



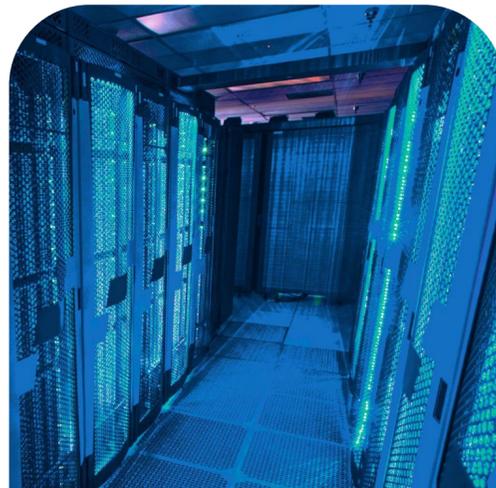
Centre de Calcul et de Données  
de l'Institut Pierre-Simon Laplace (ou "ESPRI")

Sophie Cloché, Guillaume Levavasseur, Karim Ramage



**01**

**ESPRI, l'essentiel**



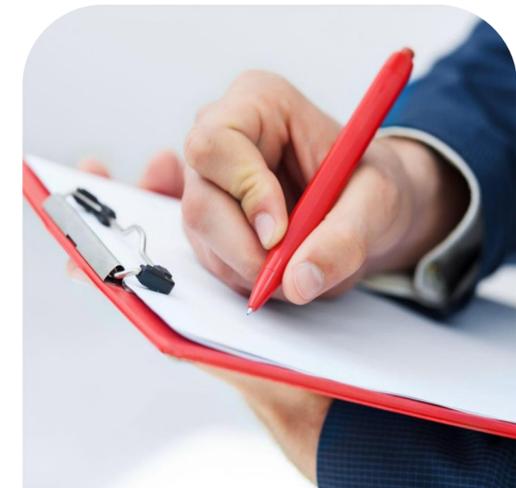
**02**

**Services autour du calcul**



**03**

**Services autour des données**



**04**

**Pour finir**



**01**

**ESPRI, l'essentiel**

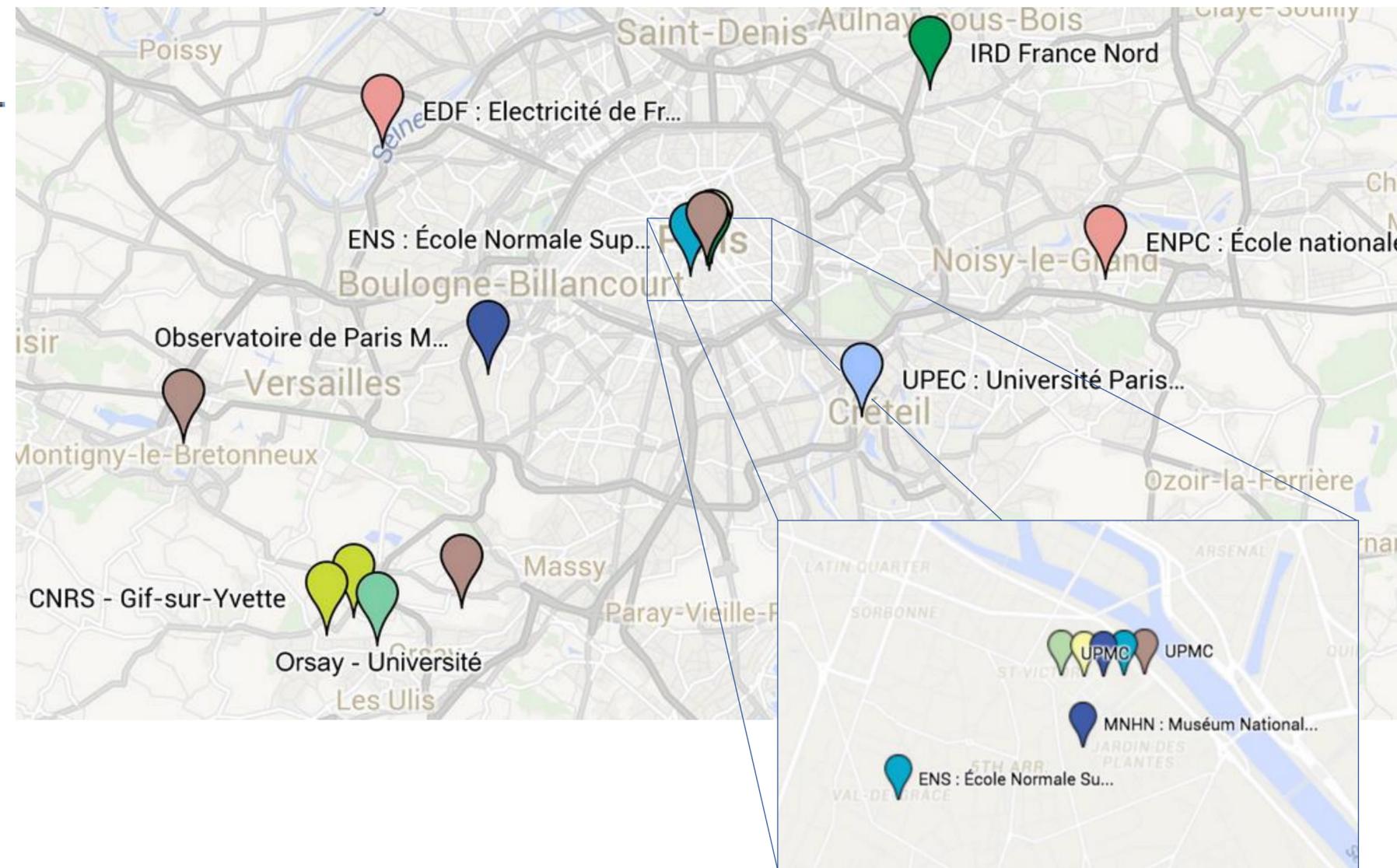


# Qu'est-ce que l'Institut Pierre-Simon Laplace (IPSL)?

L'IPSL travaille sur le climat de la terre et des planètes, ses interactions avec l'environnement et le changement climatique

## IPSL en Ile de France

-  Fédération IPSL
-  LMD
-  LOCEAN
-  CEREA
-  LATMOS
-  LSCE
-  GEOPS
-  Equipe TASQ
-  LISA
-  METIS
-  ENS Equipe S&R



~1500 personnes



