EGI: Advanced Computing for Research

www.egi.eu
@EGI_eInfra

La fédération EGI

Services, utilisateurs et EOSC

Baptiste Grenier
EGI Foundation

The work of the EGI Foundation
is partly funded by the European Commission under H2020 Framework Programme

Contenu

- La Fédération EGI
- Focus sur quelques services EGI
  - Check-in
  - Cloud
  - Stockage
  - Notebooks
  - Application on Demand
- Exemples de communautés d’utilisateurs
- EOSC, European Open Science Cloud et EGI

EGI: Calcul avancé pour la recherche

Mission

Créer et **fournir des solutions « ouvertes et libres » pour la science** et les infrastructures de recherche en fédérant les outils, ressources et expertises numériques des communautés par delà les frontières.

Vision

Les chercheurs de toute les disciplines profitent **d'un accès simple, intégré et libre** aux **outils avancés de calcul scientifique, ressources et à l'expertise** nécessaires pour collaborer et réaliser **des recherches utilisant des données et du calcul intensif** et innover.

EGI est une fédération de > 250 centres de calcul et de stockage répartis à travers l'Europe et le reste du monde.

47 Pays

> 71,500 utilisateurs

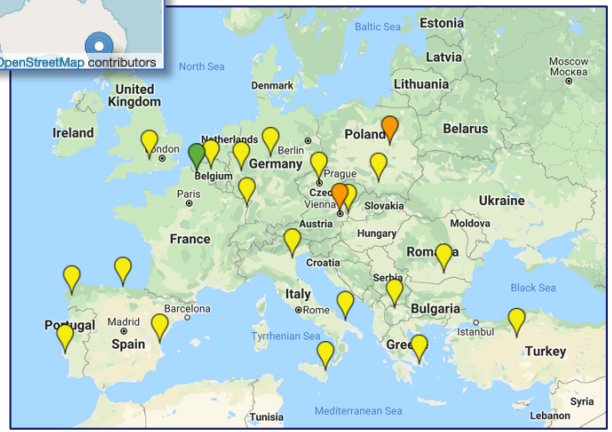
3,000 Publications en accès libre en 2018

12 e-Infrastructures intégrées

31 collaborations de recherche de grande taille

La Fédération EGI (Sept. 2019)

Centres de ressources



Fournisseurs de cloud

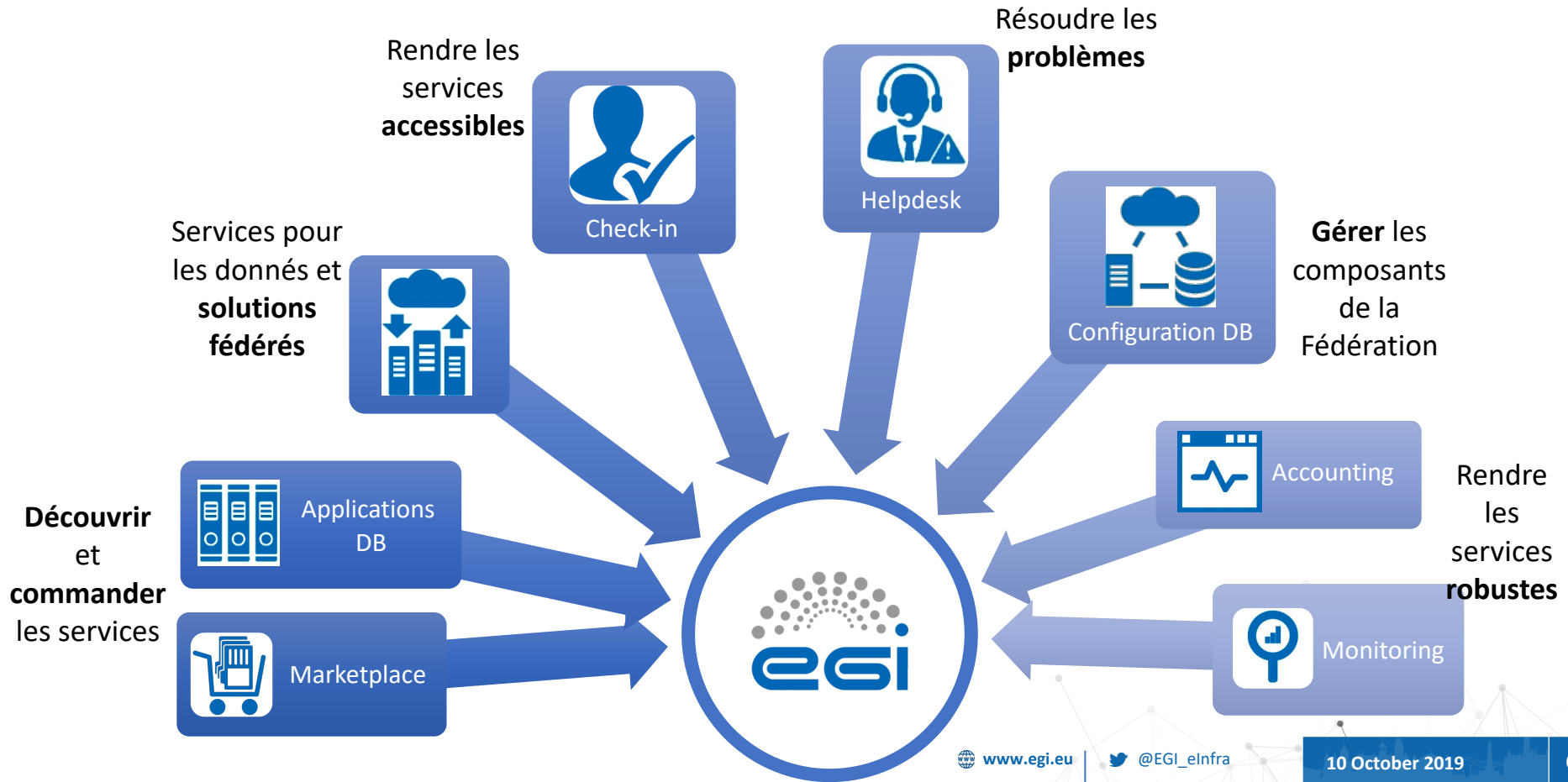
5.0 Milliards CPU core wall time / an

> 1 Million Coeurs de calcul en 2019

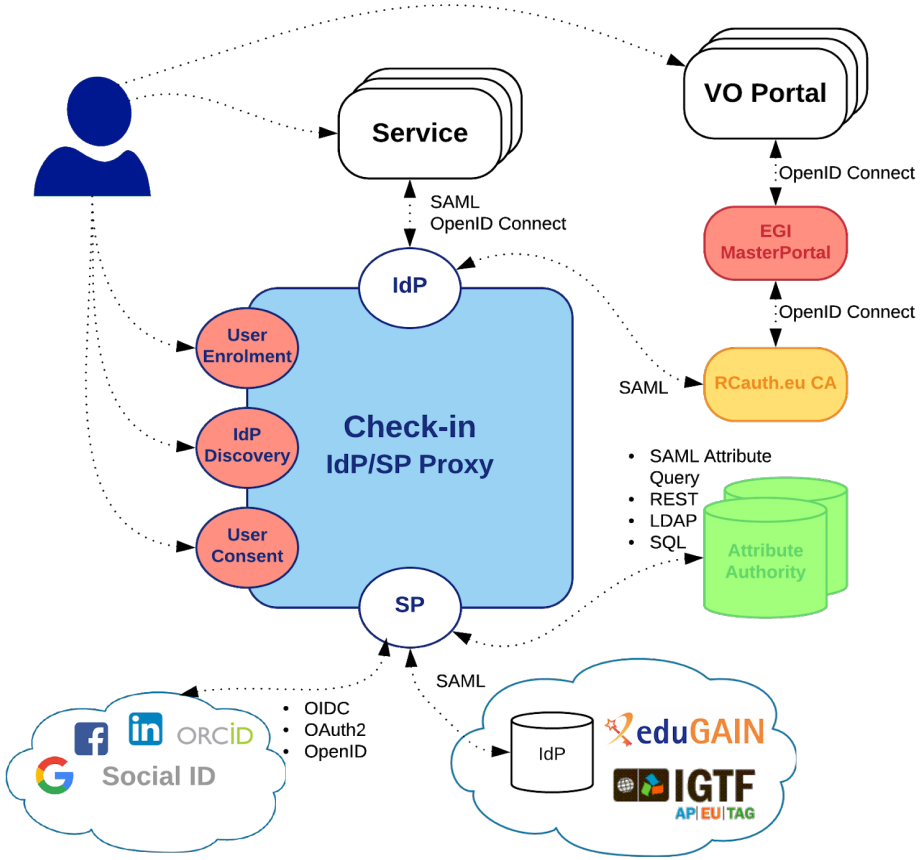
> 740 PB disque & tape

2,915 service end-points

Services pour le calcul distribué



Check-in - AAI



Une implémentation de l'architecture **AARC**

- Point d'intégration unique pour IdP et SP
- Tout les services finaux connectés peuvent avoir un seul IdP de configuré
- Pas besoin d'avoir un service de découverte d'IdP (WAYF)
- SP reçoivent des identifiants et des attributs d'autorisation harmonisés
- Expression du niveau de confiance de l'identité (ex: X509 > média social)
- Intégration avec RCAuth.eu Online CA pour fournir des certificats proxy X.509 (TTS)
  - **Intégration avec DIRAC** en cours (pour EISCAT_3D, SKA/AENEAS,…)

