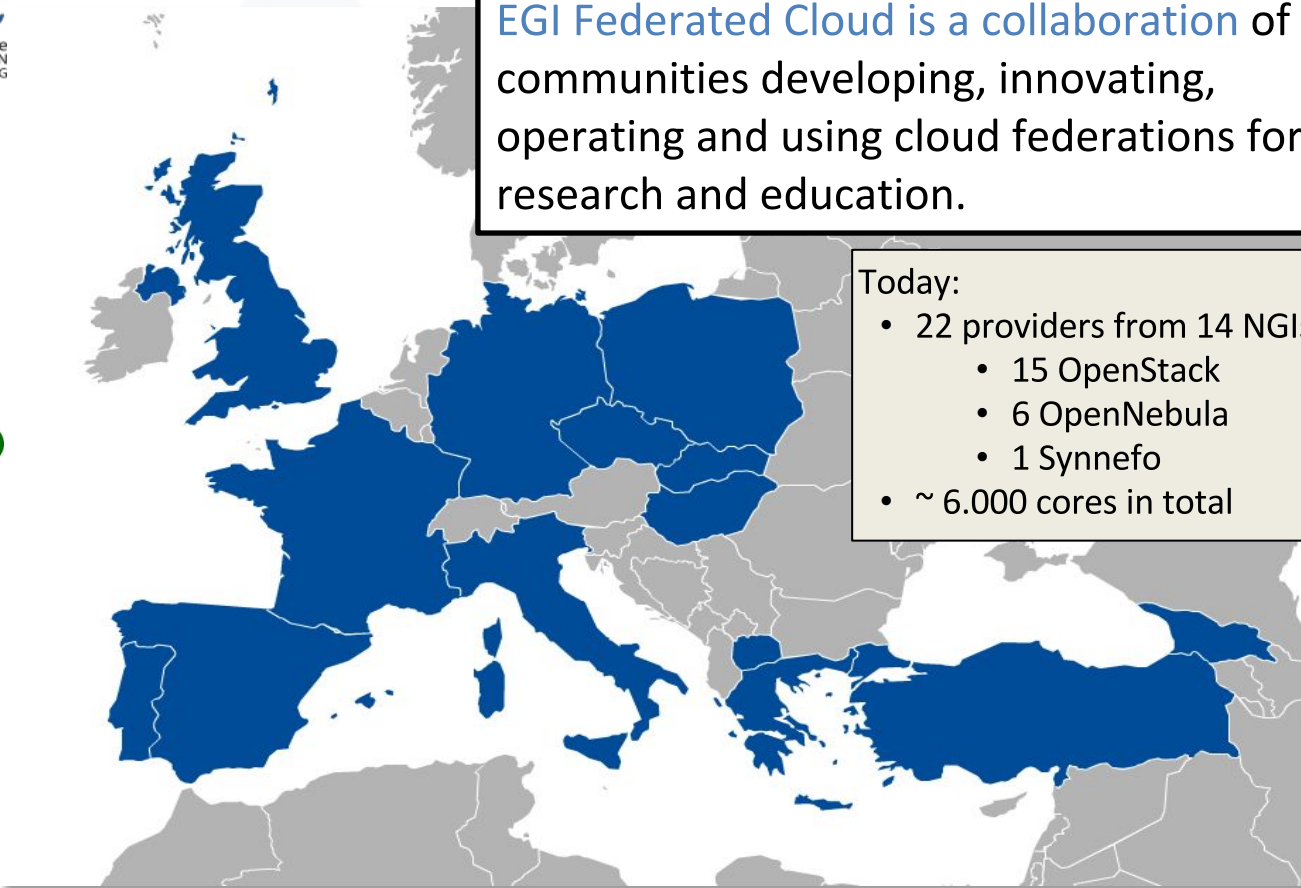- Infrastructure → Access AND