# L'IPSL, 1 UAR , organe de coordination, d'animation et de services



**8 laboratoires,  
2 équipes associées**

**des structures fédératives**

**des services mutualisés  
dont ESPRI**



## ESPRI : Centre de Calcul et de Données de l'IPSL pour les sciences du Climat, regroupant les activités :

- ✓ infrastructures informatiques pour le calcul et le stockage de données scientifiques
- ✓ gestion des données de simulation et d'observation des sciences du Climat

Depuis plus de 20 ans, le centre de calcul et de données de l'IPSL fournit un Ensemble des Services Pour la Recherche à l'IPSL – **ESPRI**.





**IPSL**  
ICEO & ICMC & laboratoires

**Partenaires & tutelles**  
ClimatSuds (IRD)  
Dir. des services num. locales

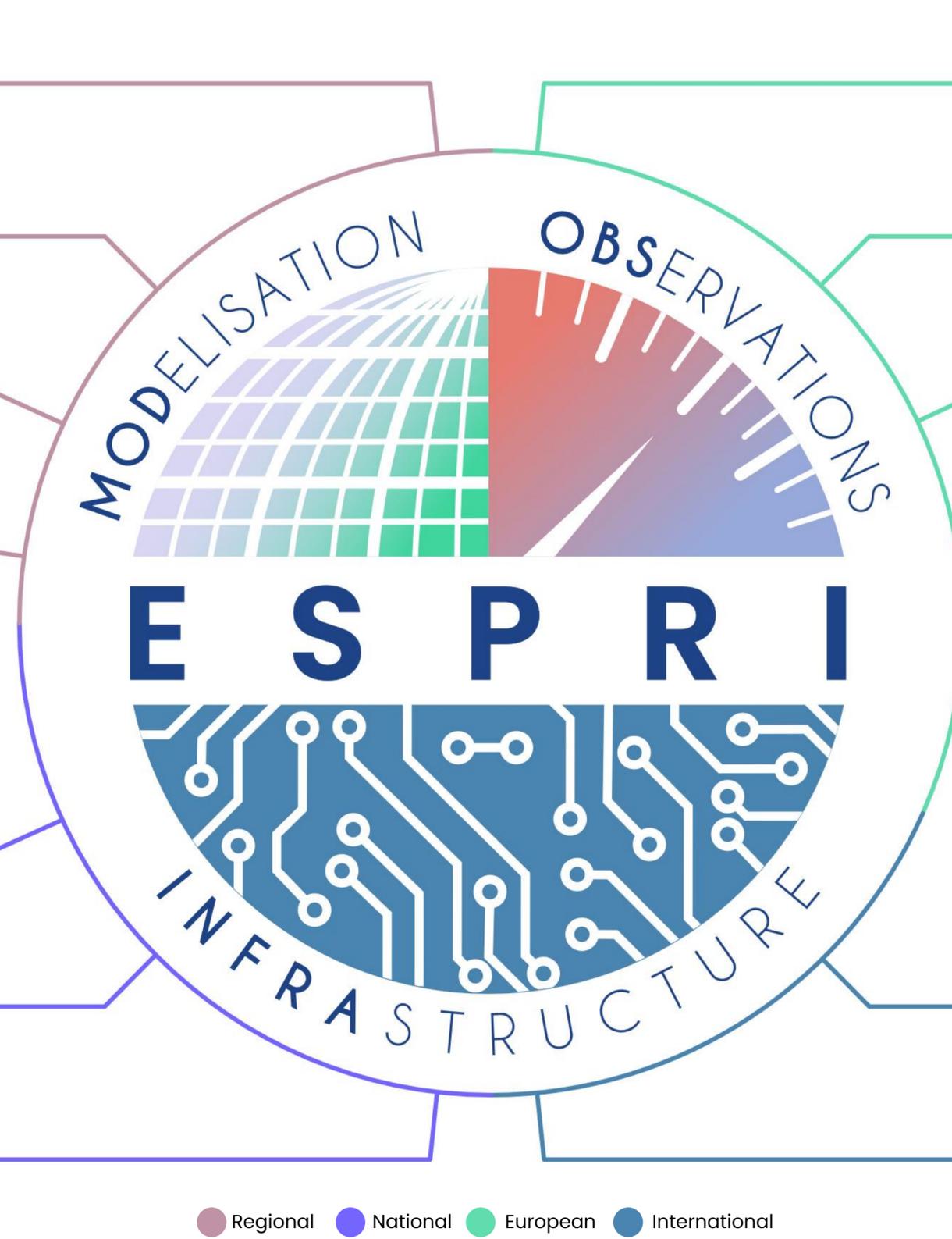
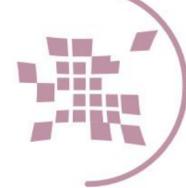
**OSUs franciliens**  
Ecce Terra  
OVSQ