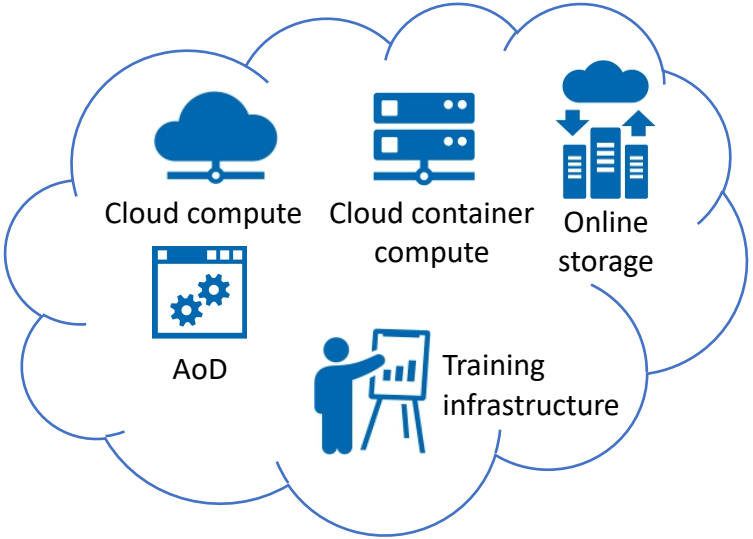
EGI Federated Cloud

Multi-cloud IaaS, identité unique via Check-in

Exécuter des Machines Virtuelles a la demande

Technologies:

- **OpenStack**,
- OpenNebula
- Synnefo



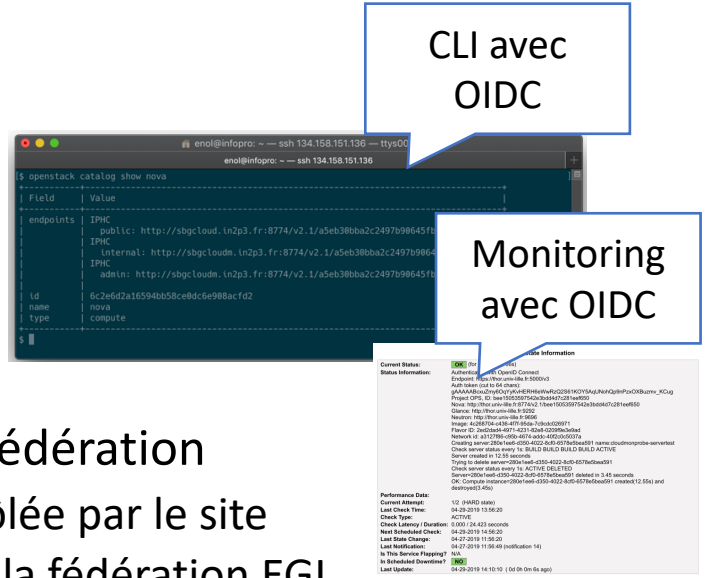
Fonctions Additionnelles

- Identité Unique via Check-in
- Bibliothèque d'Appliance
- Découvertes de ressources
- Dashboard GUI unique
- Calcul près des données

Evolution du Cloud Compute

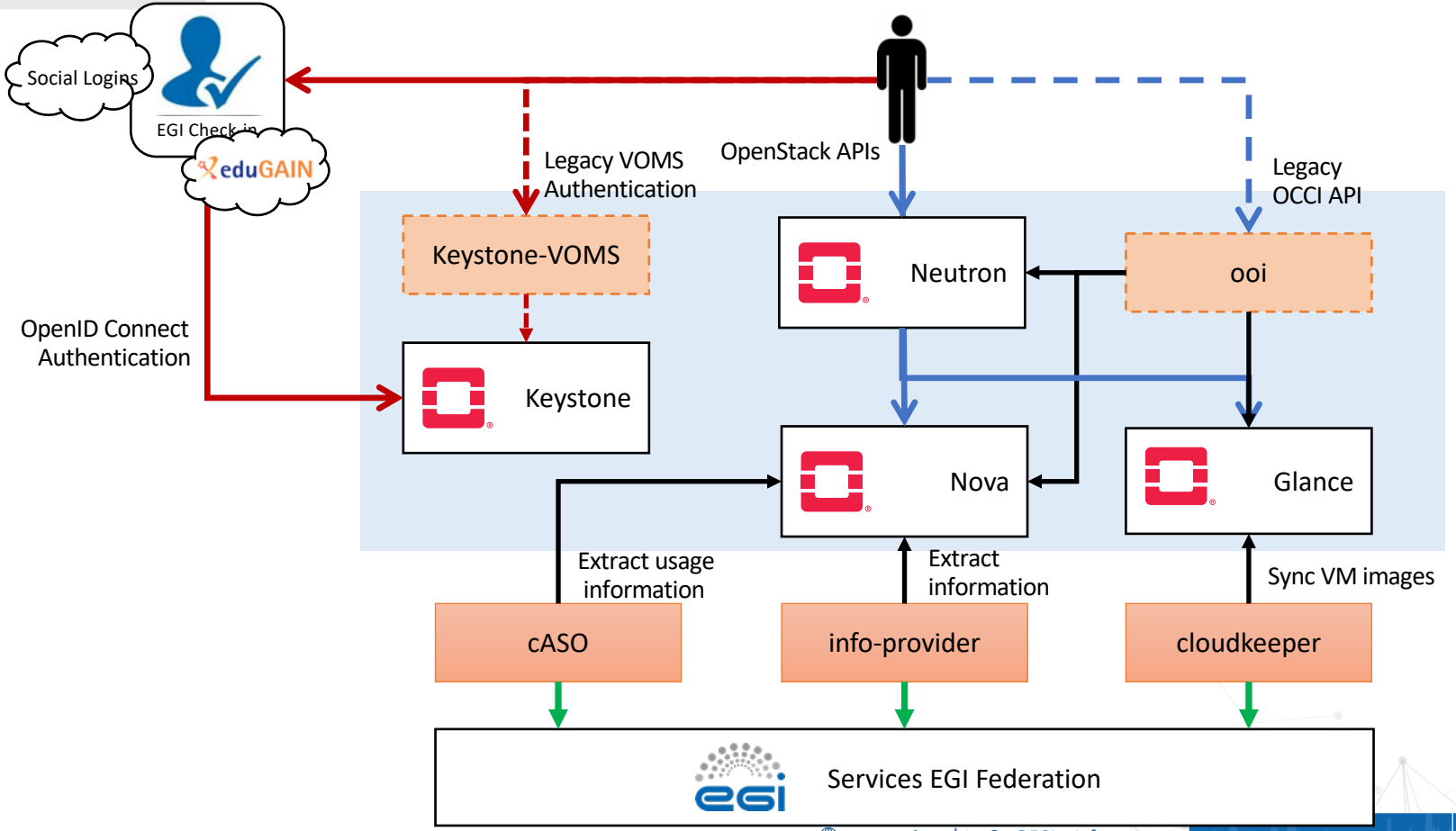
AAI et opérations centralisée

- Transition de X.509 vers OpenID Connect
  - 100% des sites OpenStack supportent OIDC !
  - OpenNebula pilote
  - AppDB finalise le support fin 2019



- Transition vers une gestion centrale des outils de fédération
  - Intégration AAI avec Check-in (OIDC) toujours contrôlée par le site
  - Mais les sites n'opéreront plus de connecteurs pour la fédération EGI
    - Sous le contrôle de l'équipe EGI Cloud opérations

