IBERGRID

INFRASTRUCTURE STATUS

Jorge Gomes
on behalf of the IBERGRID collaboration

ES-NGI

INCD

What it is

a) Federated infrastructure joining Iberian (PT+ES) research and academic organizations mainly focused:
   - Cloud Computing
   - Grid Computing
   - Data Processing

b) Enables a joint participation in international research e-infrastructures and initiatives for the benefit of Iberian researchers
   - Supports the Iberian participation in EGI, WLCG, EOSC and ESFRIs
   - Provides regional services, operations coordination and integration

c) Forum for common activities and sharing of knowledge.
   - Participation in EU projects both research and infrastructure oriented
   - IBERGRID conference series since 2007

Governance

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

REPÚBLICA PORTUGUESA
Ministério da Ciência, Tecnologia e Ensino Superior

CSIC
Consejo Superior de Investigaciones Cien

fct Fundação para a Ciência e a Tecnologia

ES-NGI — Infrastructure

IFCA Instituto de Física de Cantabria — Tech coordination

INCD — Infrastructure

LIP — Tech coordination

IBERGRID

Iberian summit of Valladolid in Nov 2018

2. Na área de Computação Distribuída, os Signatários pretendem promover a colaboração nas seguintes áreas:

a) Apoiar a realização anual da conferência IBERGRID, realizada alternadamente no território de ambos os Estados dos Signatários;

b) Reforçar a colaboração entre os Signatários, com vista a otimizar e apoiar a participação da infraestrutura Ibérica IBERGRID em infraestruturas e iniciativas internacionais de computação distribuída e repositórios de dados, entre as quais se destaca o *European Open Science Cloud* (EOSC) e o *European Grid Infrastructure* (EGI);

c) Apoiar o desenvolvimento e integração de serviços temáticos de interesse para a comunidade científica a disponibilizar através da infraestrutura Ibérica IBERGRID;

d) Apoiar e fomentar a utilização da infraestrutura Ibérica IBERGRID no apoio à participação em projetos científicos estratégicos de interesse comum tais como a participação no CERN, o suporte aos ESFRIs e o AIR Center.

Assinado em Valladolid, no dia 21 de novembro de 2018, em dois originais, nas línguas portuguesa e castelhana, sendo ambos os textos igualmente válidos.

Pelo Governo
da República Portuguesa

O Primeiro-Ministro

Antonio Costa

Pelo Governo
do Reino de Espanha

O Presidente do Governo

Pedro Sánchez Pérez-Castejón

Adenda **2023** ao
PROTOCOLO DE ADESÃO DO CENTRO OPERACIONAL
INCD À REDE NACIONAL DE COMPUTAÇÃO AVANÇADA

CLÁUSULA PRIMEIRA

Recursos computacionais

1. A INCD obriga-se a disponibilizar os seguintes recursos computacionais, para utilização pelos sistemas de I&D&I nacionais:

   …

   d) Recursos computacionais que satisfaçam a participação nacional nas infraestruturas internacionais **EGI** e **IBERGRID**, bem como a participação na iniciativa EOSC;

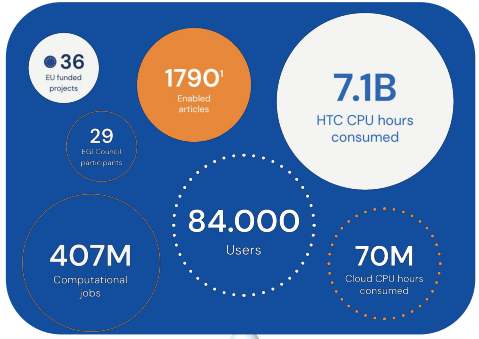
2. A INCD obriga-se ainda a:

   …

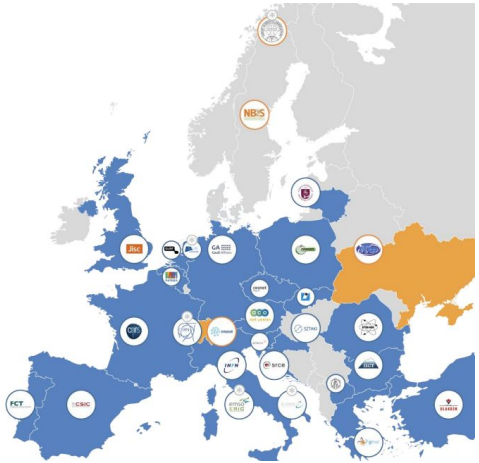
   b) Co-organizar a Conferência **IBERGRID**;

IBERGRID and EGI

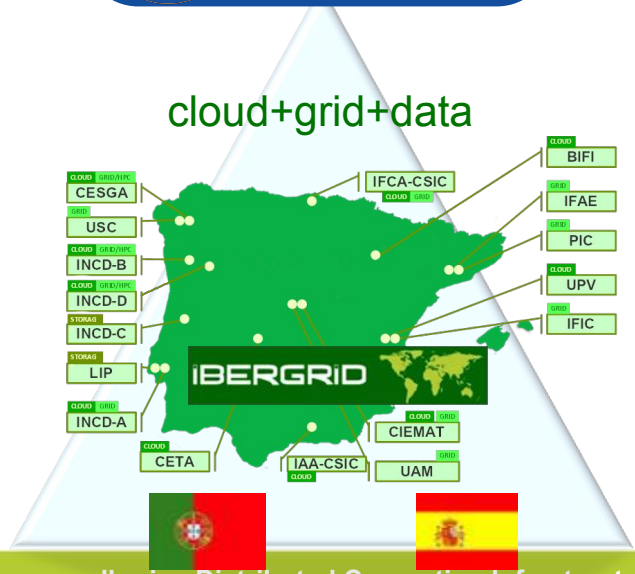
IBERGRID is a regional infrastructure within EGI since 2010 ⇒ NGI_IBERGRID



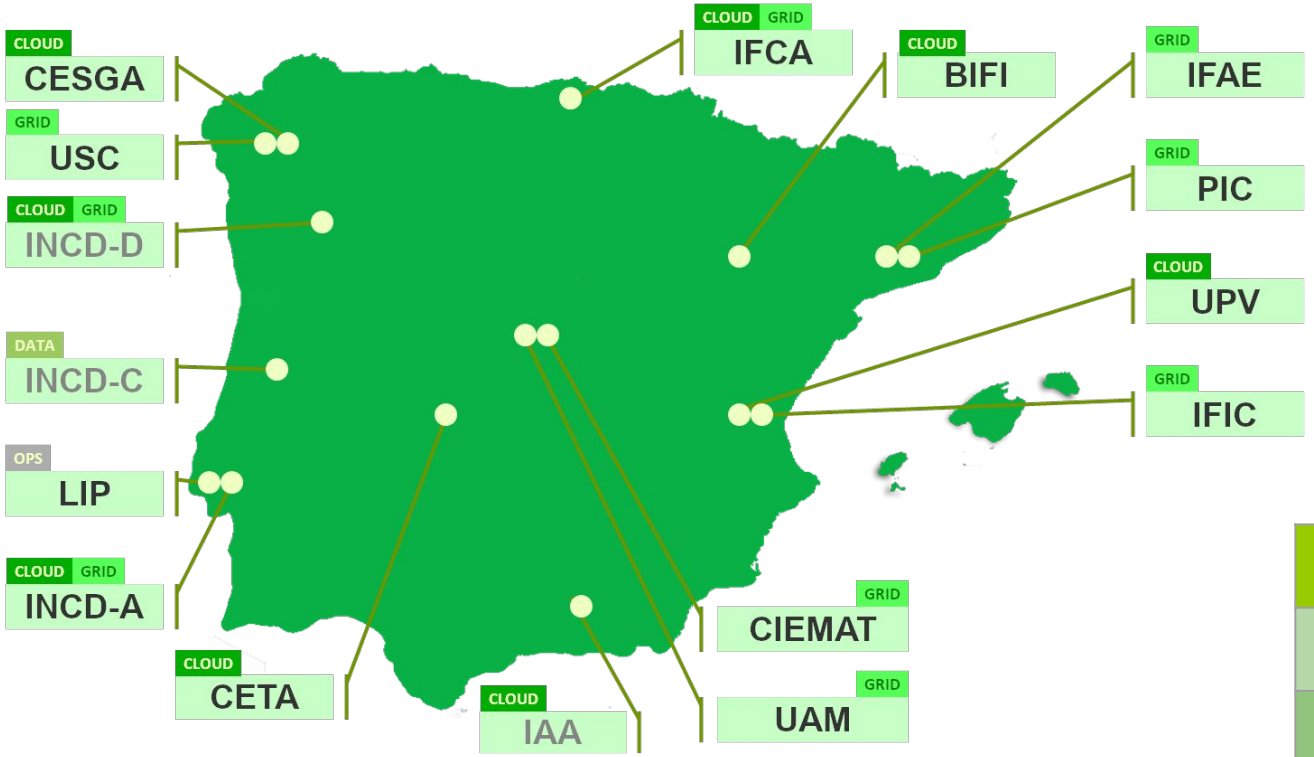
- Resource provider
  - cloud
  - grid
- Technology provider
  - software
  - services
- Community support
  - enabling usage
  - thematic services



22 countries



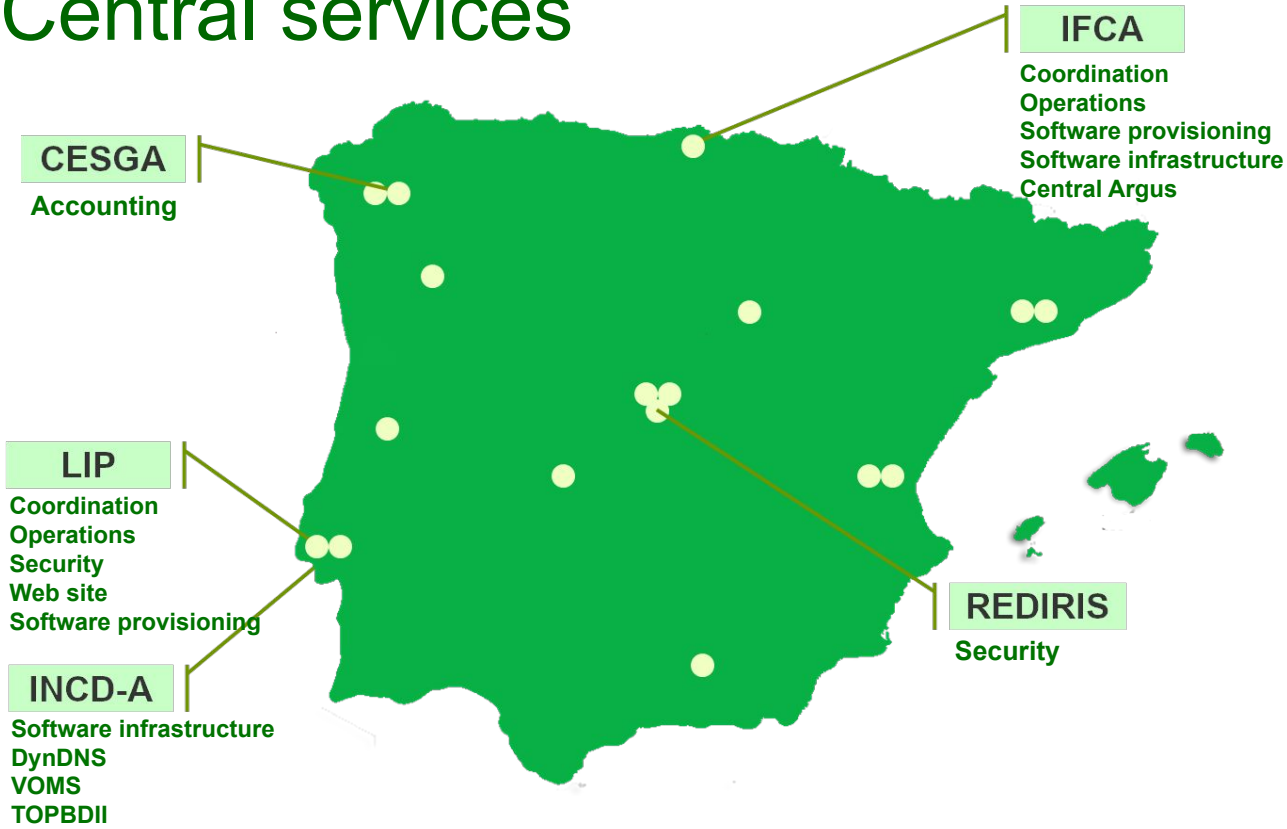
Infrastructure sites (cloud + grid)



