



## France Grilles dans l'EOSC

V. Breton

## 2000: contexte et enjeux

- Contexte: traitement de grands volumes de données scientifiques à haut débit dans des collaborations internationales
  - 2010: démarrage du LHC
- Enjeu: manipulation, stockage et analyse de grands volumes de données sur des ressources de calcul distribuées ou à travers des environnements virtuels de travail



## 2010 - verrous

- **Verrous:**

- Identification: comment permettre à des utilisateurs de s'identifier sur une fédération de sites ?
- Sécurité: comment gérer la sécurité d'une infrastructure fédérée nationale ou internationale ?
- Interopérabilité: comment permettre à des codes issus de multiples disciplines d'exploiter ces infrastructures ?
- Compétences: comment simplifier l'accès à ces ressources fédérées pour des non-experts? Comment former des ingénieurs à l'administration des sites?

## Objectifs généraux

- Mettre à disposition des ressources de calcul et de stockage ouvertes à toutes les communautés scientifiques en Europe et au-delà
- Former les ingénieurs à des technologies ouvertes pour le traitement à haut débit de grands volumes de données
- Promouvoir les développements issus de la recherche informatique française

## Qu'apporte France Grilles aujourd'hui ?

- **Au niveau européen :**
  - Des infrastructures distribuées de calcul et de stockage
  - Un porte-feuille de services à haut TRL(Technology Readiness Level) et interopérables pour le traitement des données
  - Une infrastructure d'authentification et d'autorisation
  - Une coordination opérationnelle en cybersécurité
  - Un lien entre les communautés nationales et EGI
- **Au niveau national**
  - Un réseau d'experts en calcul distribué (cloud, stockage) et d'administrateur de sites
  - Une expérience de médiation technique entre ces experts, les communautés d'utilisateurs et la communauté de recherche en informatique
  - Une expérience de formation des fournisseurs et des utilisateurs de services

## EOSC

- Données sans calcul ou calcul sans données ne font pas de sens
- EOSC = infrastructure distribuée de portails de données sur une infrastructure de calcul pour leur traitement



# Cas d'utilisation: apport d'EOSC face à une nouvelle pandémie

Epidemiological data

Biomolecular data (virus sequences)

Biostructural data (virus proteins)

Sociological data (Non Pharmaceutical Interventions)

(Meta-) data aggregation and semantic enrichment

Data integration

Data exploitation = workflow with periodic rerun of analysis

Epidemic models

Variant monitoring

New candidate drugs

Public health policies

## THE TRUE TOLL OF COVID

"Excess deaths" measures the impact of COVID by including people whose deaths were indirectly caused by the pandemic.

Official count of COVID deaths

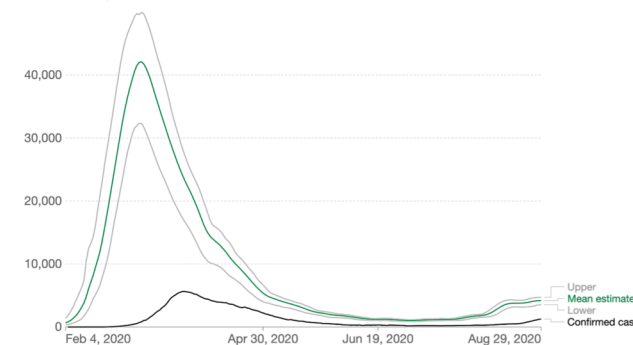
5.9M

Estimated global excess deaths

Low 17.1M High 19.6M Estimated 18.2M

Source: IHME (Data from January 2020 to December 2021)

Daily new estimated COVID-19 infections from the IHME model, Italy  
Estimates of the true number of infections. The "upper" and "lower" lines show the bounds of a 95% uncertainty interval. For comparison, confirmed cases are infections that have been confirmed with a test.