Evolution: perspective pour les sites



EGI Cloud Container Compute

Exécution de conteneurs dans EGI Cloud Compute

- Service en beta depuis un certain temps
- Notre compréhension de l'exécution de containers a évoluée
  - Début avec « utilisons nova-docker », transition par « déployons des conteneurs dans no VMs » vers « fournissons des déploiements Kubernetes a la demande »
  - EC3 est maintenant près pour déployer des cluster Kubernetes élastiques
- Et la suite ?
  - Des tests, des tests et des tests !
  - Pousser la configuration EC3 / Kubernetes en production
  - Probablement pas de gestion automatique complète du cycle de vie des clusters K8S (mise a jour et gestion des erreurs)

EGI Gestion de données

- **Archive storage:** archivage à long terme
  - Initialement basé sur du stockage a bande
    - Plusieurs participants EGI l'utilise en interne
  - Partenariat avec le CINES en cours de discussion
- **Online storage**
  - "Grid" Storage Elements
    - DPM, dCache, StoRM…
  - "Cloud" object-storage
    - OpenStack Swift

SWIFT
an OpenStack Community Project

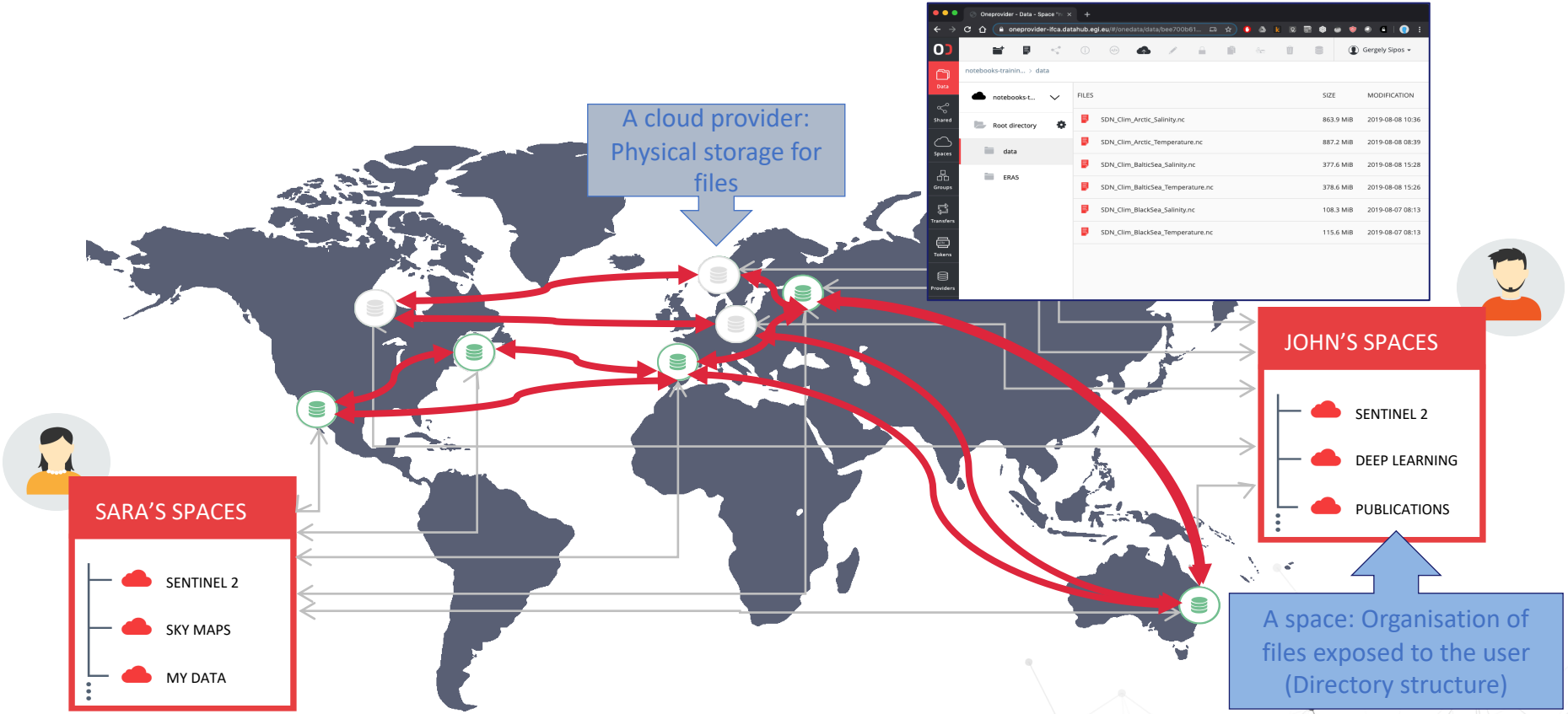
dCache

DPM
Disk Pool Manager

StoRM

- **Data Transfer**: un service pour gérer les transferts de façon sure et efficace
  - OLA avec CERN et STFC/UKRI pour FTS et/ou WebFTS

Fédération de données multi-cloud: DataHub



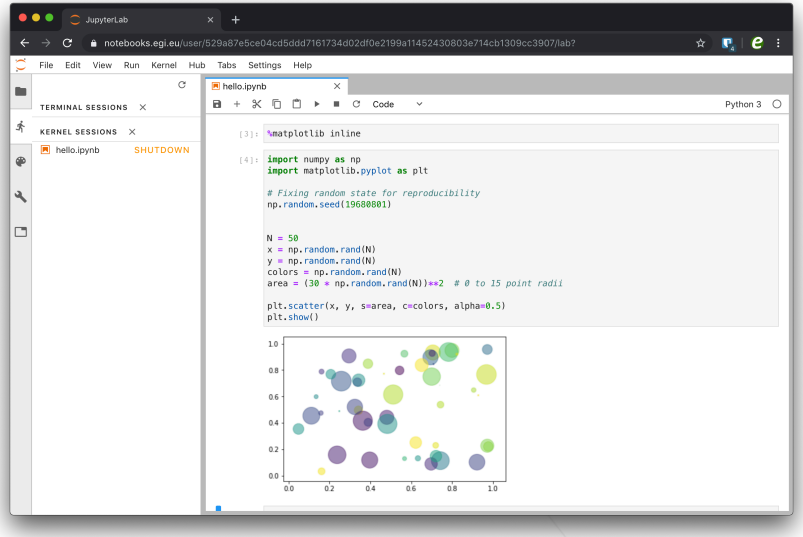
Exploration Technologique

- **Rucio:** Gestion de données distribuée
  - Une solutions distribuée
  - Un moteur basé sur des règles pour automatiser
  - Data Transfer / FTS pour planifier et gérer les transferts
  - Un service central se connectant aux endpoints de stockage existants
    - Pas de nouveau composant a déployer au niveau du site

- **Invenio:** développer des entrepots de données
  - Un **framework** pour développer une solution
    - DOI/PID, découverte, citation
    - Plusieurs “implémentations”: Zenodo, CERN Videos, CERN Open Data, Reana, B2SHARE,…
  - En attente des résultats du projet InvenioRDM
    - Une solution open source de gestion de données de recherche prête a utiliser

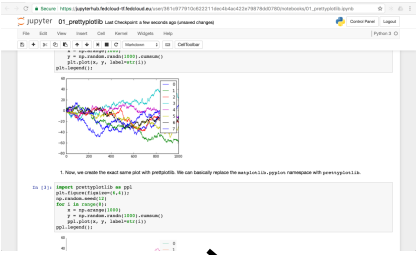
EGI Notebooks

- Exécution de blocs notes Jupyter depuis le navigateur:
  - **EGI Check-in**
  - **Stockage persistant**
  - **Calcul « élastique » dans EGI Cloud**
- Instance publique et gratuite avec 1 vCPU/1 GB RAM par utilisateur: https://notebooks.egi.eu
- Déploiements spécifiques pour des communautés basés sur des accords (signature de SLA)