EGI CERTIFIED GOCDB
NCG-INGRID-PT
CESGA
IFCA-LCG2
UAM-LCG2
PIC
BIFI
IFAE
UPV-GRyCAP
USC-LCG2
IFIC-LCG2
CIEMAT-LCG2
CETA-GRID
CESGA-CLOUD

	Current	Planned
Portugal	2	+2
Spain	11	+1

Central services



IFCA
Coordination
Operations
Software provisioning
Software infrastructure
Central Argus

CESGA
Accounting

LIP
Coordination
Operations
Security
Web site
Software provisioning

INCD-A
Software infrastructure
DynDNS
VOMS
TOPBDII

REDIRIS
Security

Coordination through weekly oversight shifts and VC meetings

Coordination by LIP and IFCA with CESGA support

Resources changes

- **BIFI** in Zaragoza integrated cloud capacity with increased usage in 2022.
- **CESGA** dropping grid and focusing on cloud.
- **CIEMAT, IFIC, IFAE** and **PIC** increased grid capacity mainly for LHC.
- **IAA** in Granada aims to integrate cloud resources for SKA.

- **INCD-A** in Lisbon slight improvements in 2022 and 2023 both grid and cloud.
- **INCD-D** new site in Vila Real aims to integrate cloud resources in 2023/2024.
- **INCD-C** site in Coimbra aims to integrate a data storage service.

Virtual Organisations on-boarding

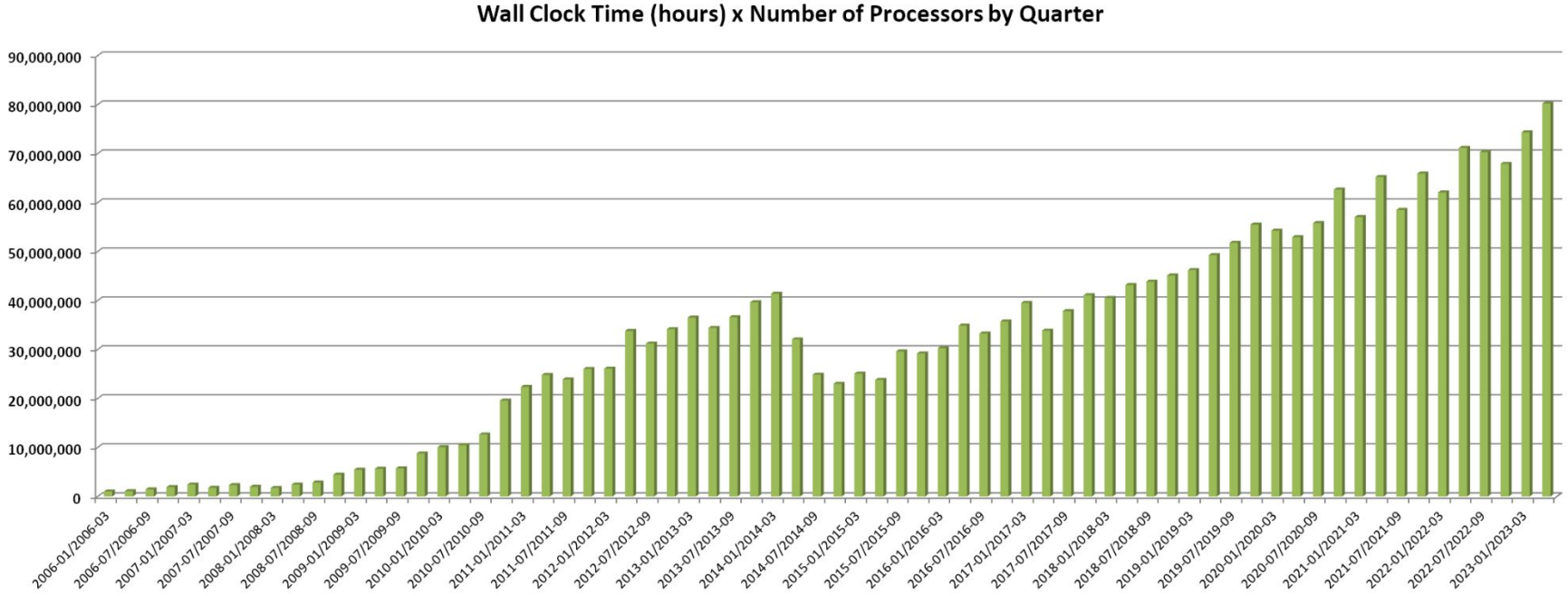
- IceCube (2022/IFCA) - neutrino observatory experiment in the south pole
- cesga.es (2022/CESGA) - makes available CESGA cloud resources
- gltfca (2023/NCG) - use case on earth observation from C-Scale project
- ispravision.egi.eu (2022/IFCA) - in situ photo recognition & augmented vision in EGI-ACE
- vo.ai4eosc.eu (2023/IFCA/NCG) - AI, ML and DL for EOSC
- vo.bd4nrg.eu (2022/BIFI) - big data management challenges in the energy sector
- vo.digitbrain.eu (2022/IFCA) - easy access to Digital Twins for manufacturing
- vo.imagine-ai.eu (2022IFCA/NCG) - imaging data and services for aquatic science
- vo.phiri.eu (2022/BIFI) - population Health Information Research Infrastructure.

Grid Computing

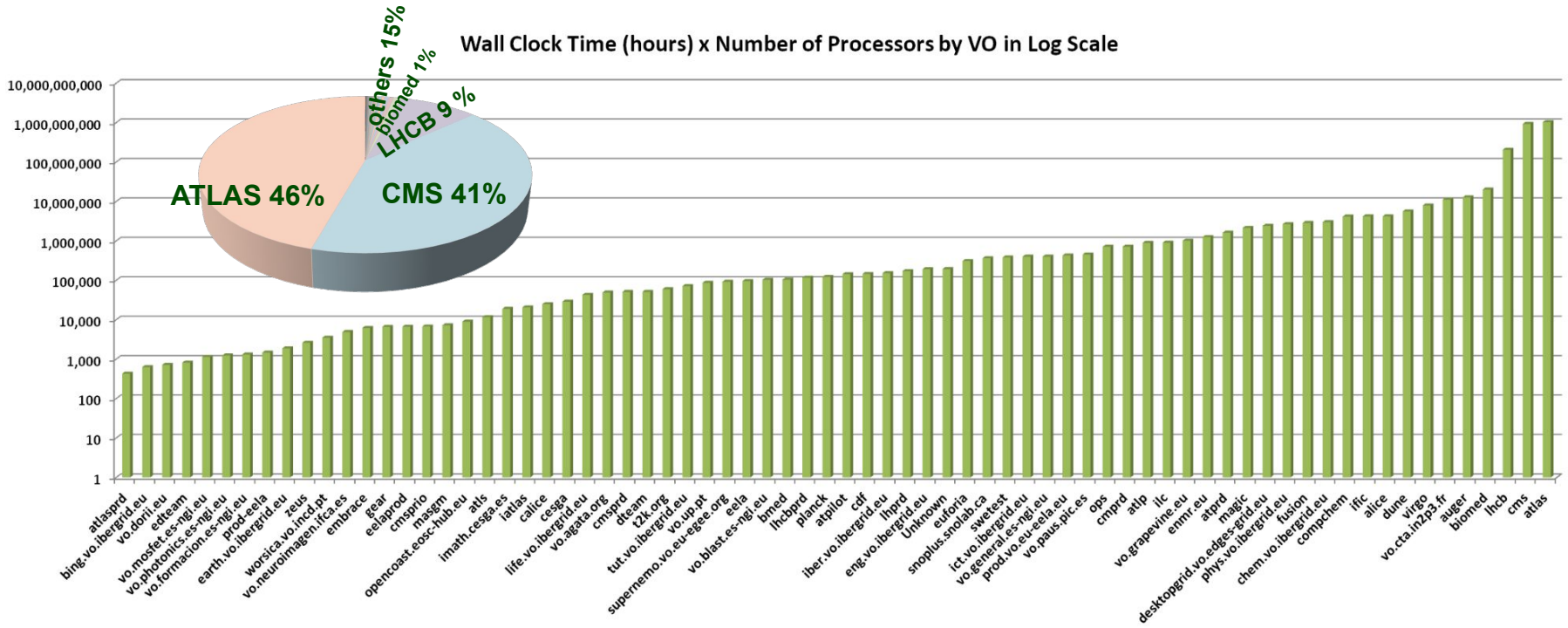
Last 12 months

- 8x grid sites
- 14x VOs
- 9,370,708 jobs
- 301,053,872 hours
- 3,217,805,562 HS23 normalised hours
- 82% CPU efficiency
- ~ 40,000 CPU cores