**GENCI**  
Centres de calcul nationaux  
(IDRIS & TGCC)

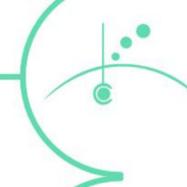
**Data Terra**  
C<sup>tr</sup>e de Donnée et Service AERIS  
Direction technique adjointe

**CLIMERI-France**  
Plateforme d'analyse nationale  
Direction technique adjointe

**GAIA-DATA & TRACCS**  
Centre ossature  
Pilotage de WP et projets ciblés



**ENES**  
Membre de la Data Task Force



**ACTRIS**  
Resp. Trace Gaz Remote Data Centre



**Copernicus**  
Fournit le Climate Data Store



**EOSC**  
« Provider » de l'European Open Science Cloud



**EUMETSAT**  
Représente CNRS au Climate Monitoring SAF



**CMIP & ESGF**  
Co-pilote de Task Teams  
Serveur de rang 1



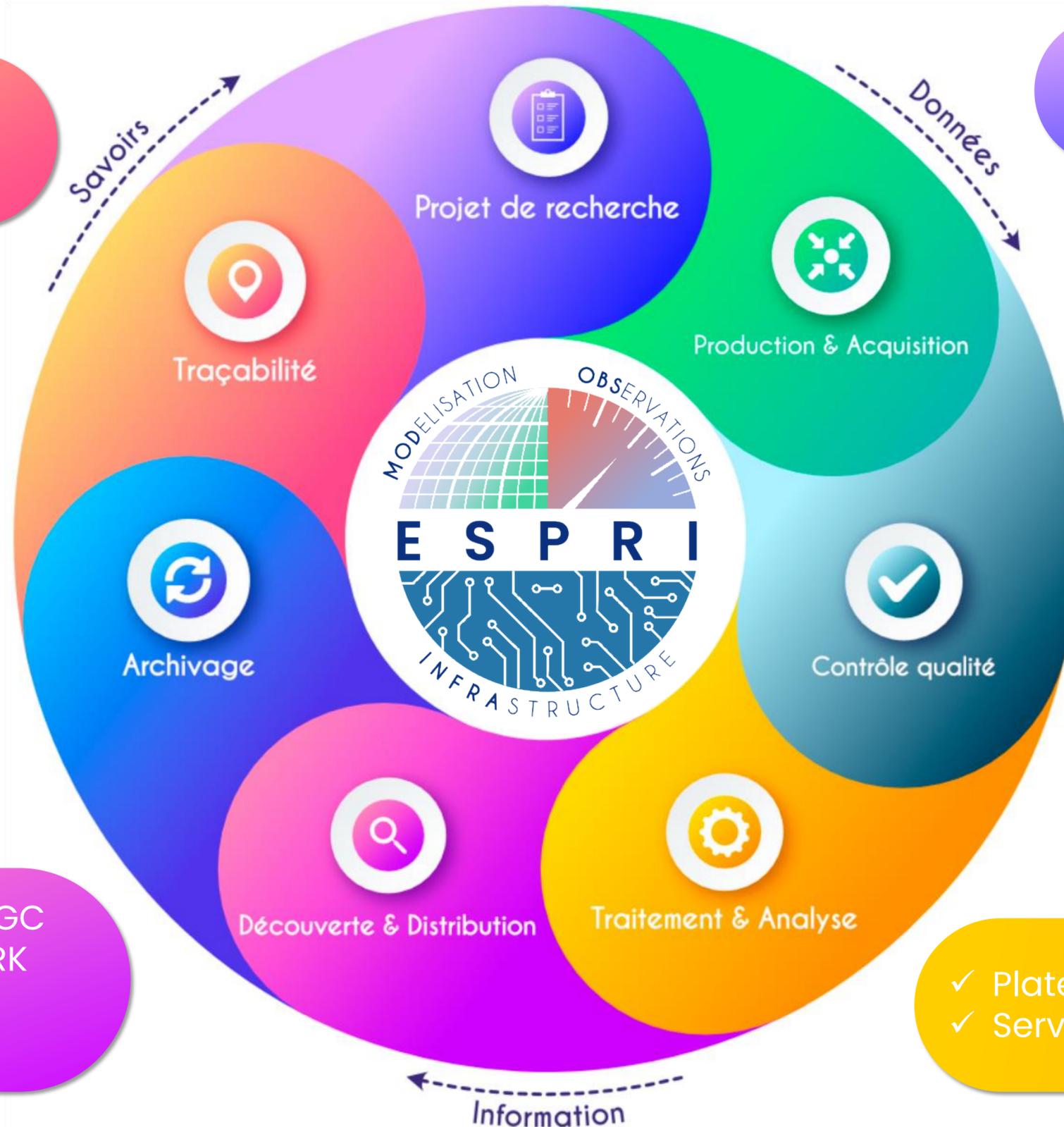
**Research Data Alliance**  
Certification CoreTrustSeal

● Regional ● National ● European ● International

- ✓ Documentation & errata
- ✓ Données référencées avec PID
- ✓ Service de DOI

- ✓ Stockage redondant
- ✓ Stratégie de réplication
- ✓ Sauvegarde long-terme

- ✓ Distribution en standards OGC
- ✓ Metacatalogue GEONETOWRK
- ✓ Publication ESGF