  technology → Deploy
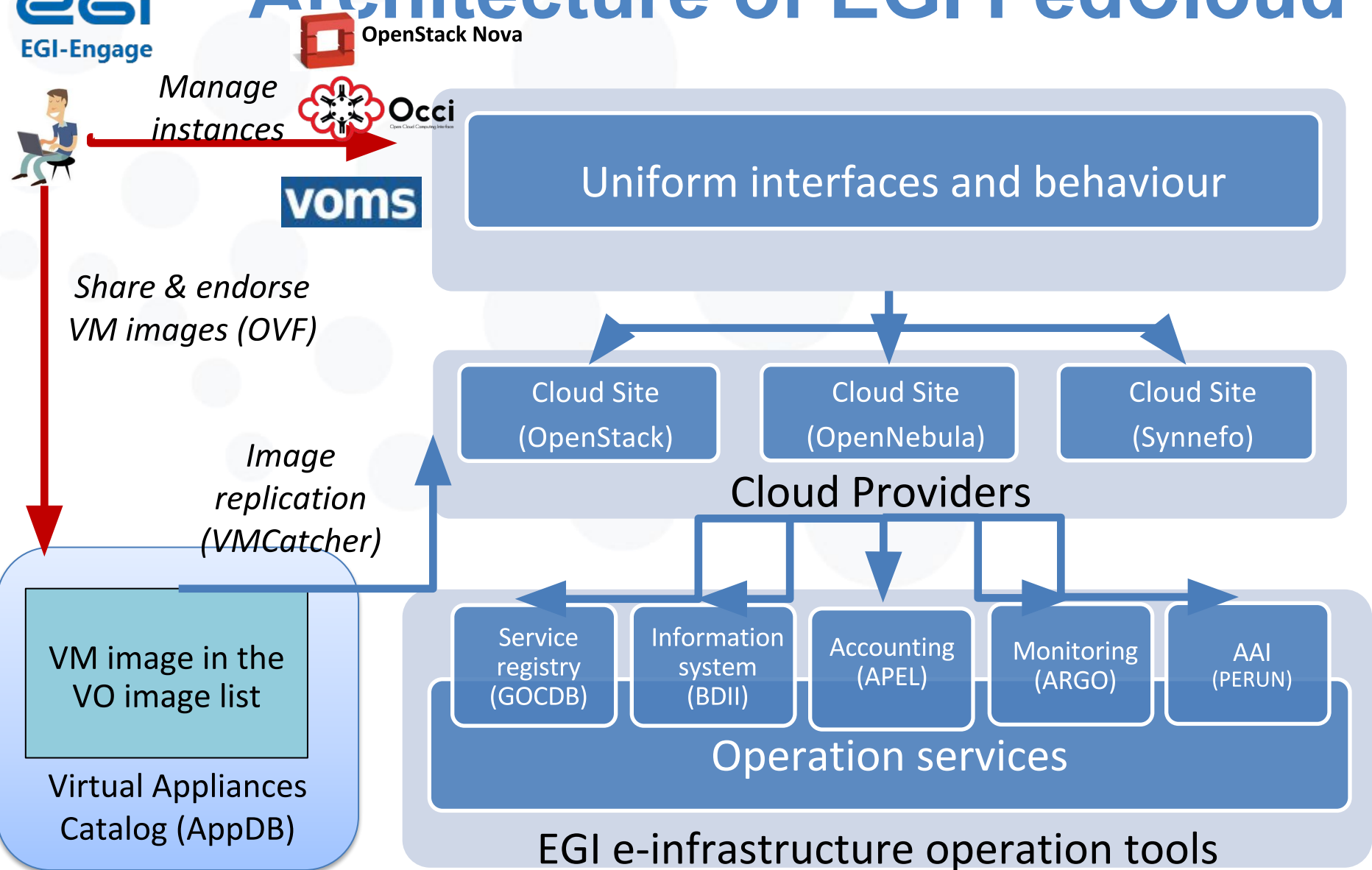
EGI Federated Cloud