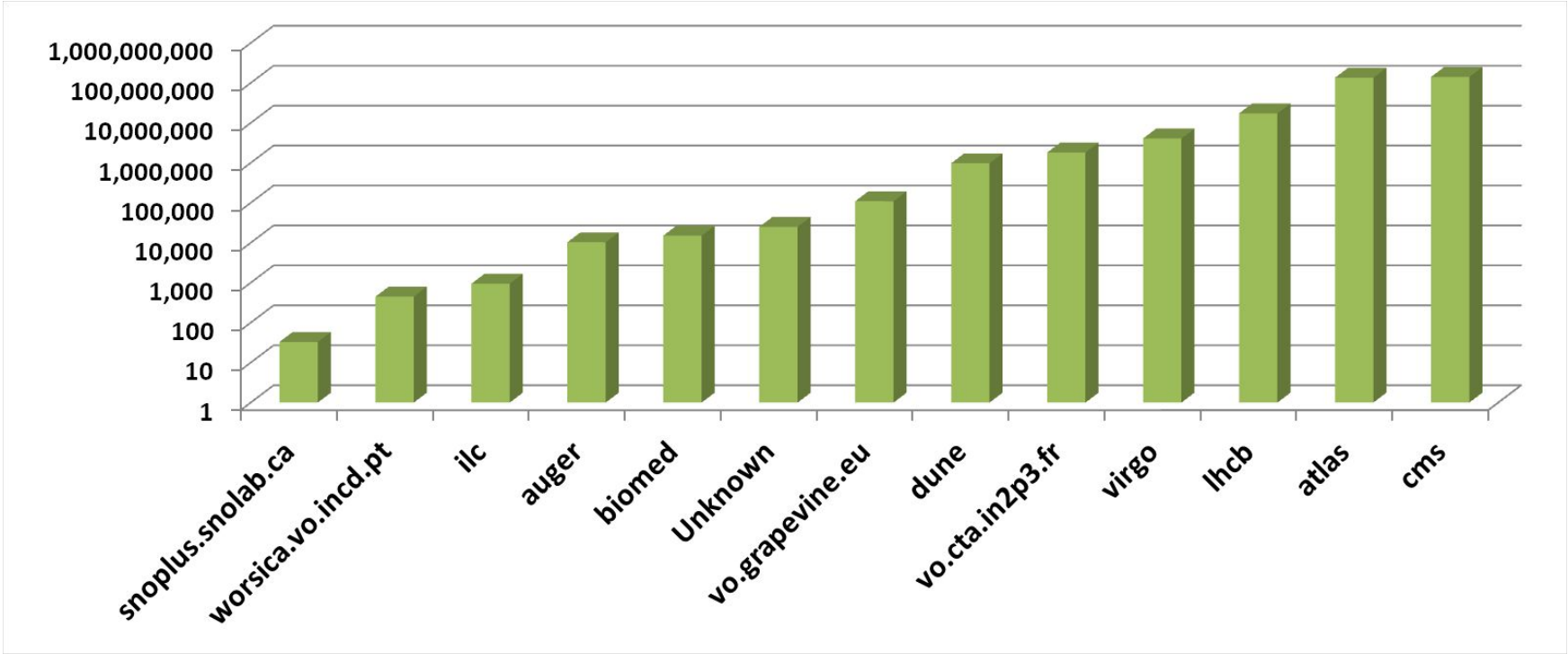
316 million grid jobs since January 2006
2,262 million processing wall clock hours



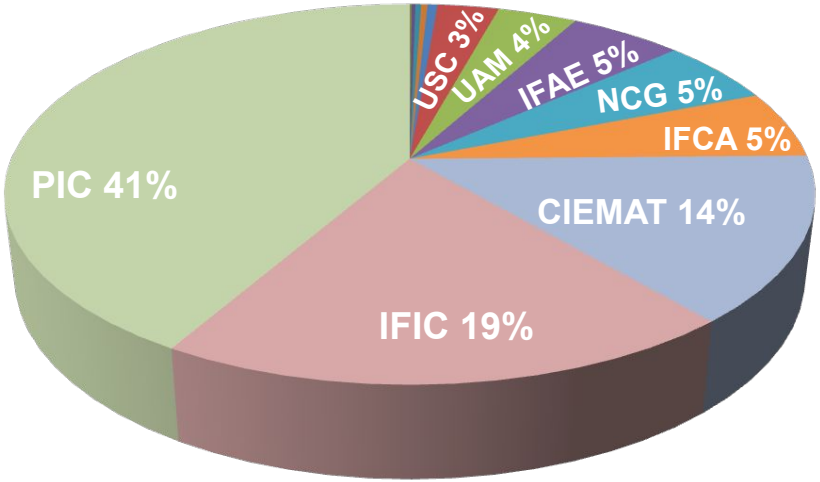
Grid Wall Clock hours by VO Jan 2006 to Aug 2023



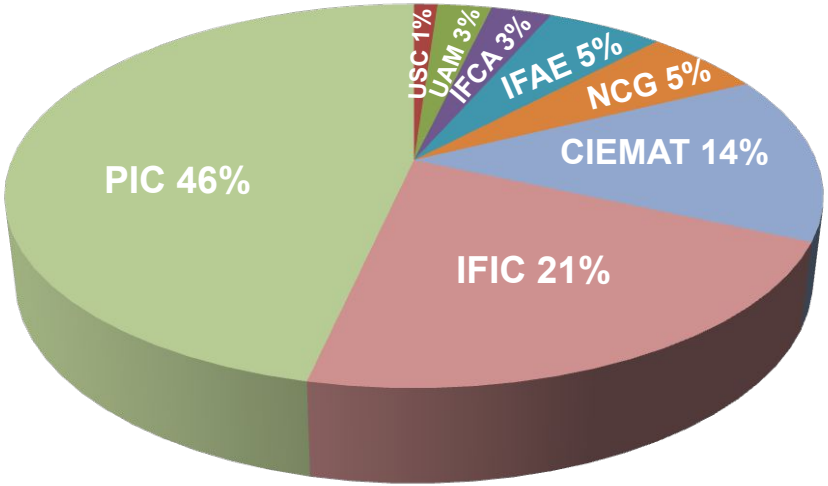
Grid Wall Clock hours by VO Sep 2022 to Aug 2023