- ✓ Serveurs THREDDS



- ✓ Plan de gestion de données
- ✓ Support utilisateur

- ✓ Données interdisciplinaires
- ✓ Réseau haut débit dédié
- ✓ Accès transparent à 3Po
- ✓ Service à la demande

- ✓ Nomenclature & standards
- ✓ Vocabulaires contrôlés
- ✓ Versionnement
- ✓ Curation

- ✓ Plateforme d'analyse multi-modèles
- ✓ Services ESPRI



~20 ETP



## ESPRI-OBS

CDS AERIS  
Gestion des données du CEO/IPSL  
Traitement de données Applications web et systèmes d'informations  
Support opérationnel campagne



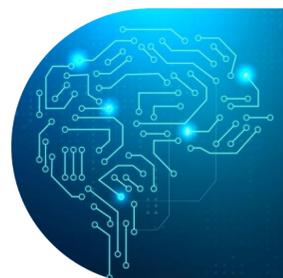
## ESPRI-MOD

Plateforme nationale d'analyses CLIMERI  
Gestion des données du CMC/IPSL  
Documentation et évaluation des modèles  
Réplication des simulations climatiques  
Applications de correction des biais



## ESPRI-INFRA

Infrastructure de calcul et stockage  
Infrastructure partagée et distribuée  
Réseau et sécurité  
Centres de calcul nationaux et régionaux  
Services informatiques généraux



## ESPRI-IA

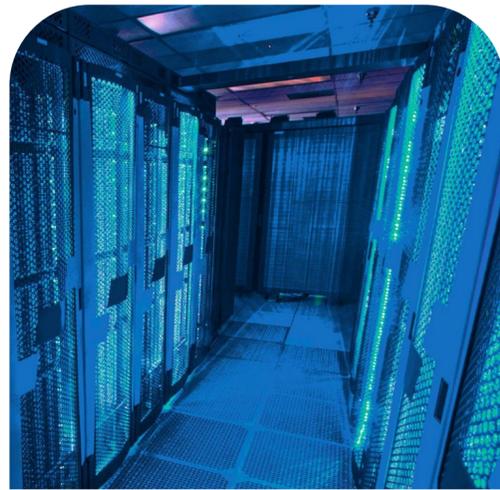
Ingénierie et support pour la communauté IA de l'IPSL