Notebooks (+ DataHub): architecture

https://notebooks.egi.eu



EGI CheckIn

AuthN/AuthZ

EGI DataHub

Space 1

Space 2

Space N

AuthN/AuthZ

Récupérer
les espaces supportés

Mise en cache des données

Notebook
Interactif

jupyterhub

jupyter

jupyter

jupyter

Oneprovider

Notebooks
Stockage
Persistent

Cluster Kubernetes dans EGI Cloud Compute

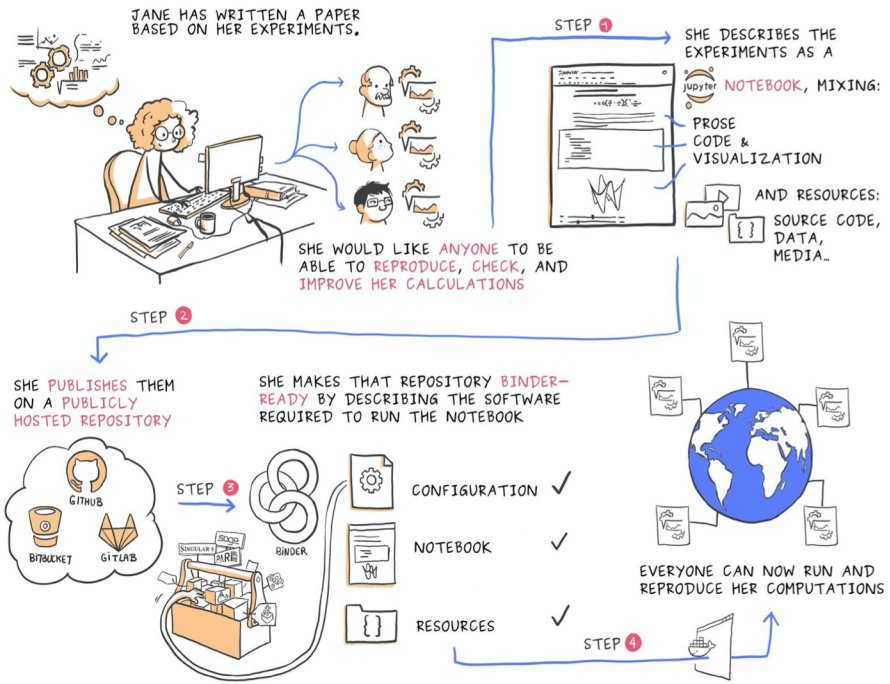
EGI Notebooks : stockage

- EGI Notebooks peut également interagir avec d'autres systèmes de stockage :
  - D4Science VRE Workspace (en **production** pour AGINFRA+)
  - B2DROP (tests effectués pour EPOS-ORFEUS)
  - NFS server pour des données partagées (en **production** pour LABSS)
  - Volumes OpenStack Cinder (tests effectués pour Max-CoE)

- Peut potentiellement intégrer tout les stockages qui peuvent être exposés en tant que volume kubernetes :
  - B2SAFE (e.g. en utilisant l'interface webdav iRods)
  - CEPH
  - ...

EGI Notebooks pour la reproductibilité

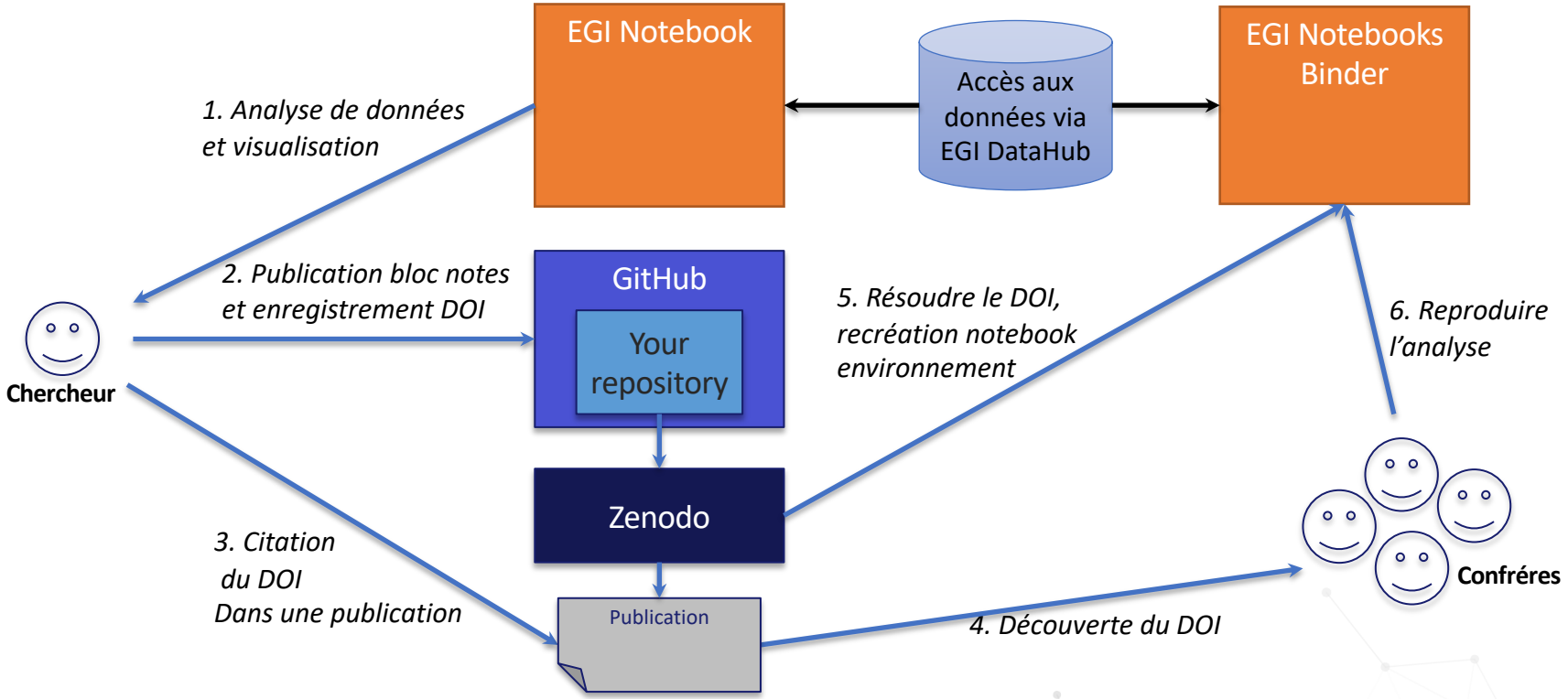
- Binder: transformer un dépôt en notebook interactif
- Des technologies modernes pour l'orchestration Cloud (Kubernetes), le calcul interactif (Jupyter), le calcul scientifique (l'écosysteme open-science)
- En cours d'intégration avec le service EGI Notebooks service



Par @JulietteTaka

Open science avec EGI Notebooks + DataHub

Jupyter-Binder-Cloud-DataHub-Zenodo-GitHub

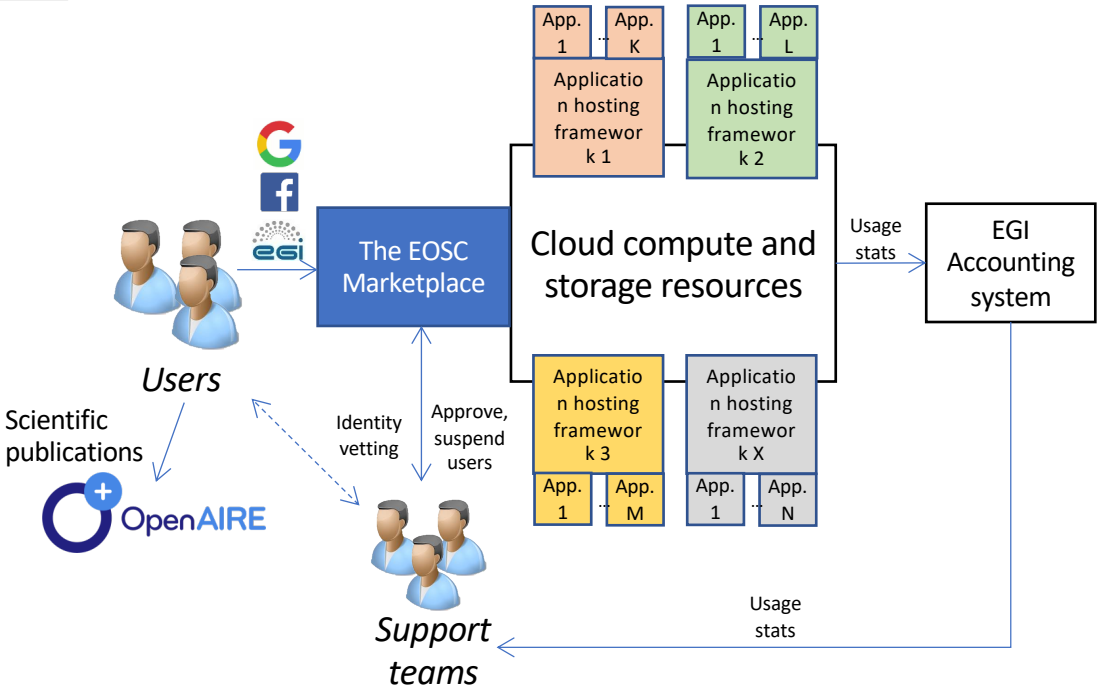


Applications on Demand

*Une **plateforme ouverte** pour partager et réutiliser des applications*

- Portails applicatifs, Science Gateways, VREs « as a Service »
  - Elastic Cloud Computing Cluster (EC3): Déploiement d'infrastructure sur du IaaS
  - EGI VMOps dashboard web pour gérer des machines virtuelles
  - Des « Science Gateways »
    - INDIGO-DatCloud FutureGateway Science Gateway (FGSG)
    - WS-PGRADE/gUSE: Sceince Gateway orientée workflow (gUSE)
  - Instance Catch-all de EGI Notebooks
  - Applications scientifiques (déployées dans les gateways/portails/VREs)
    - Chipster, Galaxy, AutoDock Vina, NAMD, RStudio, REPAST,…
- Un pool de ressources
  - 3 fournisseurs de Cloud supportent le service (CESGA, IN2P3-IRES et INFN-CATANIA-STACK) avec 210+ vCPU cores, 700GB+ de RAM et 1.2TB de stockage
- Un système de gestion d'identité simplifié EGI AAI (Check-In + robot)
- Du support utilisateur

AoD: architecture haut niveau



La Marketplace est le point d'entrée pour accéder au service et applications disponibles

Et les utilisateurs?