Grid Normalised Hours by Center

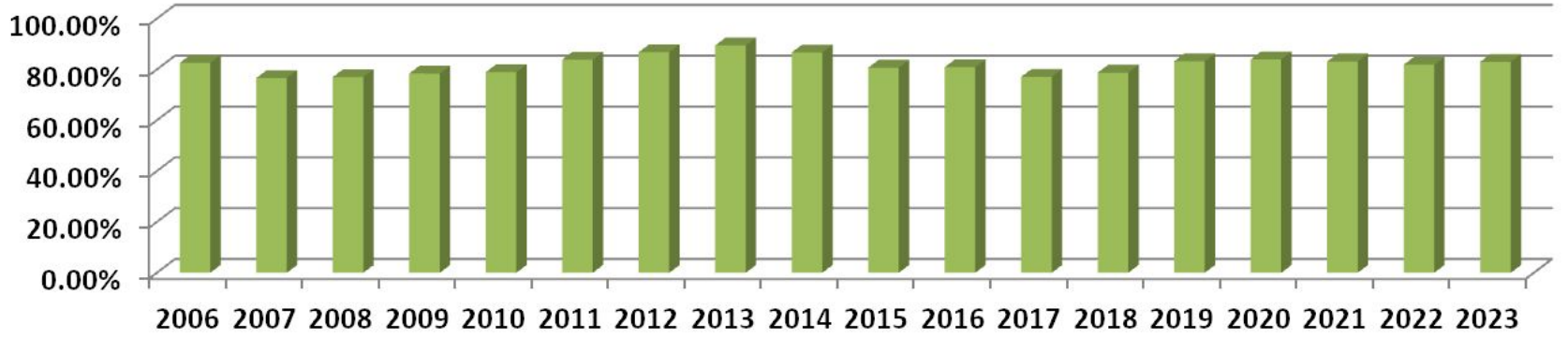


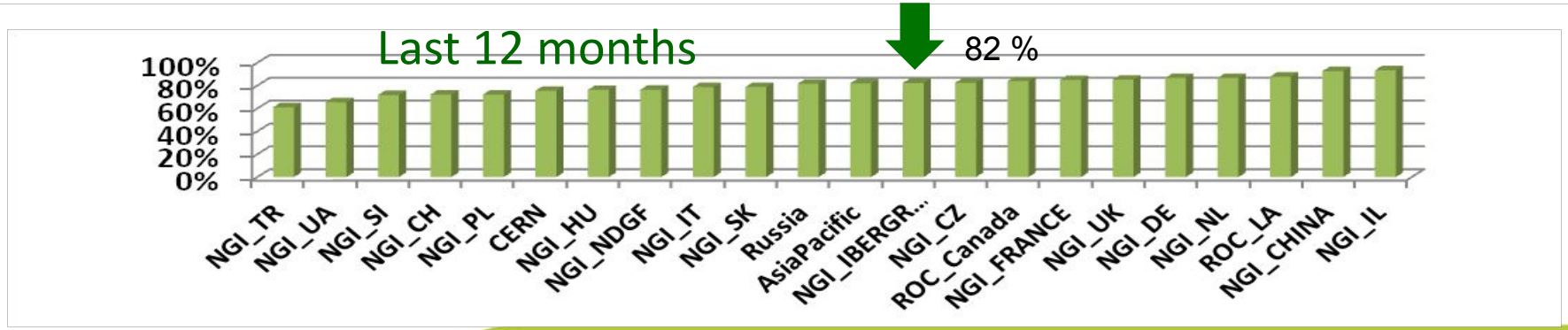
Jan 2006 - May 2023



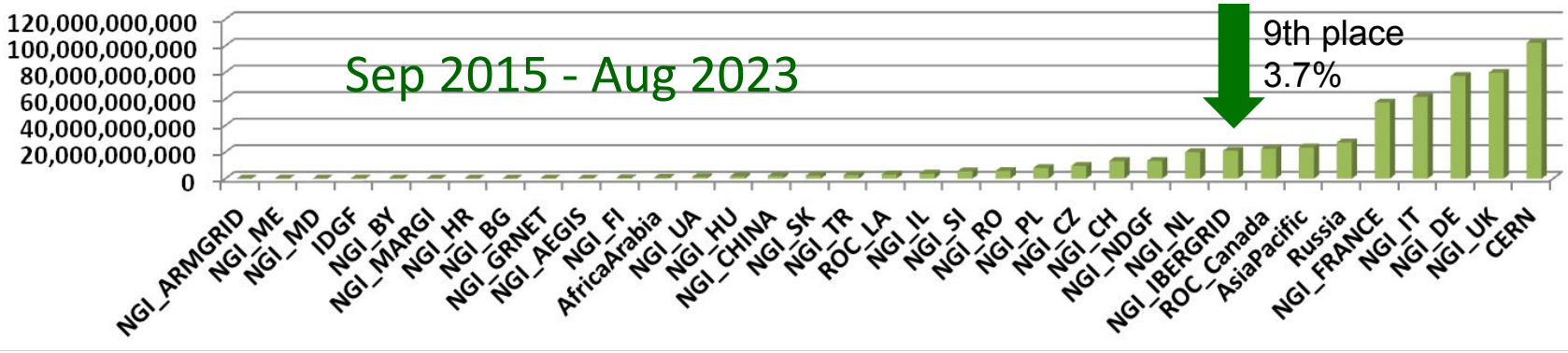
Sep 2022 - Aug 2023

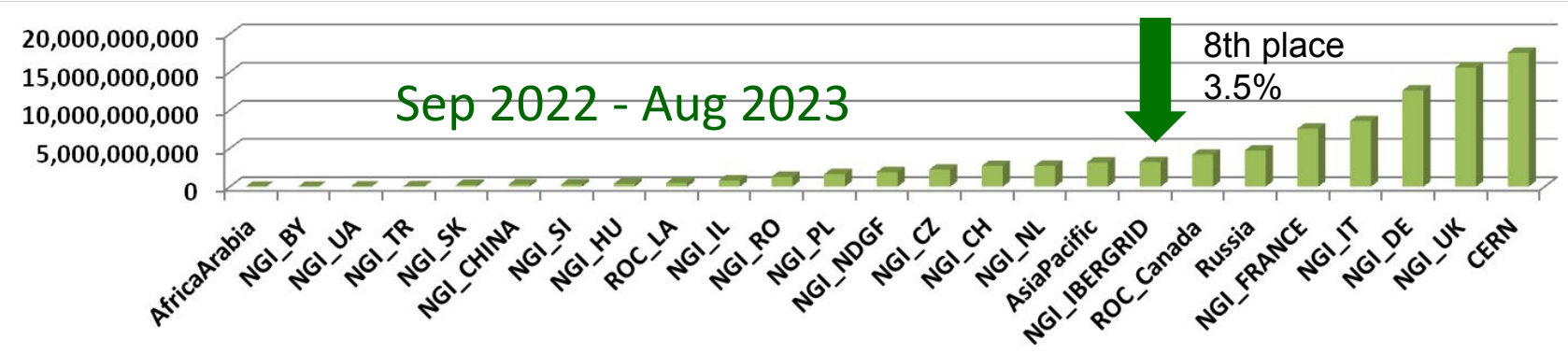
Grid CPU efficiency



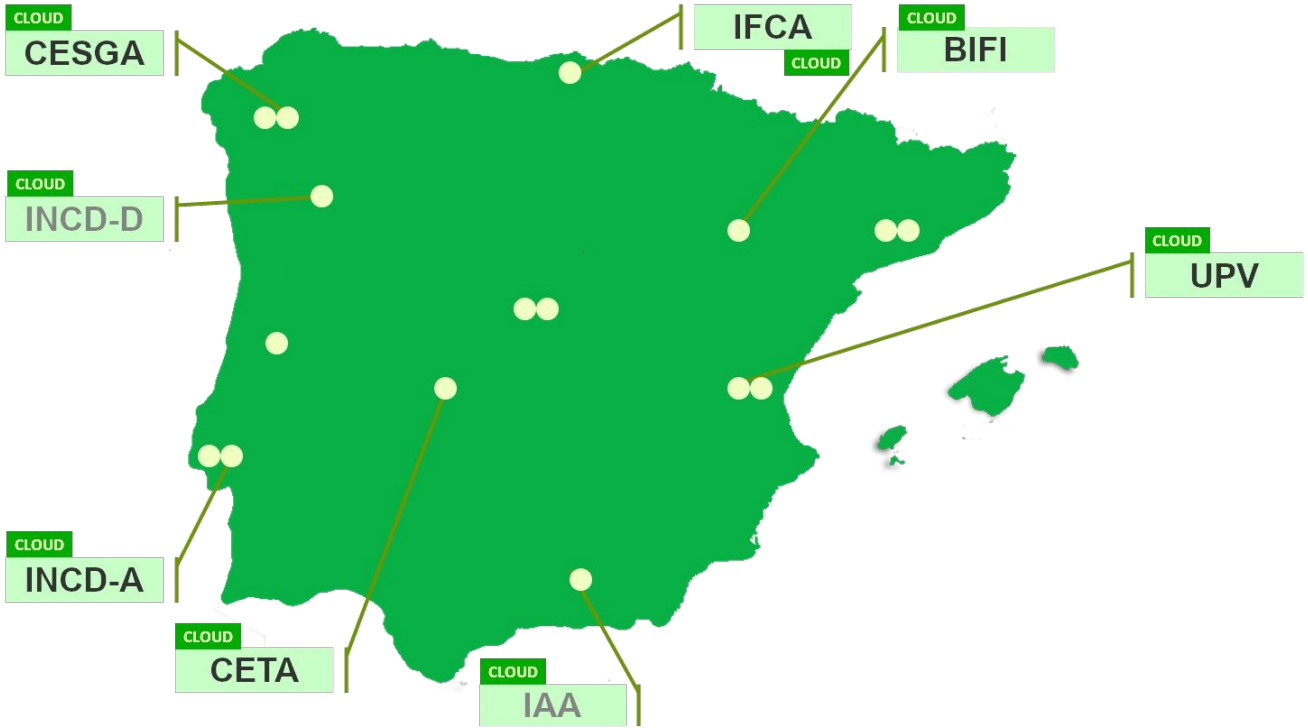


Grid Normalised Hours by NGI





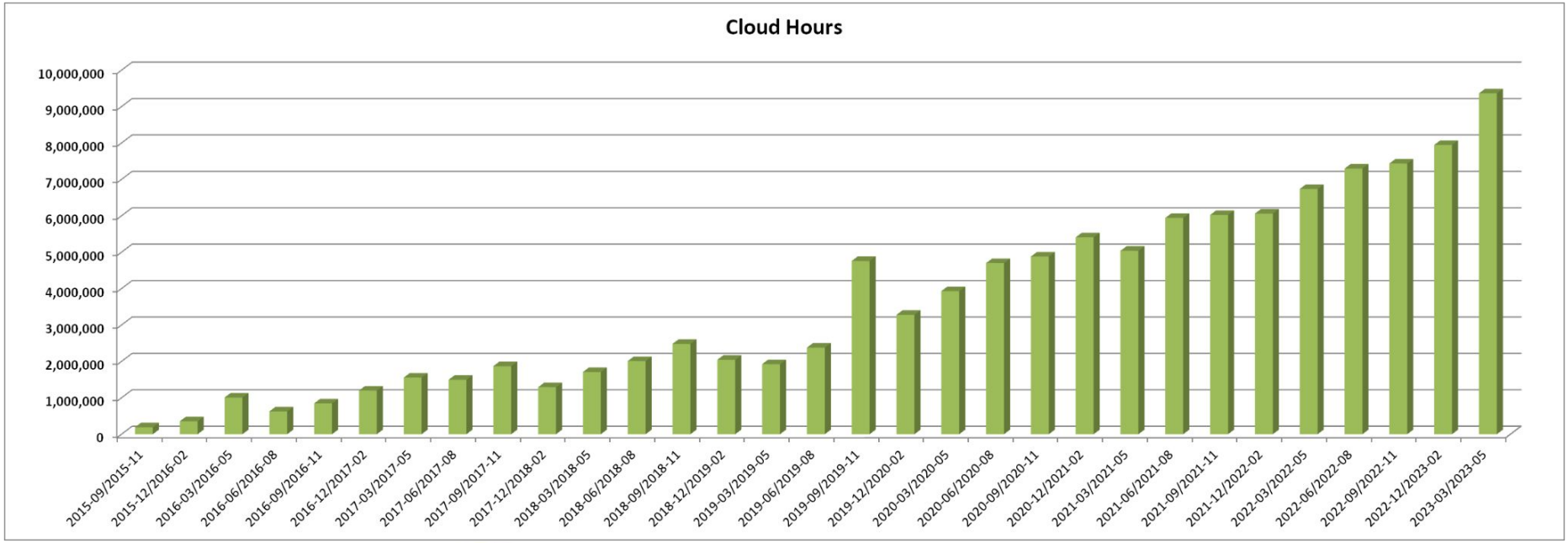
Cloud Computing



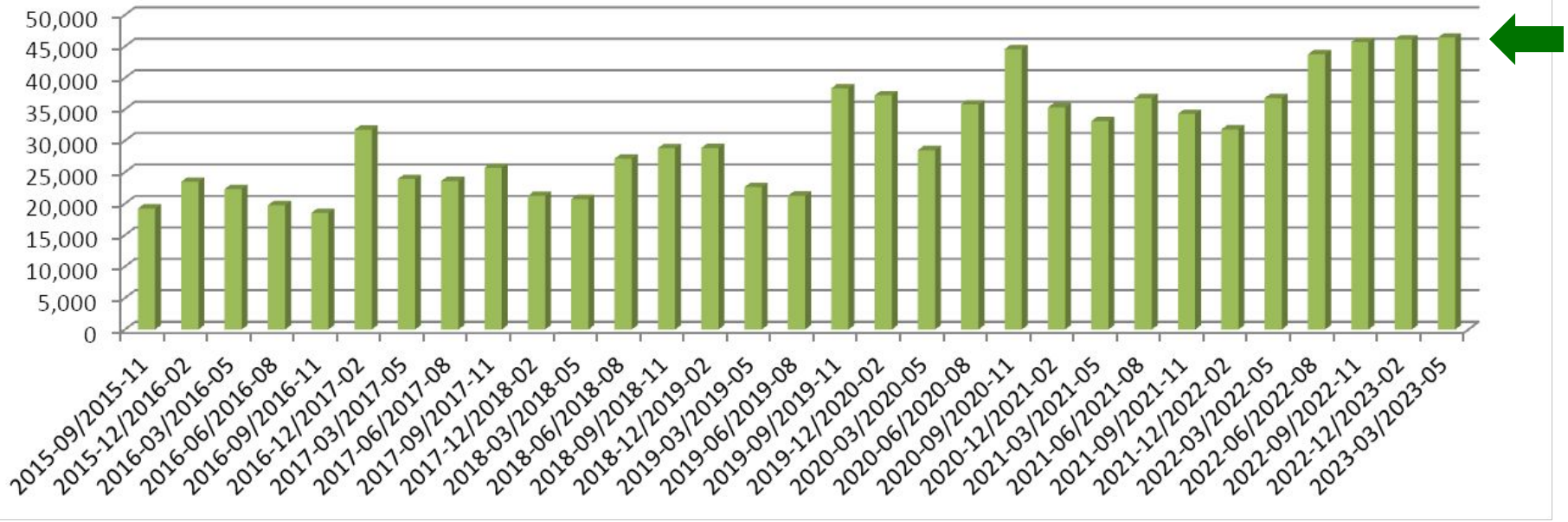
Last 12 months

- 6x cloud sites
- 30x VOs
- 170,029 VMs
- 30,844,817 hours

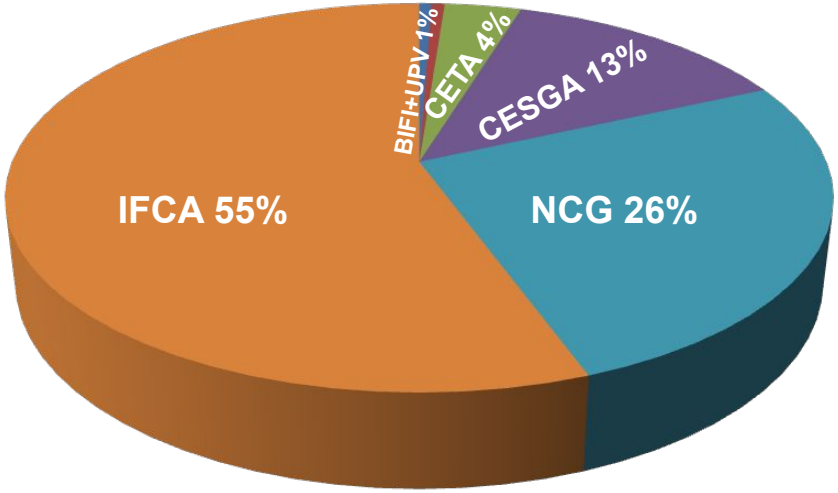
Cloud Wall Clock Hours
117 million hours since Sep 2015