## E4C DataHub

Plateforme interdisciplinaire pour la transition énergétique

Autres services en lien avec ESPRI :



**02**

**Services autour du  
calcul**



## Géographiquement distribué sur 4 sites

- Sorbonne Université
- Ecole polytechnique
- Université Versailles St Quentin
- IDRIS

## Calcul

- 3 000 coeurs CPU - 13To de RAM
- 5376 GPU Tensor cores (12 GPU - 24 GB RAM/GPU)
- Interconnexion réseau Infiniband (100 Gbps / Faible latence)

## Stockage

- ~ 6,5 Po Dataspace + 2,2 Po Userspace (Lustre@IPSL)
- ~ 4 Po Dataspace (Scality@IDRIS)
- Par utilisateur : 32Go sur /home + 1To de workspace + 2To de scratch

## Environnement logiciel

- Compilateurs (GNU, Intel, etc.)
- Bibliothèques et logiciels scientifiques (netCDF4, CDO/NCO, etc.)
- Containerisation via Apptainer

## Accès

- SSH, SSHFS
- Jupyterhub

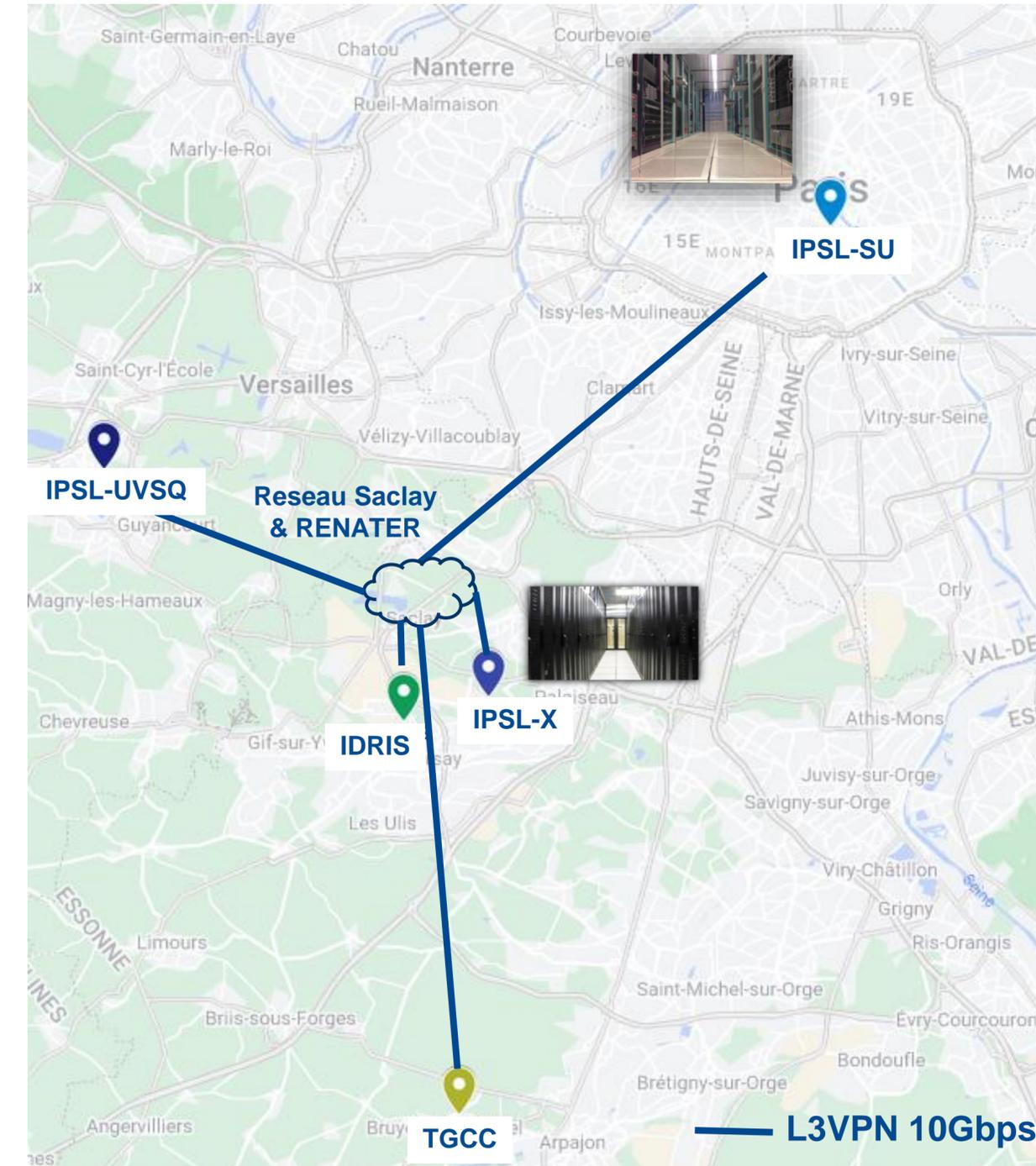
## Partage

- HTTPs, FTP, Thredds/OpenDAP, GridFTP
- DOI/PIDs

## Services

- Plateformes de virtualisation pour l'hébergement de services
- Plateforme d'hébergement Web (sites applicatifs, informatifs, wordpress, ...)