EGI Federated Cloud is a collaboration of communities developing, innovating, operating and using cloud federations for research and education.
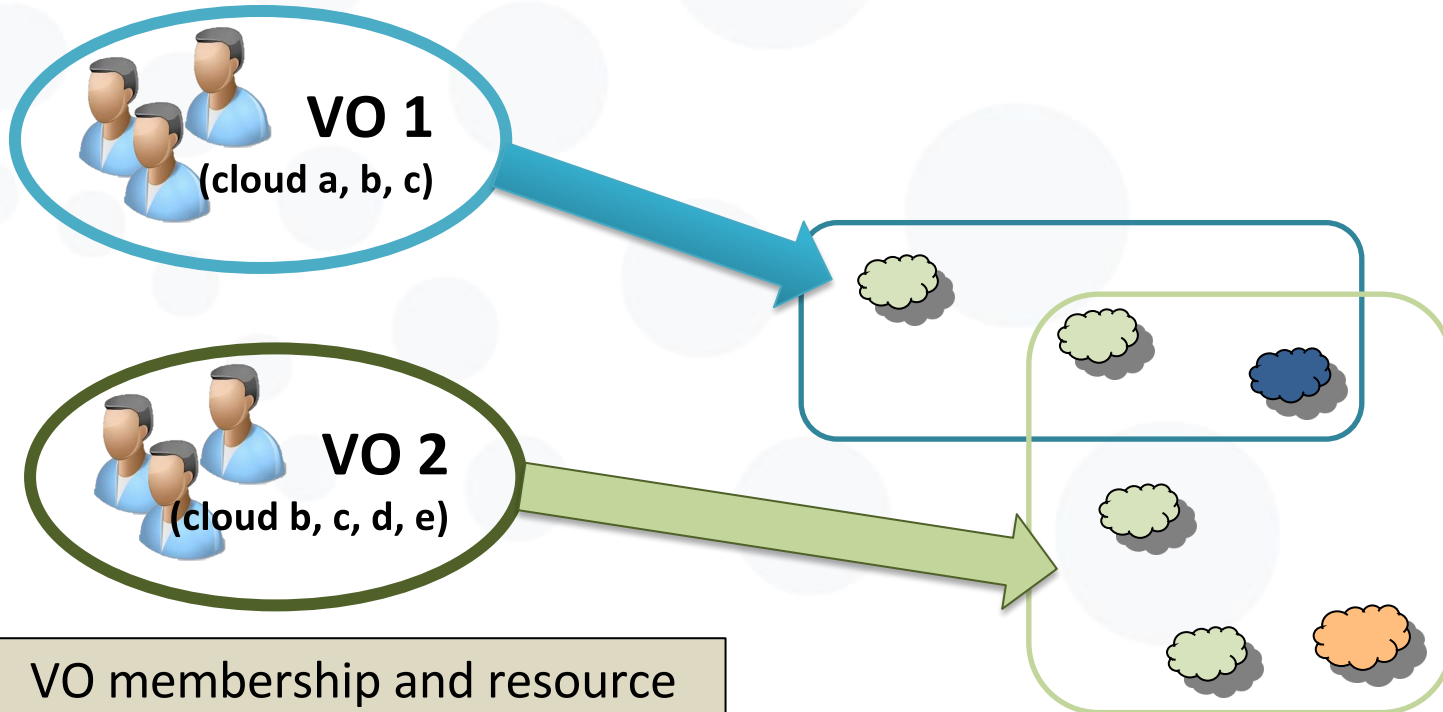
Today:
- 22 providers from 14 NGIs
  - 15 OpenStack
  - 6 OpenNebula
  - 1 Synnefo
- ~ 6.000 cores in total