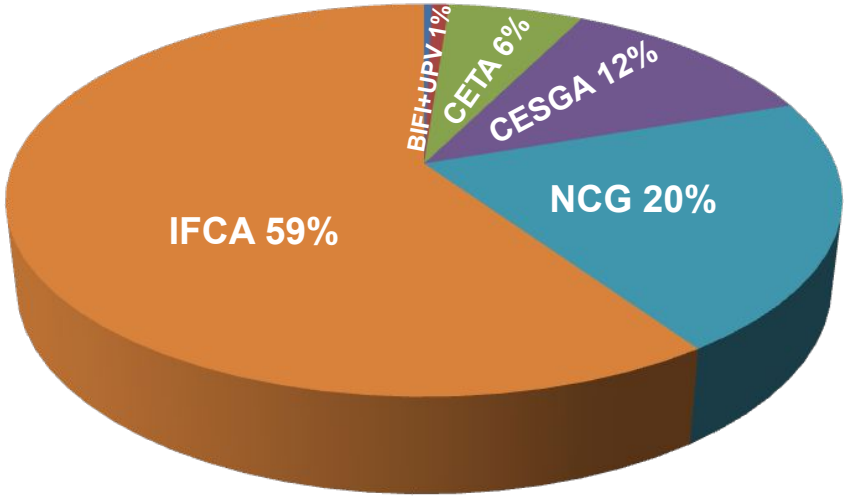
Instantiation of cloud VMs
985 thousand VMs since September 2015