## Le Mésocentre de l'IPSL





**03**

**Services autour  
des données**

## ESPRI produit, collecte ou réplique des données qui peuvent servir à tous !

### Données sol

- Base de données de campagnes
- Mesures et produits du SIRTA
- Données du réseau ACTRIS
- Données Météo-France (RADOME, RADAR, etc.)

### Réanalyses & climatologies

- Produits ECMWF/C3S (analyses, Réanalyses ERA5)
- MERRA-2
- NCEP/NCEP2
- FCDR (AMSU, SSMI, GridSat)

### Données satellites

- Produits de niveau 1 à 4
- Ceintures géostationnaires, défilants

### Et bien d'autres données

- Données aéroportées
- Ballons
- Données paléoclimatiques
- Données spectroscopiques

### Sorties de modèles couplés globaux

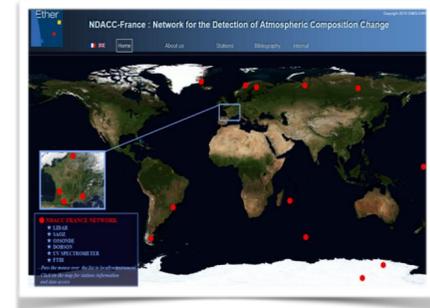
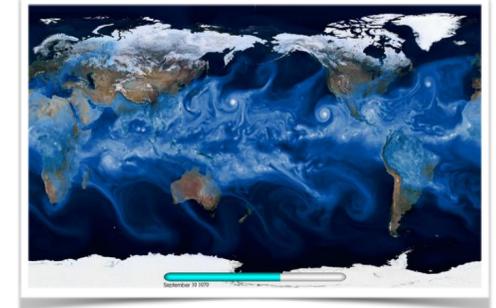
- Toute la production de l'IPSL-CM
- Répliques CMIP3/5/6
- Avec correction de biais

### Sorties de modèles régionaux

- Toute la production de l'IPSL (WRF)
- Répliques CORDEX (plusieurs domaines géo.)
- Avec correction de biais

### Données auxiliaires pour les modèles

- Observations regriddées (obs4MIPs)
- Forçages CMIP (input4MIPs)



## Ces bases de données sont entièrement gérées par ESPRI:

- ✓ Pérennisation des stockages,
- ✓ Curation des données chaudes/froides,
- ✓ Accès centralisé depuis le mésocentres ~ **8 Po**
- ✓ Complétion à la demande



## **Service de mise en production des codes scientifiques visant à automatiquement traiter les données (observations, simulations):**

- ✓ Opérationnalisation de chaînes de traitement scientifique
- ✓ Mise en production, utilisation d'outils de workflow
- ✓ Suivi de la production via des dashboards de surveillance

## **Service de post-traitements (correction des biais des simulations climatiques, etc.):**

- ✓ Mise à disposition des utilisateurs d'algorithmes de correction de biais
- ✓ Notebooks d'exemples
- ✓ Application de certaines corrections de biais à la demande

## **Environnements logiciels et d'analyse:**

- ✓ Outils d'analyse usuels en ligne de commande
- ✓ Environnements Python partagé (aussi accessible via JupyterHub) qui incluent notamment des librairies récentes et adaptées pour l'analyse de masses de données géospatiales
- ✓ Des librairies spécifiques et/ou "maison" pour l'analyse et l'évaluation des sorties de modèles climatiques