Architecture of EGI FedCloud

OpenStack Nova

Occi Open Cloud Computing Interface

voms

Manage instances

Uniform interfaces and behaviour

Share & endorse VM images (OVF)

| Cloud Site (OpenStack) | Cloud Site (OpenNebula) | Cloud Site (Synnefo) |

Cloud Providers

Image replication (VMCatcher)

VM image in the VO image list

Virtual Appliances Catalog (AppDB)

| Service registry (GOCDB) | Information system (BDII) | Accounting (APEL) | Monitoring (ARGO) | AAI (PERUN) |

Operation services

EGI e-infrastructure operation tools

VO 1
(cloud a, b, c)

VO 2
(cloud b, c, d, e)

VO membership and resource access with X.509 certificates

1. Generic VOs → Incubator for new users
2. Sector-specific VOs → Dedicated environments

- Who use EGI, how they can apply and how the SLA looks like

IT4RIs workshop @RTSS 2016, 29 Nov - 02 Dec, Porto

48,000 users

+4,000 research papers

Who Use EGI?

Federated services

BBMRI

DARIAH

EMSO (unfunded, MoU)

EISCAT-3D

Environment (disaster mitigation)
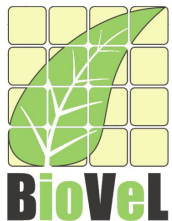
ELIXIR

MoBRAIN/ INSTRUCT

LifeWatch

EPOS

SLA negotiation

- EGI Foundation is acting as negotiator for communities
  - Finds and arranges resources that 'fit for purpose'
- Process is facilitated by
  - E-GRANT, Council surveys, broadcasting tool for sites
- SLA agreed with
  - **BILS, MoBrain, DRIHM, EXTraS, DARIAH-CC, Terradue (ESA TEPs), LSGC and Peachnote, Bioisi, EMSO_DEV**
- Finalising requests of D4Science …

Customer

SLAs

Negotiator

OLAs

Provider A

Provider B

Request submission	Matchmaking	Propose & Sign OLAs	Creating SLA	Propose & Sign SLA

FedCloud Typical usage models

- **Compute and data intensive workloads**
  - Batch and interactive (e.g. iPython-Jupyter) with scalable and customized environments
  - Cloud containers (e.g. Docker supported)

- **Service Hosting**
  - Long-running services (e.g. web server, database, application server)

- **Datasets repository**
  - Store and manage large datasets (in a storage volume)

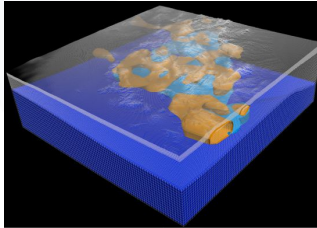- **Disposable and testing environments**
  - Host training environments, test applications

Support of RealTime Applications

- **Very specific requirements:**
  - Bandwidth
  - Response time:
    - Number of VM running after 5 minutes
    - Minimal number of VM running per hour
- **EGI SLA based on Availability / Reliability**
  - Number of VM (VCPU, RAM)
  - Total storage
- **EGI SLA's services are not based on response time**
  - If required by a community can be support at site level but EGI can't guarantee a QoS
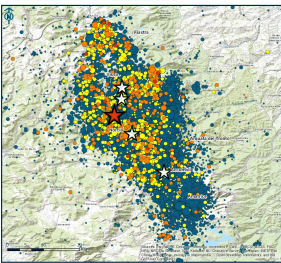   - Requires specific monitoring