Cloud Wall Clock Hours by Center

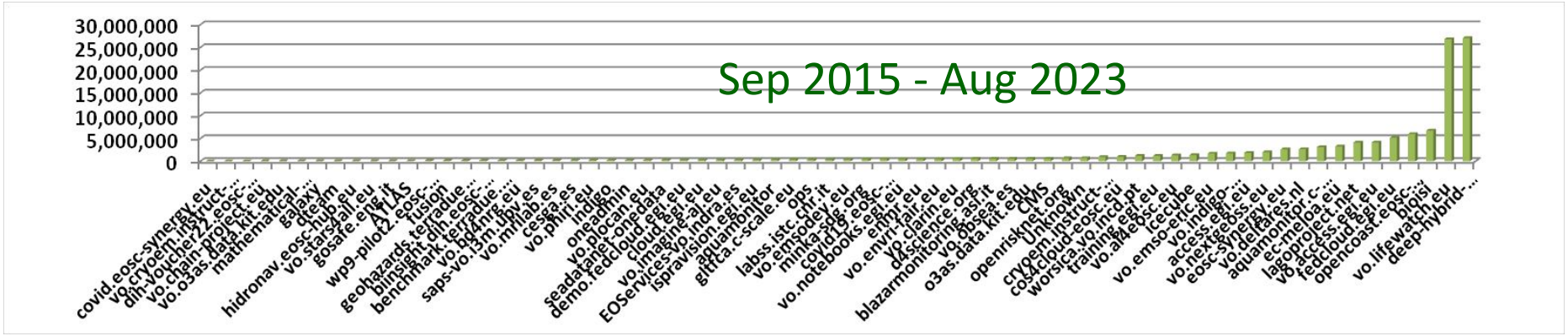


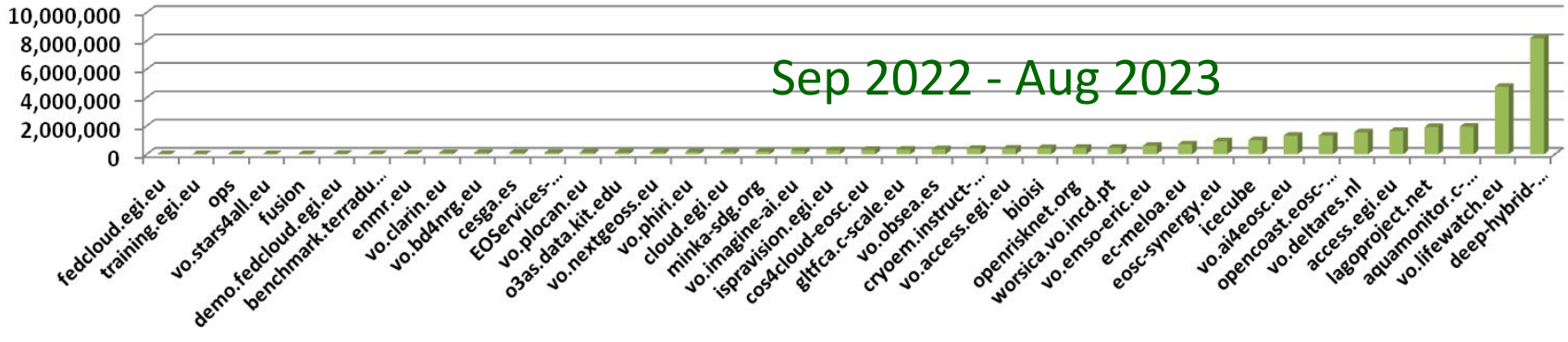
Sep 2015 - Aug 2023



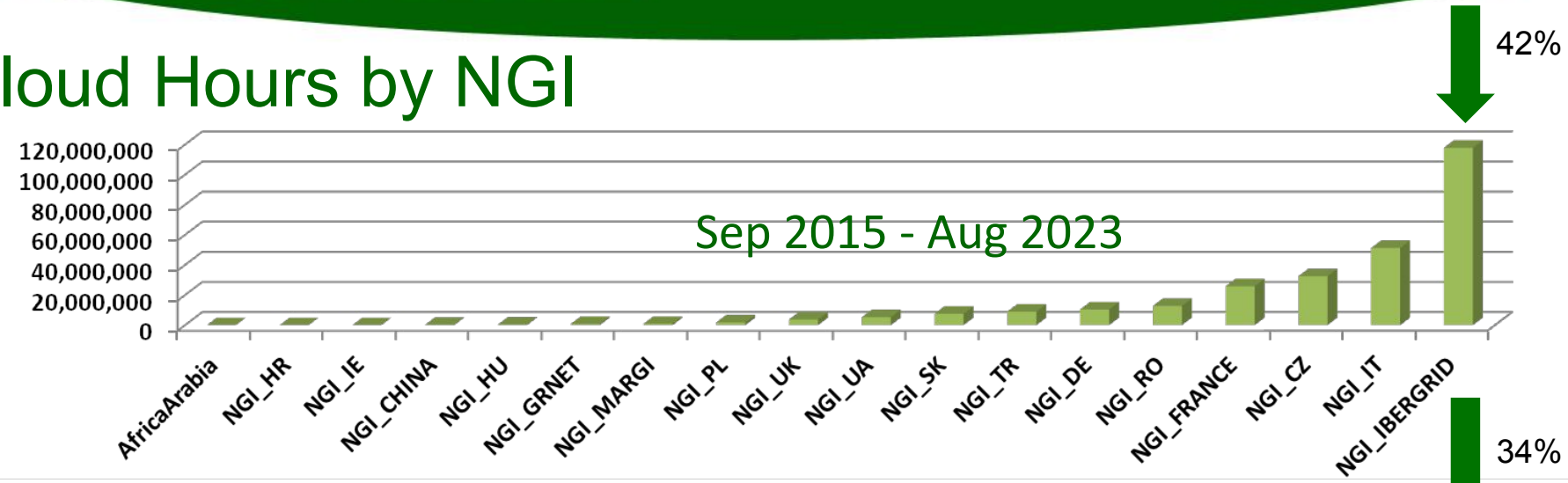
Sep 2022 - Aug 2023

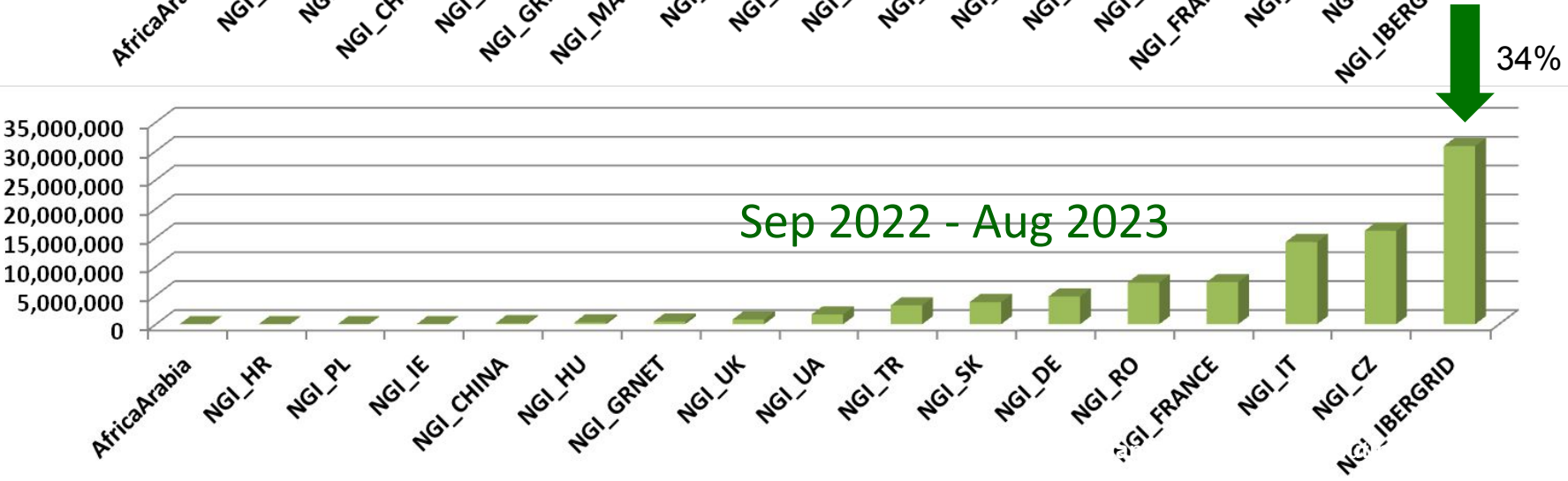
Cloud Hours by VO





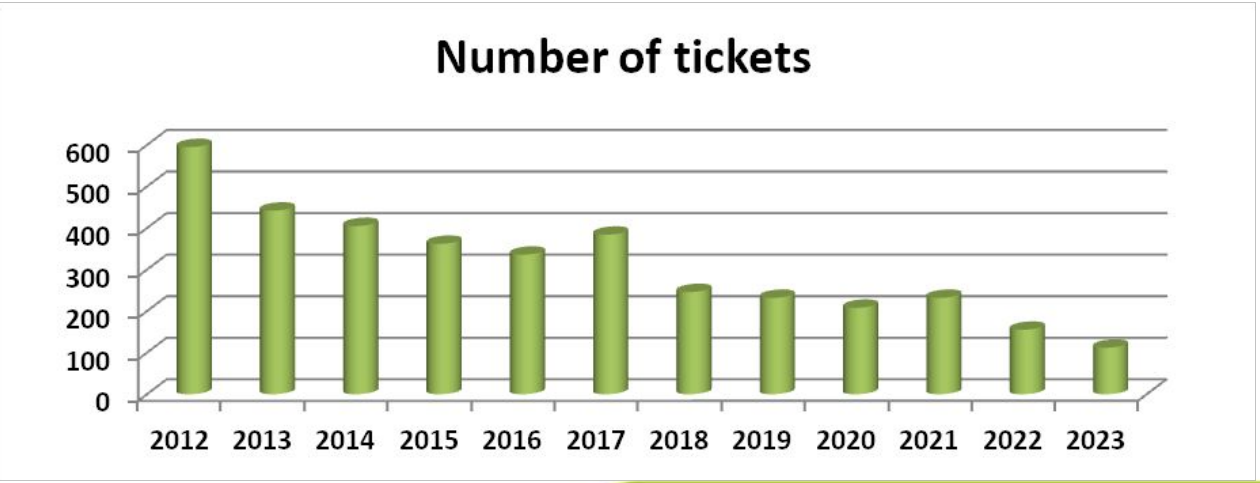
Cloud Hours by NGI

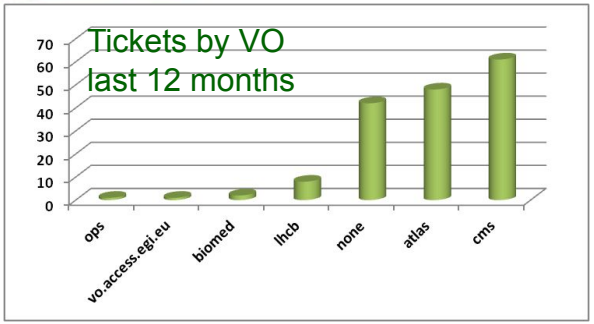


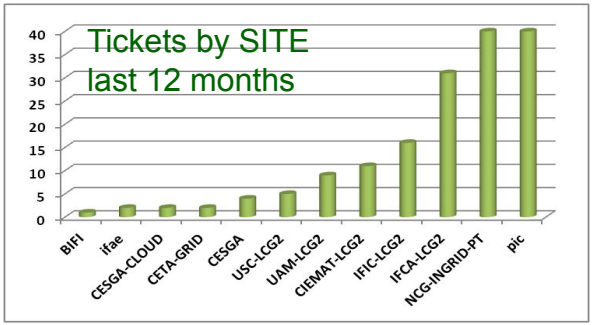


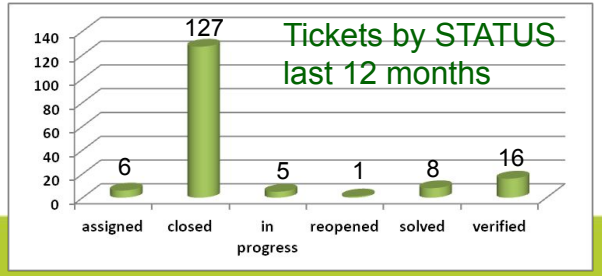
Infrastructure support

- Helpdesk: EGI GGUS, support unit: NGI_IBERGRID
- 3709 tickets since 2012
- From Sep 2022 to Aug 2023 (last 12 months)
  - IBERGRID 163 tickets, 12 in open states (7%)
  - Whole EGI 3686 tickets, 415 in open states (11%)









Infrastructure evolution

- HTC / Grid driven by WLCG
  - AAI from x.509 certificates to tokens
  - AAI moving away from VOMS / certificate to groups and attributes with OAuth 2.0
  - Dropping of legacy protocols like SRM an GridFTP
  - Adoption of WebDAV and XRootD as reference protocols for data access
  - ARC-CE and HTCondor-CE as main options to access computing resources

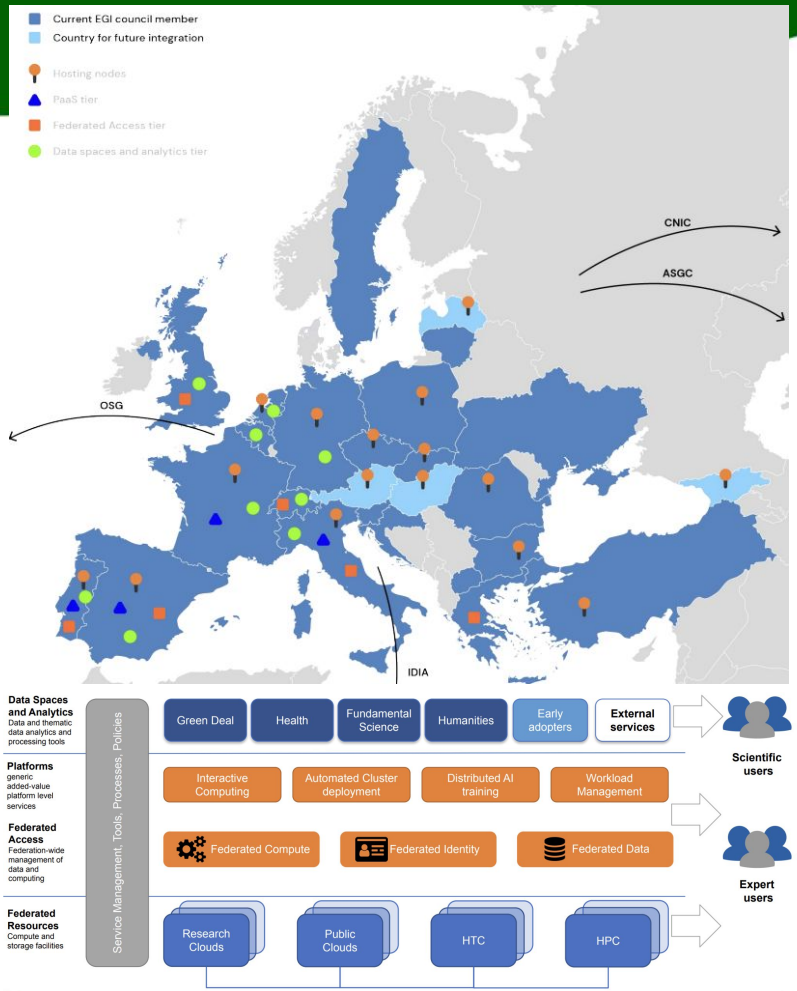
- Cloud
  - Federation replacing cloud infoproviders with native messaging …
  - Containers are everywhere and here to stay …
  - Towards serverless computing …

IBERGRID in EGI

- Regional infrastructure coordination
- Regional infrastructure support
- Accounting portal
- Orchestration services (IM)
- Software provisioning
- Software infrastructure
  - UMD repository
  - CMD repository
- Resource provisioning
  - Grid
  - Cloud
- Generic services and software:
  - IM, OSCAR, EC3
  - SQAaaS + JePL
  - udocker
  - DEEP platform
- Thematic services
  - LifeWatch / GBIF
  - Scipion
  - OPENCoastS
  - Worsica
  - aquamonitor
  - etc

EGI-ACE

- Middleware quality
  - Software provisioning
  - Software infrastructure
- Resources integration
  - HPC integration piloting
- Cloud resources provisioning
  - Thematic services
- Infrastructure services provisioning
  - IM, EC3, DEEP
- VA based provisioning of services



Software quality criteria

Good practices aimed at improving *research* software and services. Set of quality conventions oriented to DevOps.

- Software baseline
  - Criteria meant to enhance the visibility, accessibility and distribution of source code.
  - Encouraging good coding practices to improve quality, reliability and security.
- Services baseline
  - Minimum set of principles for reliable and fit-for-purpose services such as web services, web applications, platforms etc.
  - Provides common coherent quality attributes aimed to ensure functional suitability and strengthening of the services reliability and stability.

The baselines are openly developed on github.

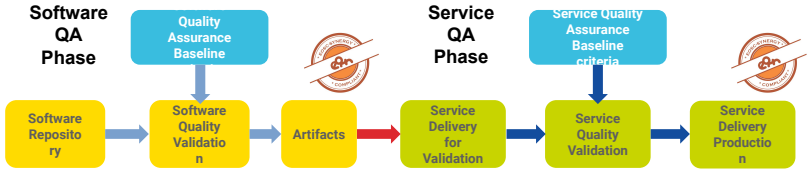
Software quality criteria

Infrastructure for Quality Research Software task force of EOSC

- Identify Quality Attributes that are appropriate for RS
- Review of publications relevant for identification of software quality attributes.

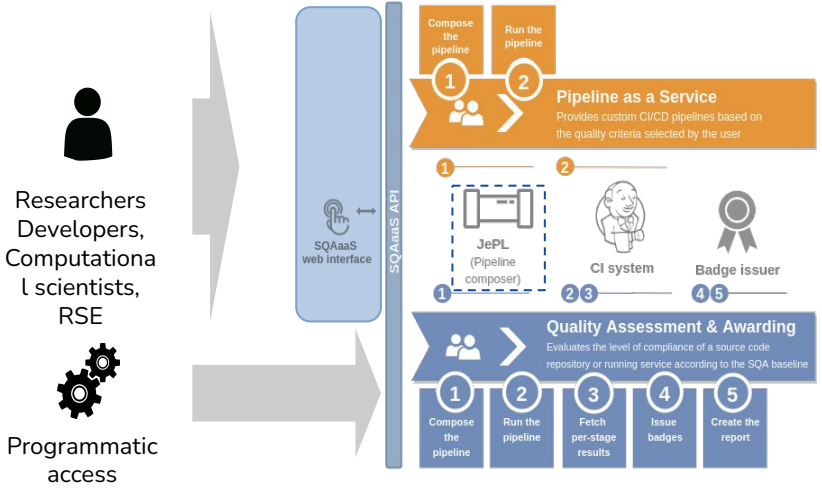
- 1st Deliverable of the Task Force Subgroup: Published in zenodo:
  - https://zenodo.org/record/8221384
- Contains: Quality Characteristics and Quality Attributes
  - From all identified Quality models
  - End up with 132 Quality Attributes
- Next step is produce recommendations of Quality Attributes for RS.

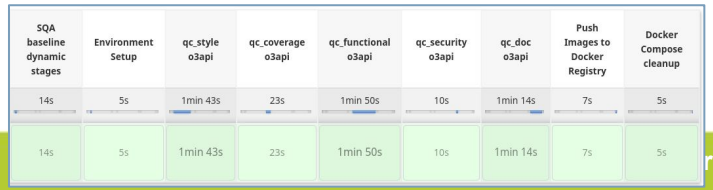
Software quality SQAaaS



Quality Assurance as-a-Service platform (SQAaaS)

- Enables the on-demand creation of CI/CD pipelines making quality verification and validation easily accessible to developers.
  - The **Pipeline as a Service** building block allows you to compose and test customized CI/CD pipelines in accordance with reference criteria.
  - The **Quality Assessment & Awarding** building block analyses, the level of compliance to the quality baselines.
- Integrates a wide range of quality verification tools that are made easily available through a friendly web interface.