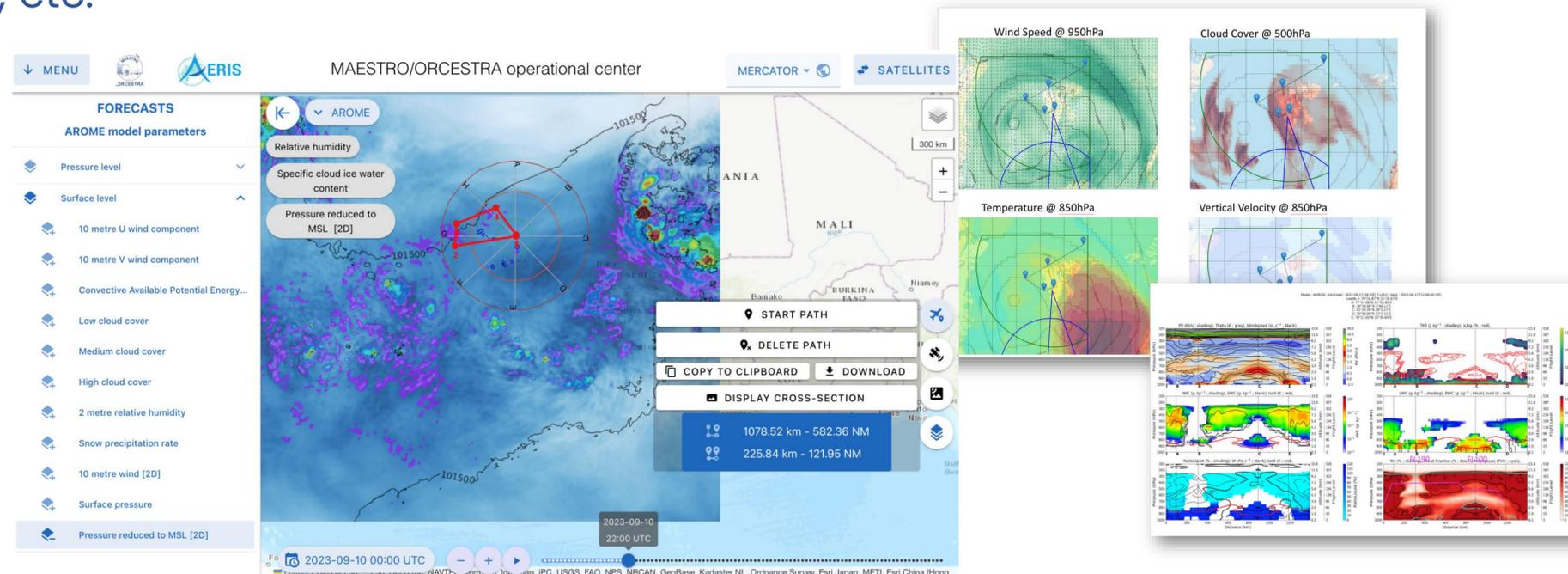
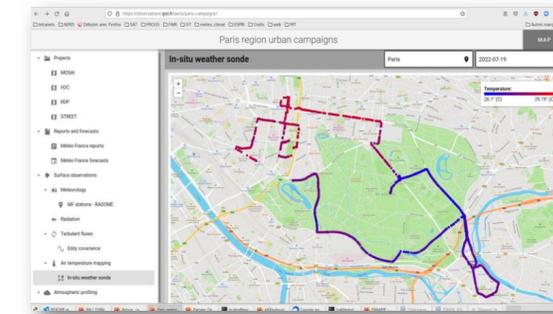
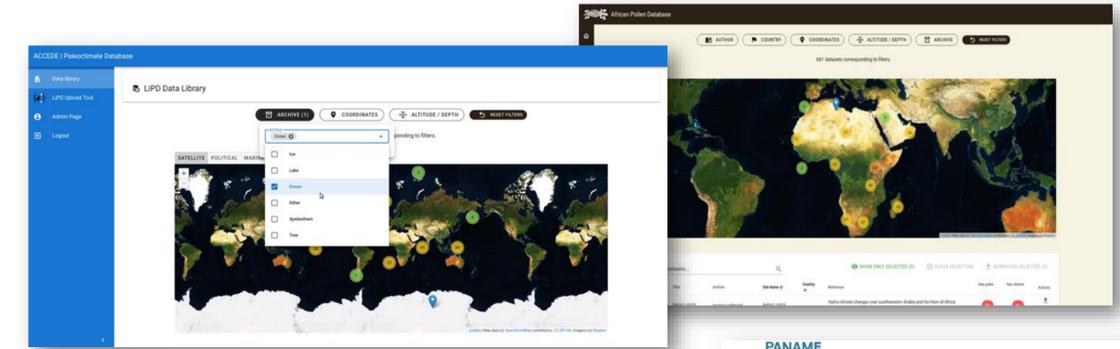
# Développement d'applications web de visualisation et d'accès avancé

## Outil de visualisation spécifique et thématique, exemples :

- ✓ BDD Paleo
- ✓ Portail climat urbain et Qualité de l'air

## Des outils de visualisation/calcul dédiés aux campagnes de mesures :

- ✓ Suivi en temps réel des modèles de prévision météo, images satellite, prévision de trajectoire ballon/satellite, calcul dynamique, etc
- ✓ Aide à la planification des vols avions, lâchers de ballons, contrôle de paramètres instrumentaux, etc.