- *Use Cases of Real Time Applications*

Near–real time earthquake simulations

mesh & seismic wavespeed
model selection

manual movie creation
with Paraview scripts

Istituto Nazionale di
Geofisica e Vulcanologia

Simulation with
Local INGV cluster
manual input setup

Earthquake
occurrence

0   ~1h   ~1.5h   ~3h   ~6h   You Tube

~4h

Simulation with

VERGE

automatic synthetic
seismograms & check
with observed data

manual input setup

manually reviewed
moment tensor seismic
source solution

automatic wavefield
propagation

INGV CENTRONAZIONALETERREMOTI

http://cnt.rm.ingv.it

Collection and Analysis of Ocean Big Data

emso DEV

European Multidisciplinary Seafloor and water column Observatory Development

**European Partners**

11
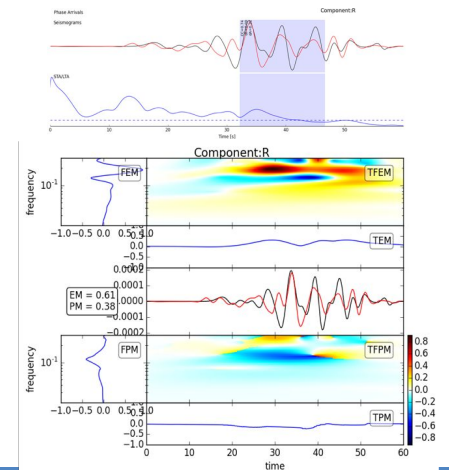
- **Building the EMSODEV Data Management Platform (DMP) using EGI Federated Cloud**
  - **Push Transfer Flow**: data is sent to a DMP service which "listens" to near-real time updates on XML files describing sensors data and measurements
  - **Pull Transfer Flow**: data is retrieved via API exposed by an OGC SOS server available at the OBSEA observatory located in Vilanova and managed by Universitat Politecnica De Catalunya.
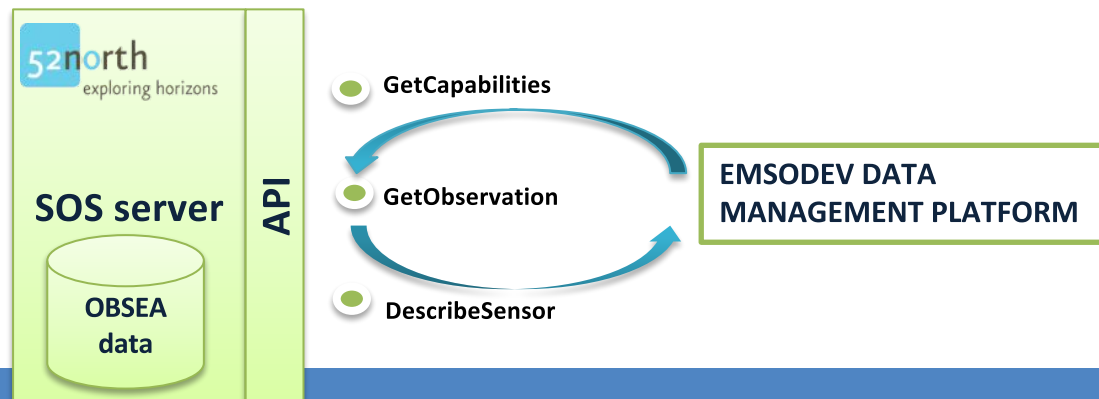
EMSODEV- Data Management Platform