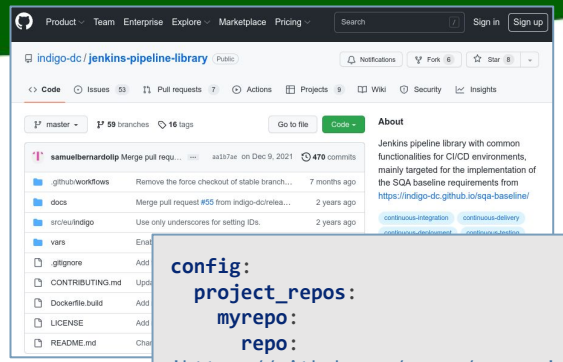
https://sqaaas.eosc-synergy.eu

Software quality JePL

Jenkins Pipeline Library (JePL)

- The library that powers the SQAaaS platform.
- Especially suitable for complex setups, you can use directly the JePL instead of the SQAaaS.
- Tech-savvy users tend to favor code over a graphical interface for the task of managing their CI/CD pipelines.
- JePL uses pipeline descriptions written in YAML.
- Just add JePL to your software repository and build your software or service quality assurance using YAML descriptions to benefit from the full set of features.
- JePL implements the software and service baselines maintained by EOSC-Synergy.

https://github.com/indigo-dc/jenkins-pipeline-library




```
config:
  project_repos:
    myrepo:
      repo:
'https://github.com/myorg/myrepo'

 sqa_criteria:
   qc_style:
     repos:
       myrepo:
         container: myrepo-testing
         tox:
           tox_file:
/myrepo-testing/tox.
```

```
|-- .sqa
|    |-- config.yml
|    |--
docker-compose.yml
  |-- Jenkinsfile
```

https://github.com/indig

```
services:
  myrepo-testing:
    image:
"indigodatacloud/ci-images:python
3.6"
    hostname:
"myrepo-testing-host"
    volumes:
      - type: bind
        source: ./myrepo
          target:

  /myrepo-testing
```

IM

- Virtual Infrastructures on IaaS Clouds
- on-premises, public and scientific clouds
- Multiple provider interfaces supported:
  - OpenNebula, OCCI, EC2, GCE, Azure,
  - Docker, K8s, Openstack, libvirt, EGI, etc
- Spanning providers
- Recipes for common deployments
- TOSCA profiles
- CLI, web, XML-RPC and REST interfaces

https://github.com/grycap/im
https://github.com/grycap/im-client
https://github.com/grycap/im-web

EC3

- Elastic Cloud Computing Clusters
- Leverage IM to interface with clouds
  - EC2, OpenNebula, Openstack, EGI, …
- Leverage spot instances
- Clusters across clouds with VPN/SSH
- Recipes for clusters
- Web and CLI interfaces

https://www.grycap.upv.es/ec3

OSCAR

- Event-driven serverless computing model for data-processing applications
- Automatically deployed on multi-Clouds
- Execute on customized runtime environments provided by containers that run on elastic Kubernetes clusters
- An OSCAR cluster can be easily deployed via the IM dashboard

https://github.com/grycap/oscar

SCAR

- Execute containers in AWS Lambda Convenient approach to run generic applications on AWS Lambda FaaS
- SCAR is integrated with
  - API Gateway to expose applications
  - AWS Batch for long-running applications
- Serverless workflows by combining functions that run on either AWS Batch or AWS Lambda

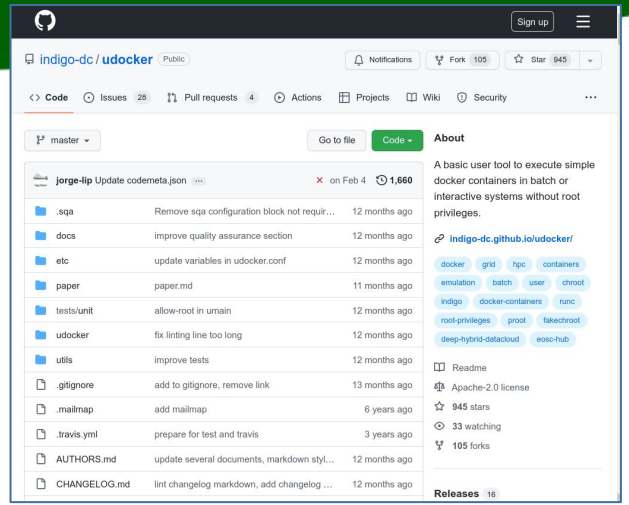
https://github.com/grycap/scar

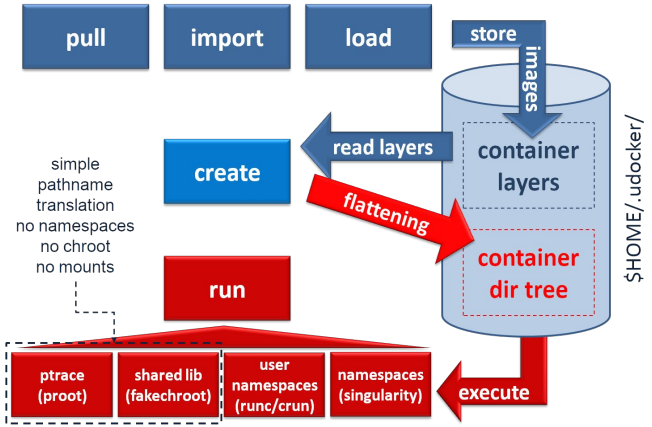
udocker

User tool to execute docker containers in user space.

- Fully user space.
- No root privileges required to use or install.
- Does not require compilation.
- Download and execution of docker containers by non-privileged users.
- Suitable for Linux batch systems and interactive clusters managed by other entities such as grid infrastructures.
- Does not require Linux namespaces.

https://github.com/indigo-dc/udocker

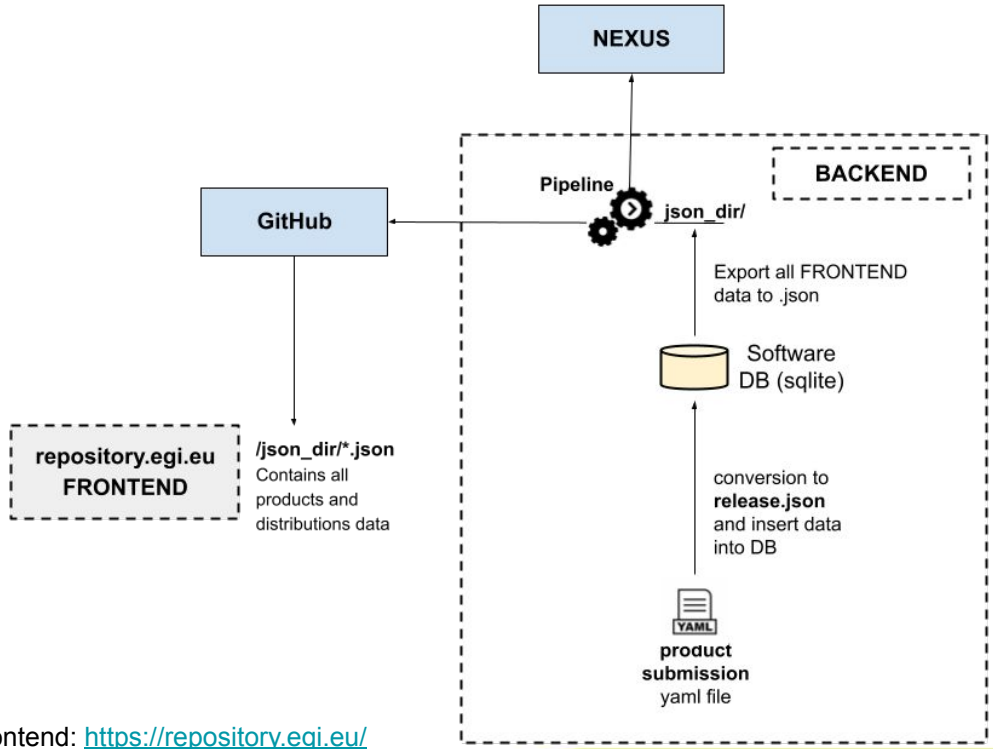




Software provisioning and Infrastructure

- Both sw provisioning and infrastructure are provided by IBERGRID since 2021
- Software provisioning
  - Verify, test software for EGI distributions
- Software Infrastructure
  - EGI Software repositories service
  - Composed by a frontend and backend integrate with the Software provisioning
- Complete redesign of the service:
  - New infrastructure based on Nexus Repository OSS
  - Release workflow implemented in GitHub and Jenkins pipelines.
  - New public webpages
- Quality assurance process
  - Full-automated approach
  - Faster release cycle with higher automation and less manual intervention

Software Infrastructure (repositories)



New setup:

- **Complete redesign of the service**
- Allow full automatization of the process
- Easier to add new functionalities like: add new OSes, add containers support
- Better integration with Jenkins
- New Frontend layout based on software products rather than releases.
- **In production for UMD-5 ONLY**
- **UMD-4 still uses the old infrastructure**
- **CA's will be moved soon**

New functionalities:

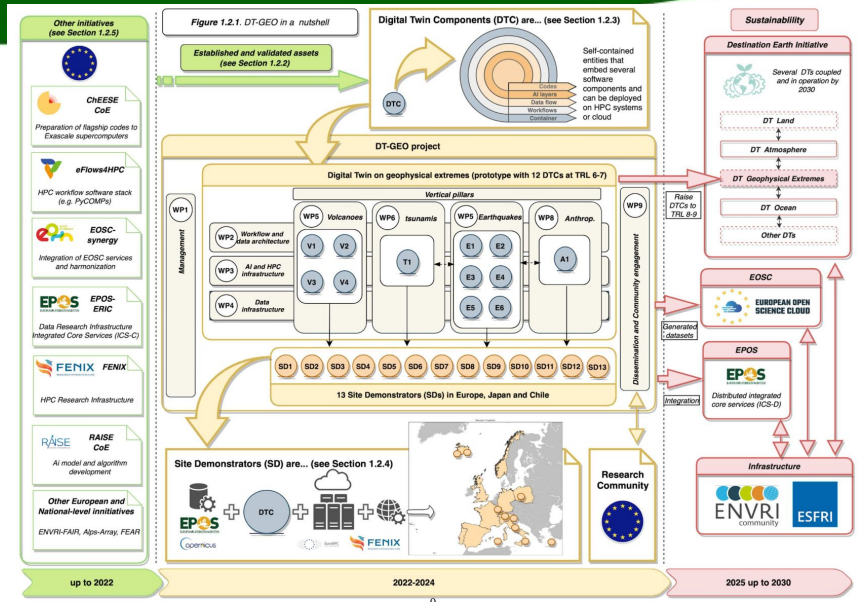
- IPv6
- High Availability (HAProxy)
- Tracking of workflow based on Git
- User interface

Frontend: https://repository.egi.eu/
Backend: https://repository.egi.eu/software/umd/5/

DT-Geo

- **DT-Geo**
  - Digital Twin of geophysical extremes.
  - Analyse and forecast the impact of geohazards from **earthquakes, volcanoes, tsunamis and anthropogenic seismicity.**
  - For integration in **Destination Earth** Initiative.
  - Urgent computing, early warning forecast and rapid post-event assessment.
  - Started in September of 2022.