(Une des façons de) Supporter les communautés d'utilisateurs

Négocier l'accès aux ressources en prenant en compte les besoins (type, nombre, taille, coût, disponibilité,…)

Type, nombre, taille, coût, disponibilité, etc.

Représentant du Projet/Communauté

Besoins pour l'Instance

Négociateur

Conditions

Fournisseur Applic.

Fournisseur Stocakge

Fournisseur Cloud

Fournisseur Grille/HTC

Formation

Support

Service Level Agreement

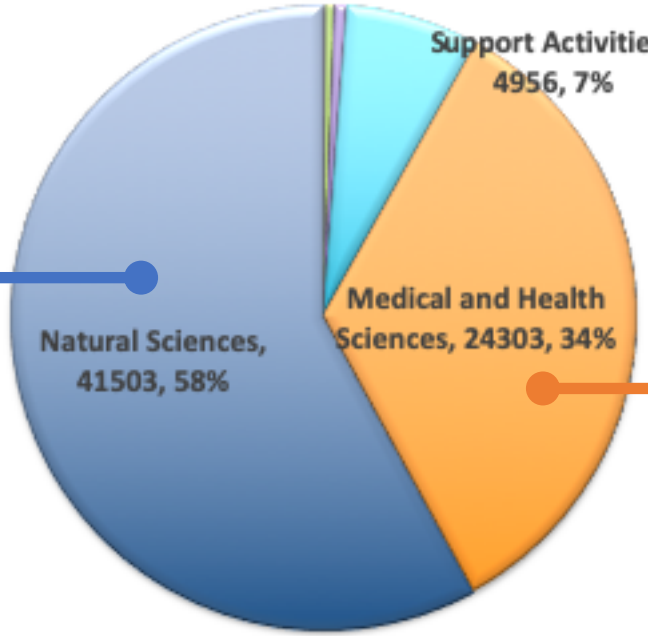
Operation Level Agreement

Accord sur les ressources requises

Réservation des ressources

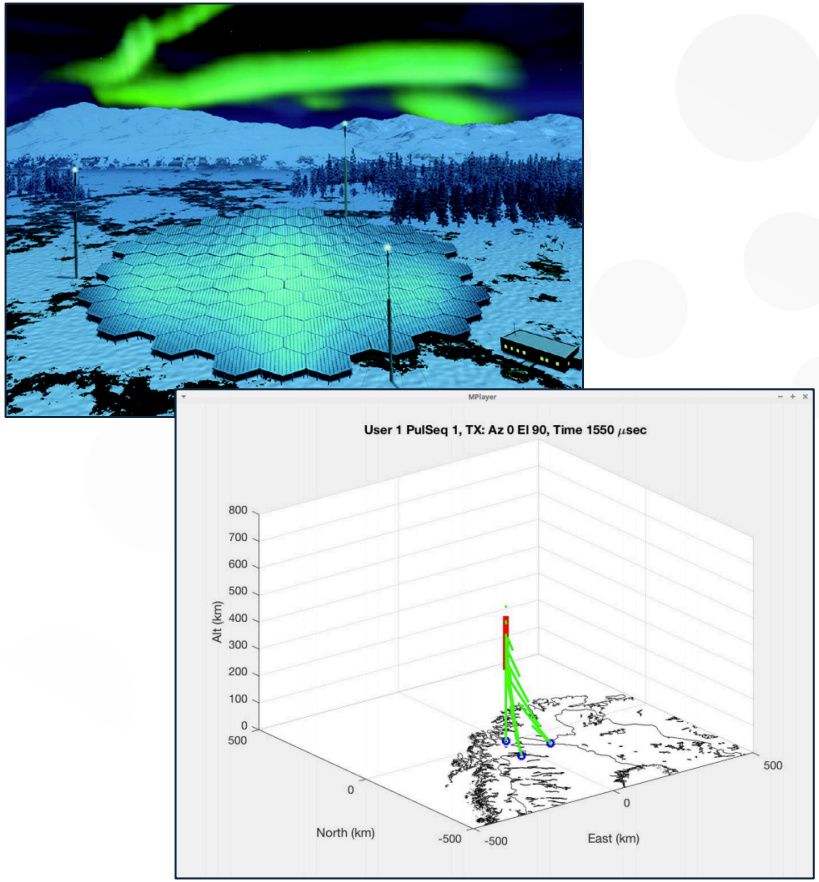
EGI Statistiques Utilisateurs (Sep 2019)

- Physical Sciences
  LHC, BELLE, ICECUBE, LOFAR, VIRGO, CTA, XENON, LSST, SKA, KM3NET, ..
- Biological Sciences
  LifeWatch
- Earth Science
  EISCAT, EMSO
- Chemical sciences



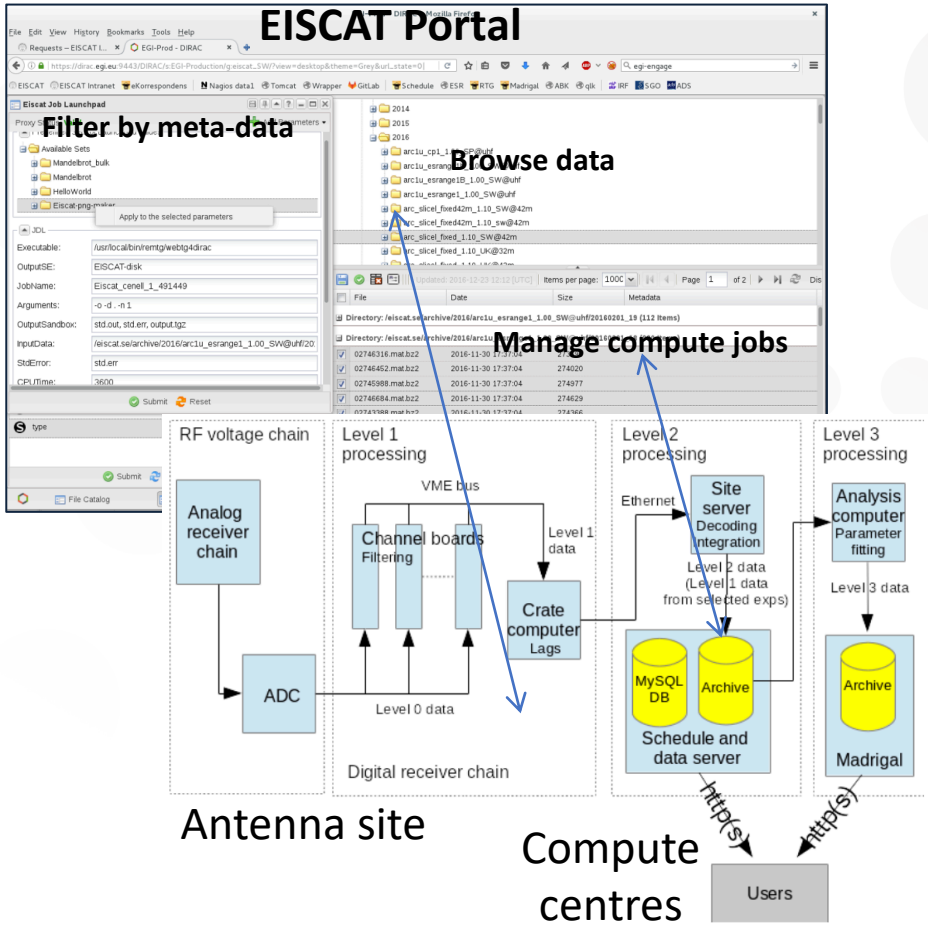
- BBMRI, ELIXIR
- Bioinformatics
- Neuroimaging
- Structural Biology
- Hematology