DATA MANAGEMENT PLATFORM
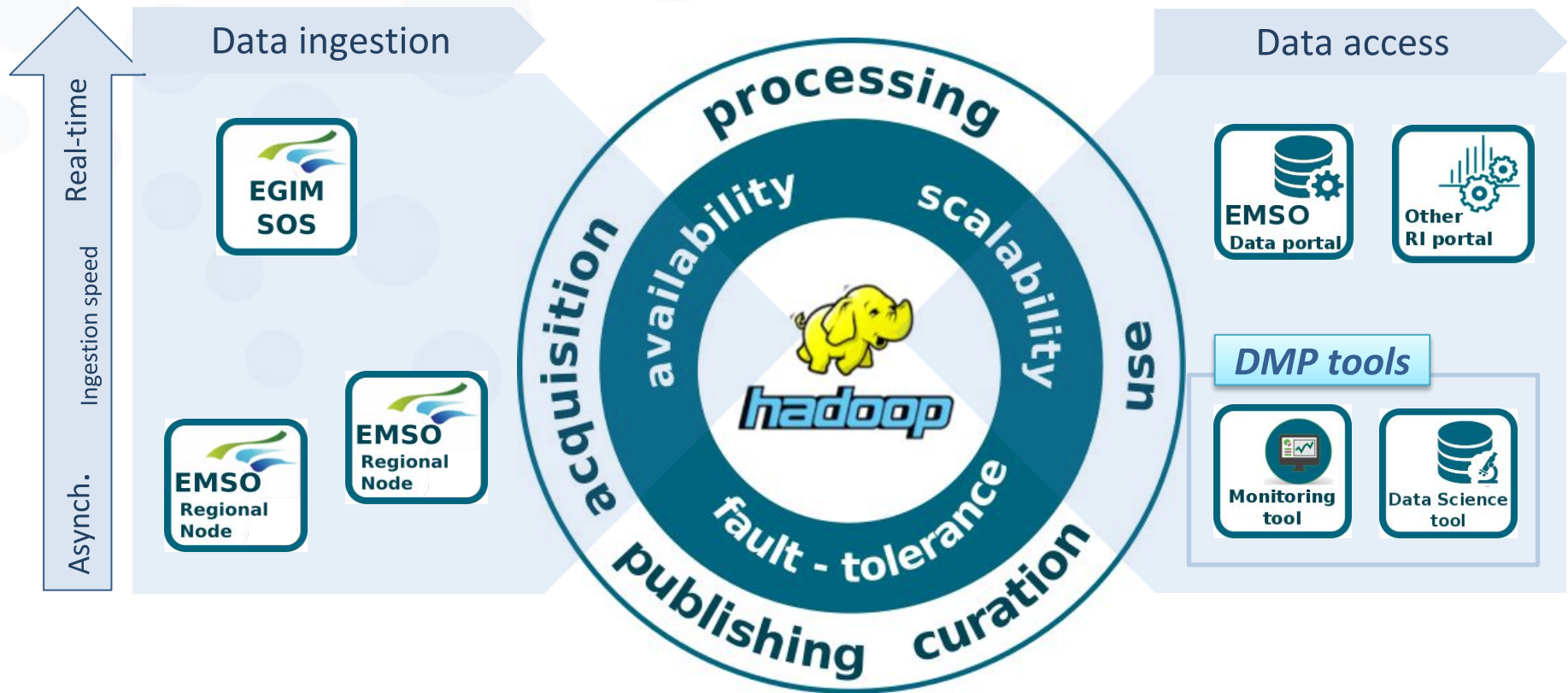
Production VO: vo.emsodev.eu
Cloud Compute: ~10 VMs (8 CPUs + 16GB RAM + 40GB HD)
File Storage: 5 TB

- *Benefits of Using FedCloud*

Benefits of Clouds

- **More** (infinite) computing power

- **On demand** at your fingertips

- **Scaling resources dynamically** according to demand

- **Choice,** with multiple providers

- **Resulting in:** better, faster

EGI added value

- A **ready-to-use IaaS** where to deploy on-demand IT services

- Easy VM and **security management**

- **Scalable** according to community needs (within the boundaries established through **SLA**)

- Secure VM access via a mechanism (VOMS credentials) based on proxy credentials issued and **verified by EGI**

- Fast and reliable **support** (ggus.eu trouble-ticketing and by mail)

But: Cloud Challenges

- **It's new !**

- **Security** / your **intellectual property**

- Traditional software **licensing**

- **Data transfer**

- Cloud **expertise** is needed

- Losing **control** over your jobs and data

- First **access** to clouds is often cumbersome

- *future direction and roadmap*

Software Containers remove or reduce cloud challenges

CAE Cloud Challenges	Containers
Security	✓
Portability	✓
Compliance	✓
Data Transfer	✓
Standardization	✓
Software licenses	✓
Resource Availability	✓
Transparency of Market	✓
Cost & ROI transparency	✓
No Cloud expertise Needed	✓

Beta phase

Thank you for your attention

Questions?