Source: IHME (2022), JHU (2022)  
Note: This chart shows the model estimates dated 12 September 2022.  
OurWorldinData.org/covid-models - CC BY

# Les verrous de 2022 sont de même nature quoique différents des verrous de 2010

- Identification
- Cybersécurité
- Interopérabilité des services
- Compétences: comment accompagner la longue traîne de la science ?
  - >90% des 100.000 + chercheurs académiques français n'utilisent pas les services des mésocentres et IR de calcul nationales

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION



300 organismes et établissements  
115 infrastructures de recherche plus de 2500 unités de recherche



115 000 enseignants-chercheurs  
61 000 personnels de soutien dont 4 000 informaticiens<sup>[1]</sup>



2,7 M étudiants  
dont 1,6 M dans les Universités



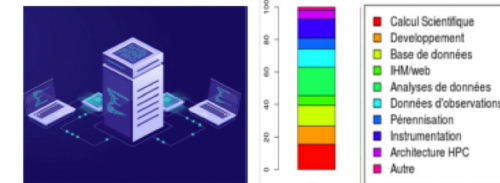
12 datacenters labellisés, 1 centre de calcul Tier 0, 2 centres de calcul Tier 1, 26 mésocentres Tier 2, 1200 salles machines et 20 000 serveurs physiques<sup>[2]</sup>,



2 000 Po de données stockées,  
15 pétaflops de capacité cumulée de calcul



Au niveau national :  
12 000 km de fibres optiques, 1 300 sites connectés,  
72 points de présence,  
150 longueurs d'onde de 1 à 200 Gb/s de débits<sup>[3]</sup>



Pratiques liées à l'activité scientifique

Des activités numériques variées : calculs, développements, bases de données, instrumentation<sup>[4]</sup>



Système d'informations et services numériques : par établissement, plus de 150 applications métiers installées et maintenues

**EOSC: rendre les données ET les services FAIR**



## De nouveaux verrous sont apparus...

- Utilisation par la société civile (entreprises et science citoyenne)
- Modèle économique
- L'enjeu écologique
- La reproductibilité des calculs et la réutilisation sur les différents types d'infrastructure
- Et le coût de l'énergie



## Les besoins et les services à construire pour la communauté scientifique

- un catalogue national de services/ressources de calcul et stockage
  - Interopérable avec l'EOSC-Catalogue européen
- Une infrastructure nationale centralisée d'autorisation et d'authentification
  - Intégrée au niveau européen
- Des environnements de gestion de grands volumes de données (data lake)
- Un accès simple aux ressources de cloud pour toutes les communautés

## France Grilles dans la construction d'EOSC

- Projets européens
- EOSC task forces:
  - Technical Interoperability of Data and Services, Researcher engagement and adoption, Long Term Data Preservation



Pioneering EOSC



Computing platform for EOSC



Regional EOSC PROJECT



Platform for Open Science  
and FAIR data



Implementing FAIR principles



Services for Earth observation

## Conclusion

- France Grilles déploie depuis 2010 des services pour le traitement de données à haut débit sur des infrastructures de calcul et de stockage distribuées
  - Basés sur des technologies Open Source et interopérables
  - Ouverts à toutes les communautés scientifiques
  - Gratuits pour l'utilisateur
- France Grilles est aussi un réseau d'experts qui participent aux développements de ces technologies
- Les services de France Grilles s'intègrent à une offre européenne pour l'EOSC et participent à la construction d'une plate-forme nationale pour la Science Ouverte
- Son offre de services est complémentaire de celle des mésocentres et du HPC

# Atelier technique EOSC-France

- Strasbourg, 24-26 janvier 2023, IPHC
- Format hybride (présentiel – visio)
- Thèmes:
  - Rejoindre EOSC (catalogue et portail)
  - Les task forces EOSC et l'intérêt pour les centres de calcul
  - Authentification et autorisation, sécurité et aspects légaux pratiques, identifiants numériques
  - Interopérabilité des services et des infrastructures
  - Projets techniques fédérés autour du calcul et de la donnée