- **Contribution**
  - Software and Service Quality assessment
  - Integration of workflows and containerisation
  - Providing tools for containerisation (udocker)

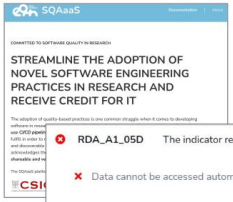


T4.1 — **Asset management (BGS, CSIC)**
Defines architecture for datasets, data products, (web)services, software and other digital assets

T4.2 — **A quality-based ecosystem for software and FAIR data (CSIC)**
Extend the EOSC-synergy project results on SQAaaS using CI/CD pipelines

T4.3 — **Integration of digital assets (BGS, CSIC)**
Automated or semi-automated workflow orchestration and provide appropriate tools for interfacing and conversion to permit integration and interoperation

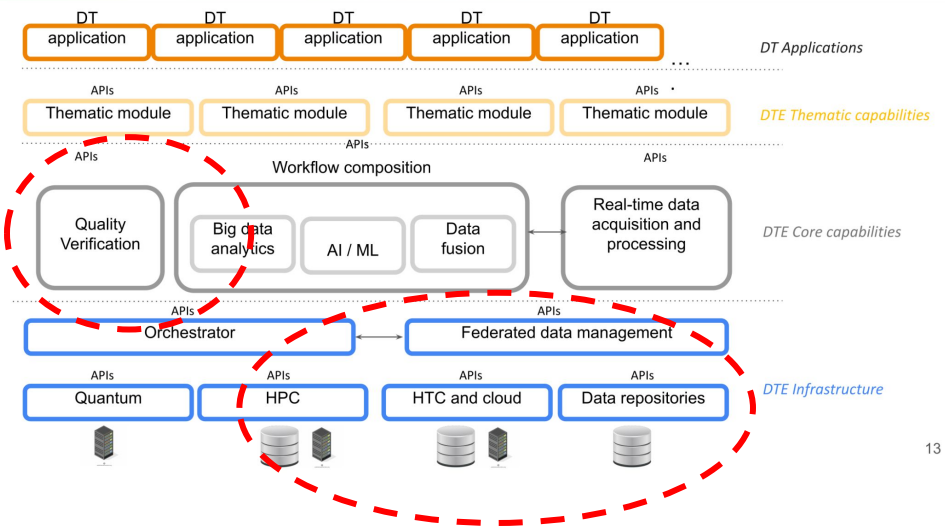
2. **Initial QA tracking** of DT-GEO software and data products through SQAaaS platform:
   - QA reports: https://gitlab.com/dtgeo/QA_reports_dtgeo
   - New features on the platform: i) support for private repos, ii) FAIR report (incl. hints for improving), iii) load previous QA reports

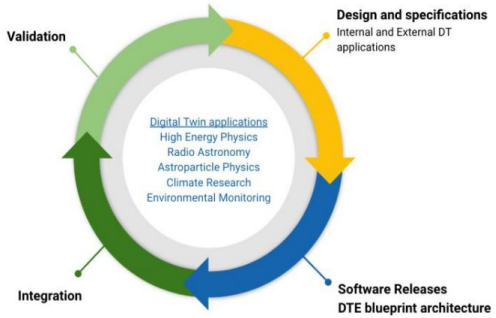


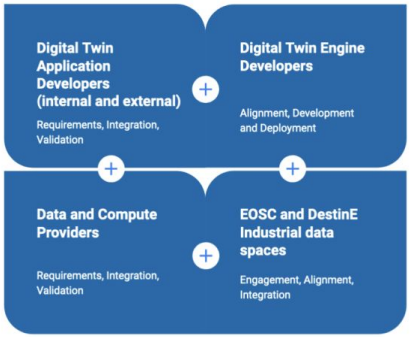
interTwin

- **InterTwin**
  - Develop a common approach to the implementation of Digital Twins (digital twin engine - DTE)
  - Applicable across the whole spectrum of scientific disciplines
  - Open-source interoperable platform
  - Software components for modelling and simulation to integrate application-specific DTs
  - Blueprint architecture for DTs
  - Liaison with Destination Earth
  - Started in September 2022

- **Contribution**
  - Software release and management
  - Quality and validation for applications, models and services







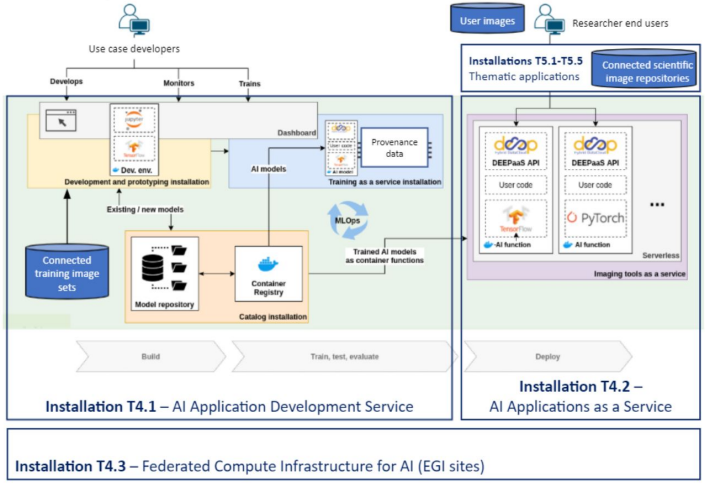
iMagine and AI4EOSC

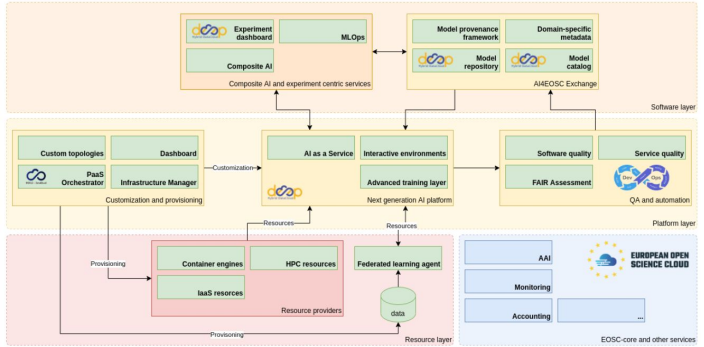
iMagine (2022-2025)

- Imaging data and services for aquatic science
- Ocean warming, and acidification
- Litter and oil spills monitoring of water surfaces
- Marine biomass estimation and preservation through real-time monitoring
- Coastal ecosystems, and beach-related human activities monitoring and analysis

AI4EOSC (2022-2025)

- Development of the DEEPaaS platform
- Advanced services for Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (DL) models and applications for EOSC.





EUCAIM

- Cornerstone of the European Commission-initiated European Cancer Imaging Initiative
- Foster innovation and deployment of digital technologies in cancer treatment and care
- Aims to enable more precise and faster:
  - predictive medicine
  - diagnostics
  - clinical decision making
  - treatment

- Establish a pan-European federated infrastructure of FAIR pan-cancer anonymized images
- Showcase how medical images can be accessed, used and pooled, while ensuring a high level of ethics, trust, security and personal data protection
- Experimentation Platform for the development and benchmarking of AI tools toward Precision Medicine in cancer diagnosis and treatment.

Earth sciences

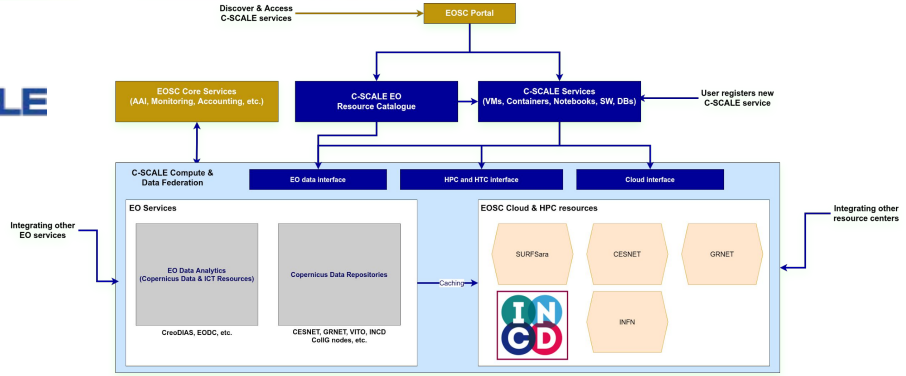
C-SCALE

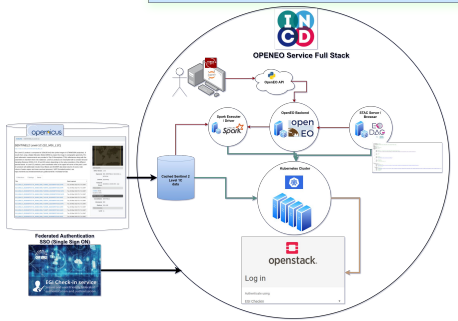
European Open Science Cloud project.

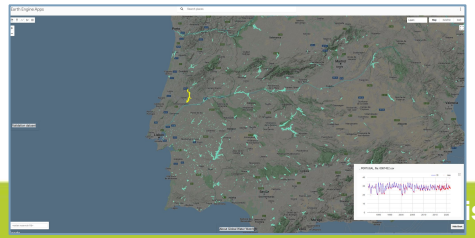
C-Scale (2021-2023) federate European EO infrastructure services:

- Copernicus DIAS and others.
- Capitalise on the European Open Science Cloud (EOSC) capacity and capabilities.
- Support Copernicus research and operations with large and easily accessible European computing environments.

Supporting EO use cases and providing cloud computing resources.







Aqua Monitor detects how the Earth's surface water has changed during the last 30 years.

Changes are detected in real-time using satellite imagery for any place on Earth.

Porting the application from the Google Earth Engine platform to the open C-SCALE infrastructure, providing an interactive (zoomable) map that displays land use changes (wet vs dry).

Relies on the top-of-atmosphere reflectance images from Landsat 4,5, 7, and 8 and will be extended to use Sentinel-2 MSI Level-1C data.

Kubernetes on INCD cloud.

HL-LHC

- High Energy Physics.
- Expected for 2029-2030.
- Looking at the particles (the very small).
- Massive increase of luminosity and detector upgrades for Run-4.
- multi-ExaByte data processing regime.
- Federations and data-lakes.
- CPU-only sites with cache-storage, opportunistic HPC and Clouds.
- Software optimisation and accelerators.

SKA

- Radio Astronomy.
- Expected for 2028-2029.
- Looking at the universe (the very big).
- Data will be taken in multiple locations.
- Rates of Terabits per second to on-site central processing facilities.
- 700 Petabytes of data per year.
- Data will be distributed to regional centres around the globe.
- Processing using HPC.

Conclusions

- IBERGRID a very active federated infrastructure growing and evolving continuously.
- Is a major regional infrastructure at European level and within EGI.
- Delivers cloud and grid capacity serving important research communities of all sizes.
- Through its federation members participates in and supports a wide range of projects.
- Develops and leverages unique software components and services.
- Has consolidated its capabilities and responsibilities in software QA and management.
- Is exploiting new paths in digital Twins, earth observation, health and other domains.
- Looking forward to new challenges …

This edition of IBERGRID is dedicated to the memory of Vicente Hernández.

PI of GRyCAP at UPV and a key founder of IBERGRID.

For further information

- http://www.ibergrid.eu/

Thank You