EISCAT_3D



- Un projet de « European Incoherent Scatter Scientific Association » (EISCAT)
- Etudier les interactions dans les régions de la ionosphère et magnétosphère (aurores boréales)
- 109 cellules de 91 antennes (+2 stations réceptrices) → Jusqu'à 100 rayon simultanés
- Production max de données: > 50 Gb/s
- Premières données prévues en 2021

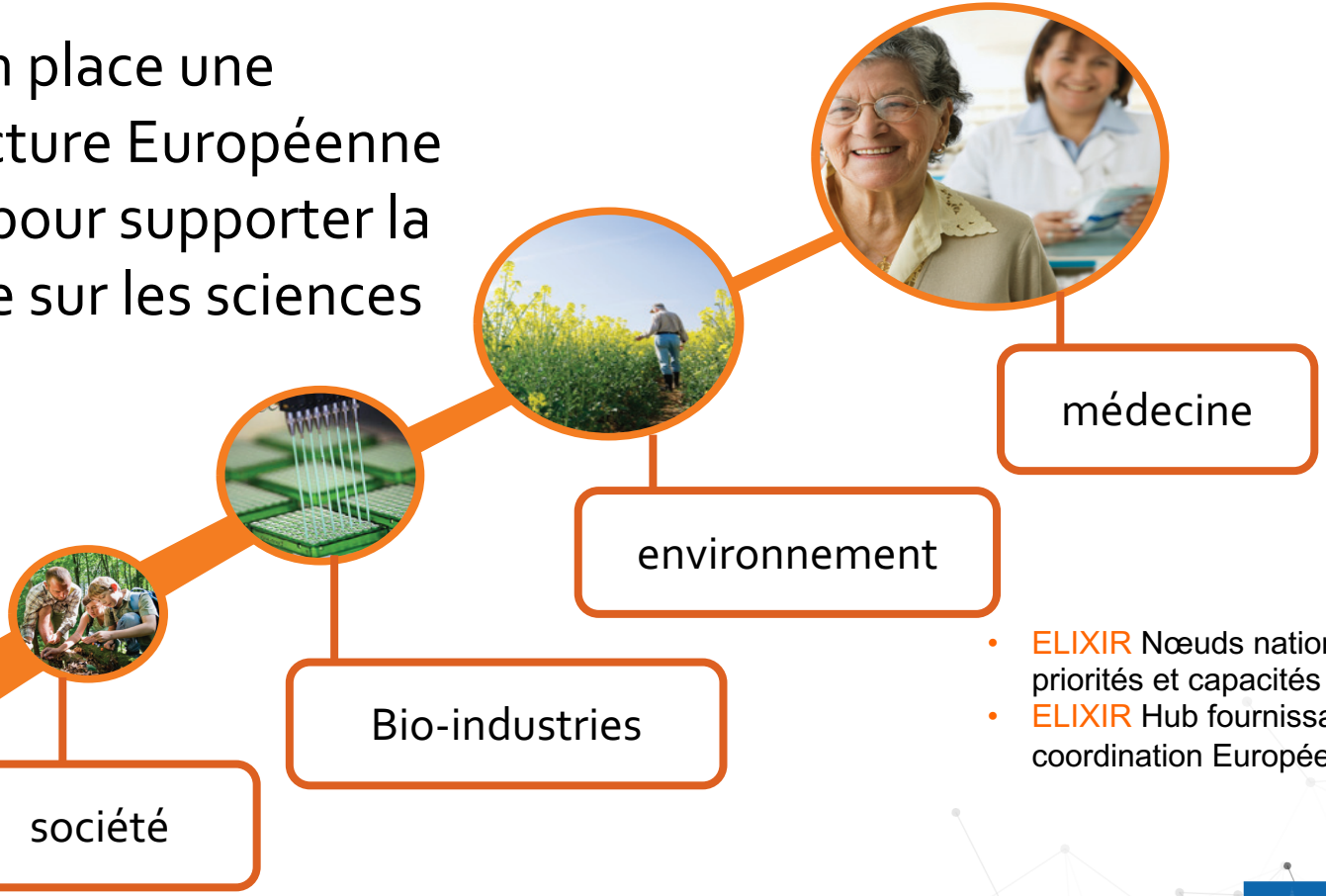
Support pour EISCAT_3D



- Intégration de services
  - Data transfer
  - Cloud computing (EGI Cloud)
  - Conteneurs
  - AAI (Check-in)
  - Portail utilisateur (DIRAC)
- Engager les utilisateur
  - Early Adopters
  - Formation technique
  - Amélioration continue du service

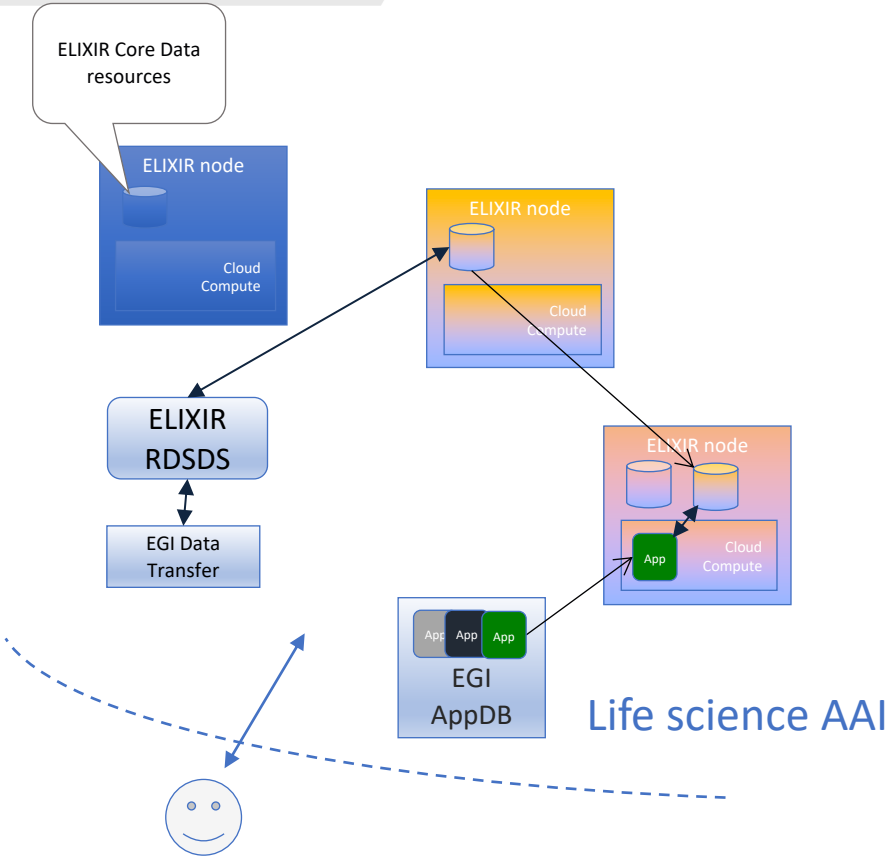
ELIXIR

Mettre en place une infrastructure Européenne pérenne pour supporter la recherche sur les sciences de la vie



- ELIXIR Nœuds nationaux basés sur les priorités et capacités locales
- ELIXIR Hub fournissant une coordination Européenne

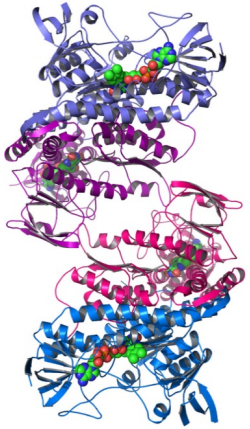
Centre de compétence ELIXIR



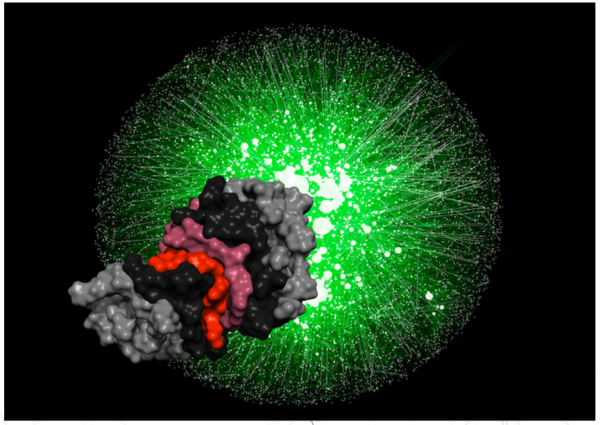
1. Intégration technique
   - EOSC-hub AAI ←→ Life Science AAI
   - EGI Cloud (Fédération de cloud)
   - EGI AppDB (Boutique d'applications virtuelles)
   - EGI Data Transfer Service (réplication de données)
2. Business model
   - Allocation de ressources aux utilisateurs et projet utilisateurs
3. Politiques d'accès aux données
   - Données sensibles
4. Formation
   - Former les centres de ressources

WeNMR

Comprendre le chemin menant à la maladie

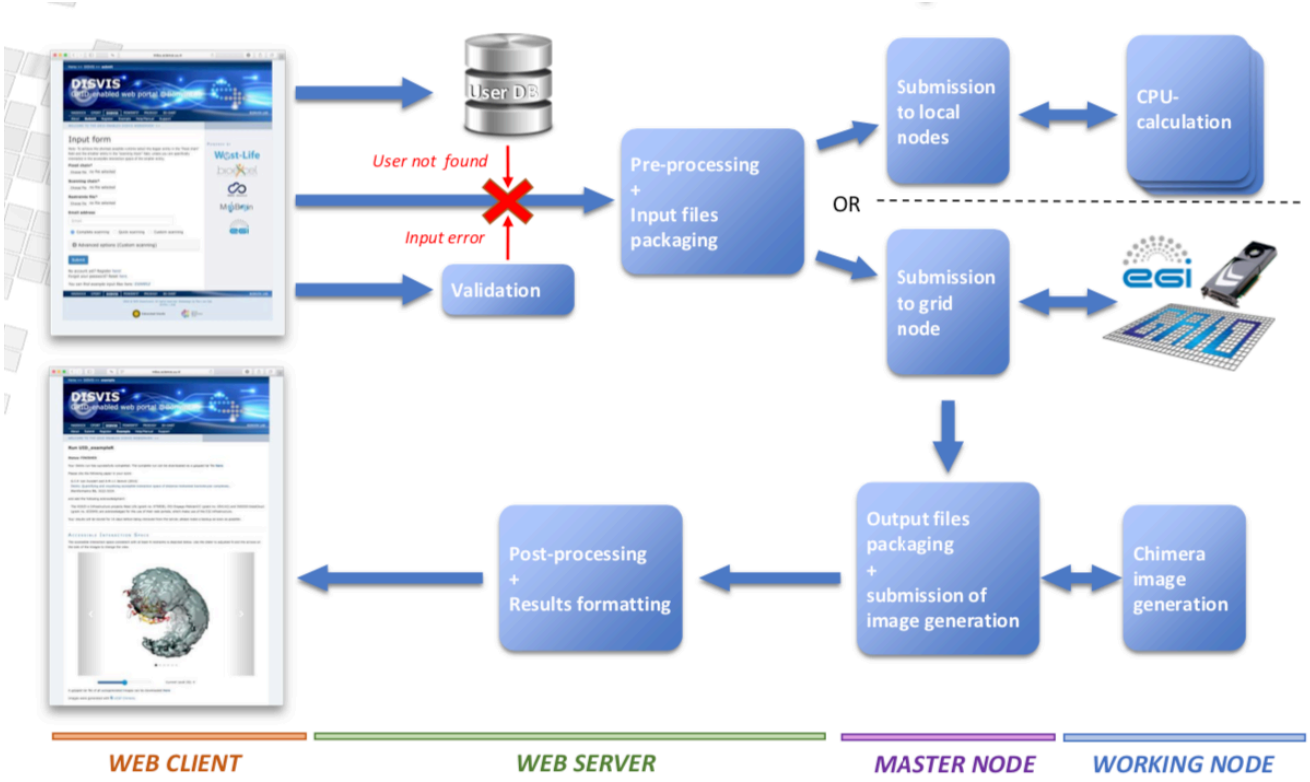


- Les protéines et biomolécules interagissent dans un réseau complexe
- Des problèmes dans se réseau peuvent causer des maladies comme le cancer
- Les chercheurs utilisent des modèles 3D pour étudier les interactions
- Cette information est essentielle pour comprendre les origines des maladies et développer de nouveaux traitements

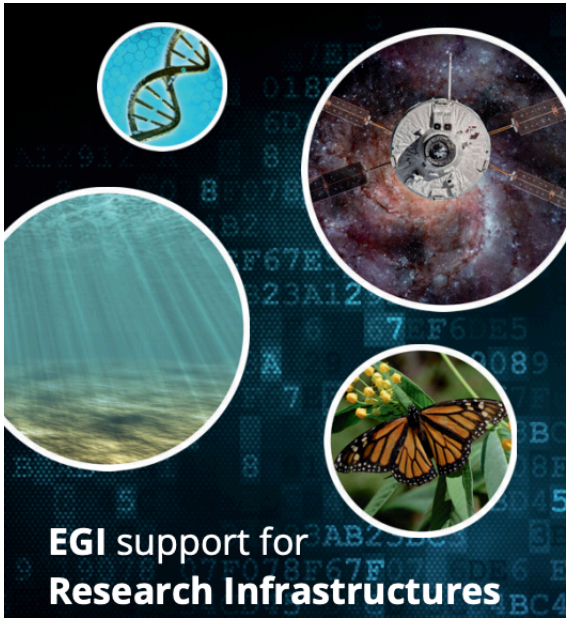


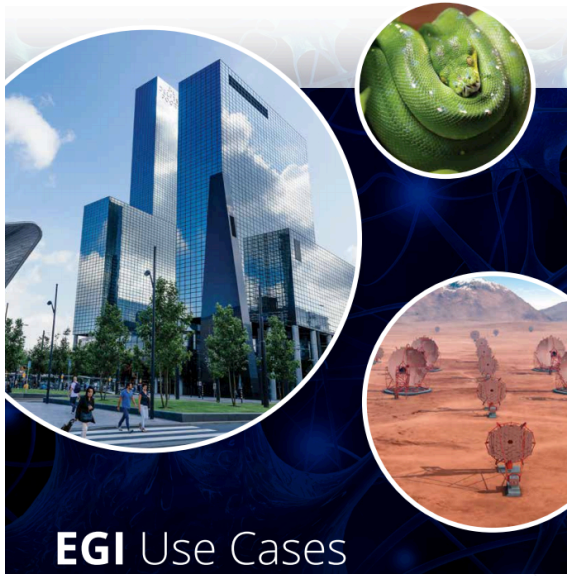
WeNMR: Architecture derrière le portail et support par EGI

- Cloud services
- High-Throughput computing
- Accès aux CPUs and GPGPUs distribués
- Outils supportant la plateforme
  - Workload manager (DIRAC)
  - AAI fédéré



« Success stories » supportées par EGI





https://www.egi.eu/about/publications/

EGI et EOSC (European Open Science Cloud)

Data, applications and computing infrastructure as integrated solution

Qu’est ce que European Open Science Cloud?

- Un environnement virtuel ouvert et de confiance avec un accès simplifié a des services (TRL 8/9) supportant le cycle de vie complet des données de recherche:
  - Fédérer et connecter les infrastructure de recherches (RIs)
  - Rendre les données FAIR, les stocker et assurer leur préservation à long terme
  - Des services pour trouver, accéder, combiner et analyser des données
  - Un environnement de travail personnalisé et protégé
- Une fédération sur plusieurs niveaux (« cœurs fédérateur », reste de l’ecosystème) joignant l’offre et la demande dans un environnement contrôlé
- Ouvert, transparent, basé sur cette règle: pas de « lock-in » par des fournisseurs de services, portabilité des données, IPR, sécurité cloud…
- Orienté utilisateur et inclusif (à travers les frontières et disciplines)
- Accessible grâce à un point d’accès simple, et non-exclusif
- Gouverné par des règles de participations claires
- Dirigé par une structure de gouvernance inclusive

Mission

Le projet EOSC-hub mobilise des fournisseurs de stature européenne offrant des services, logiciels et données pour une recherche avancée et une innovation basées sur les données.

Ces ressources sont offertes via le Hub – le système d'intégration et de management de EOSC, agissant comme un point d'entrée pour tout les participants au niveau Européen.

EOSC-hub

€ 33M

100+ partners

2018-2020

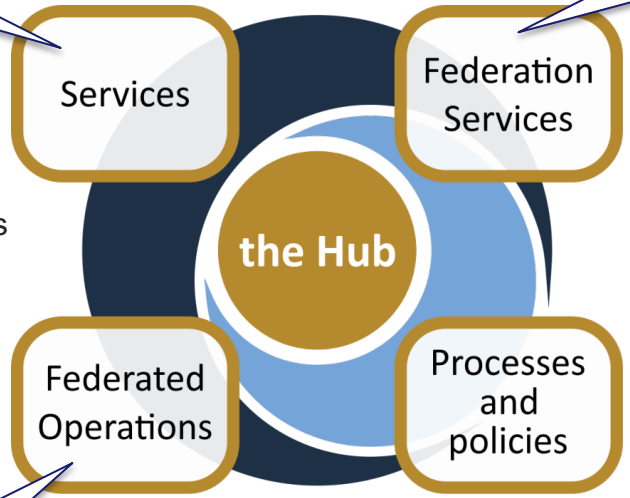
Coord. by EGI Foundation

Le projet EOSC-hub a établis le 'Hub'

Du consortium ET des **contributeurs externes**

Usage suivant les **Régles de Participation**

- Data
- Applications & tools
- Baseline services (storage, compute, connectivity)…
- Training, consultants



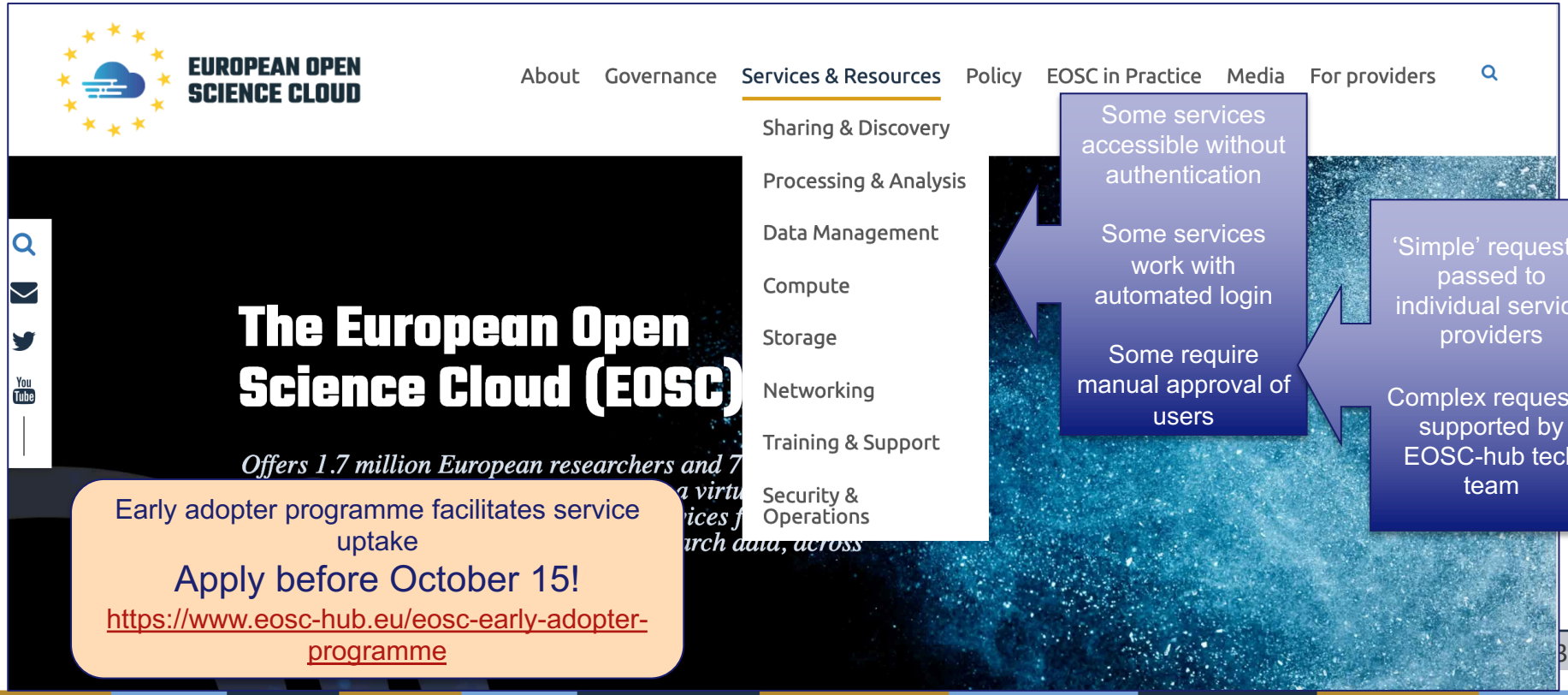
- Marketplace
- AAI
- Accounting
- Monitoring
- …

- Lightweight certification of providers
- SLA negotiation
- Customer Relationship Management
- …

- Security regulations,
- Compliance to standards,
- Terms of use,
- FAIR implementation guidelines
- …

Basé sur **FitSM**

Pour accéder à EOSC : https://eosc-portal.eu



EGI: Advanced Computing for Research

www.egi.eu
@EGI_eInfra

Thank you
for your attention.

Questions?

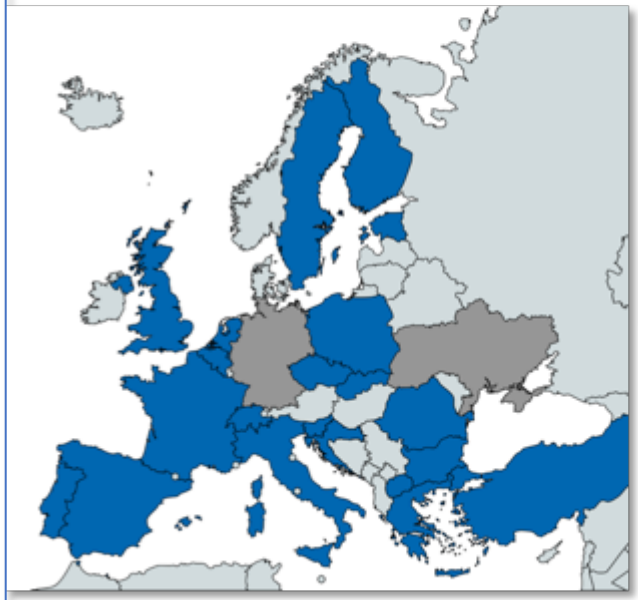
This work by the EGI Foundation
is licensed under a Creative Commons Attribution 4.0 International License.

The work of the EGI Foundation
is partly funded by the European Commission under H2020 Framework Programme

Participants EGI (2019)

23 Countries and **CERN**

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



belspo
IICT
GA Gauß-Allianz
arnes
IFIN-HH
SwiNG
grnet
srce
SNIC
CESNET
INFN
MARGI
ULAKBİM
JISC
EENet
CSIC
SURF
CYFRONET
CSC
cnrs
CERN

Partenaires Internationaux

Canada

USA

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

Latin America
Universida de Federal do Rio de Janeiro

China
Inst. Of HEP Chinese Academy of Sciences

Shanghai Science and Technology Innov. Resources Center

India Centre for Development of Advanced Comp.

Asia Pacific Region
Academia Sinica at Taiwan

> 20 SLAs, 240 Research projects

EGI Service Portfolio

Compute

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

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

High-Throughput Compute
Execute thousands of computational tasks to analyse large datasets

Workload Manager BETA
Manage computing workloads in an efficient way

Applications

Applications on Demand BETA
Use online applications for your data & compute intensive research

Notebooks BETA
Create interactive documents with live code, visualisations and text

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

Security

Check-in BETA
Login with your own credentials

Training

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

ISO 27001 Training
Learn how to manage and secure information assets

Training Infrastructure
Dedicated computing and storage for training and education

ISO 9001
Certified Quality Management System
www.tuv-sud.com/ms-cert

ISO 20000
Certified IT Service Management System
www.tuv-sud.com/ms-cert

EGI Internal Service Portfolio

Pour les membres de la Fédération

Coordination

Operations Coordination and Support
Coordinate activities to ensure seamless operations

Community Coordination
A joint approach to user engagement

ITSM Coordination
Ensures professional service management for EGI IT services

Technical Coordination
Progress and innovation through collaboration

Strategy and Policy Development
One federation, one vision, one strategy

Project Management and Planning
A joint approach to planning and management

Security Coordination
Enhance local security for a safer global infrastructure

Communications
Share your successes at a larger scale

Operations

Marketplace BETA
Expose your services to a broader audience

Accounting
Track and report the usage of your services

Collaboration Tools
IT tools for better coordination

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

Operational Tools
Integrate resources and operations in a federated ecosystem

Helpdesk
Your point of contact to ask for support at EGI

Service Monitoring
Monitor the performance of IT services

Validated Software and Repository
Benefit from a repository of high-quality software validated for the EGI infrastructure

Security

Check-in
Login with your own credentials

Attribute Management
Manage memberships and groups in communities and virtual organisations

ISO 9001
Certified
Quality Management System
www.tuv-sud.com/ms-cert

ISO 20000
Certified IT Service
Management System
www.tuv-sud.com/ms-cert