## Au niveau national

- ✓ Gestion et distribution des simulations climatiques françaises
- ✓ Stockage des simulations internationales à l'IDRIS

## Au niveau européen

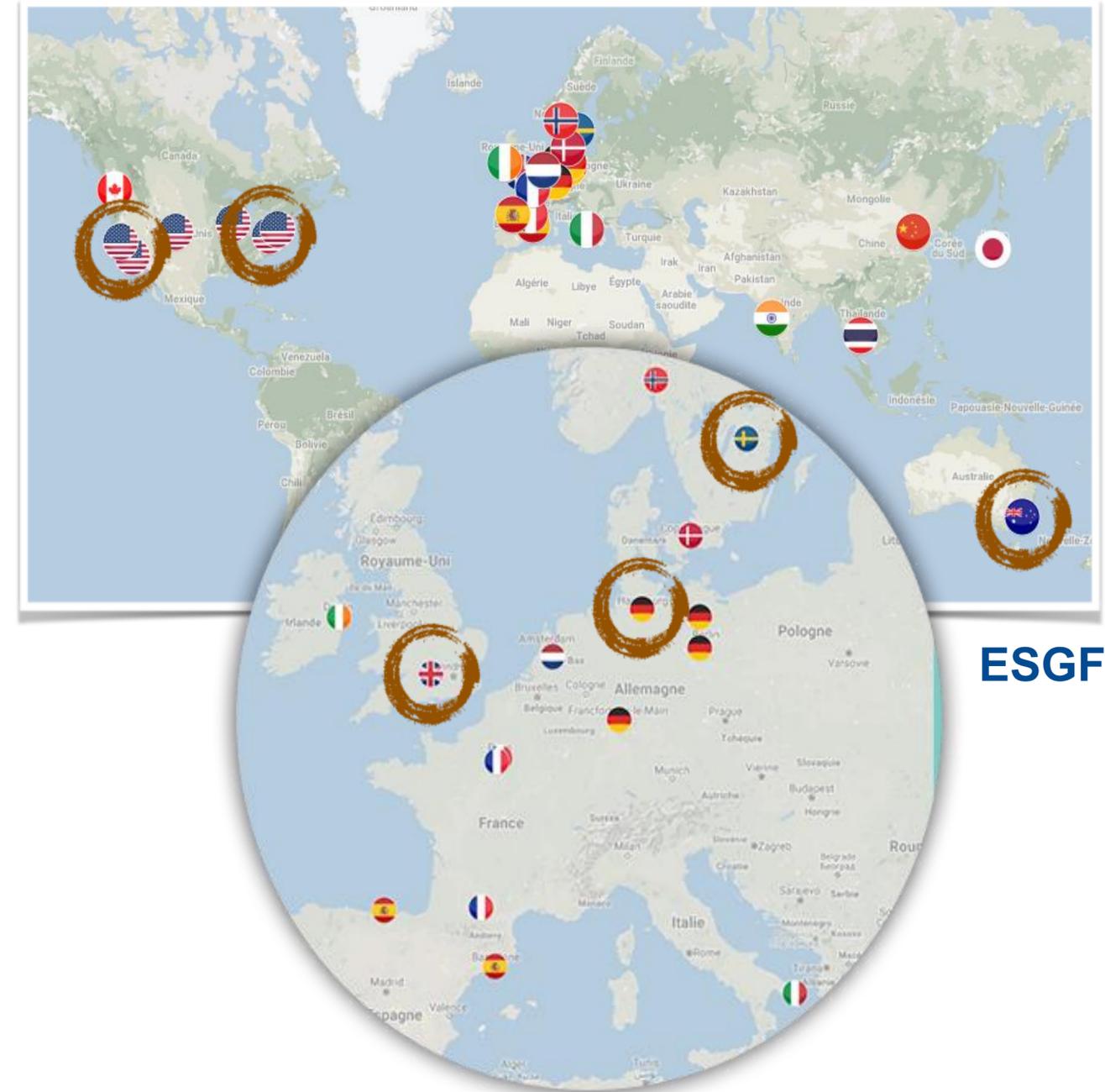
- ✓ Pilote de l'archive de données climatiques pour le service Climat du programme Européen COPERNICUS (C3S)

## Au niveau international

- ✓ Le plus grand contributeur du réseau international ESGF (> 1,4M de jeux de données  $\approx 1.7\text{Po} \approx 22\%.\text{ESGF}$ )

L'**Earth System Grid Federation (ESGF)** est un réseau de **serveurs décentralisé** et **fédéré** avec une collaboration **internationale**. C'est un système de **nœuds distribués** qui interagissent dynamiquement grâce à un paradigme Peer-To-Peer (P2P).

## Publication des simulations climatiques sur ESGF



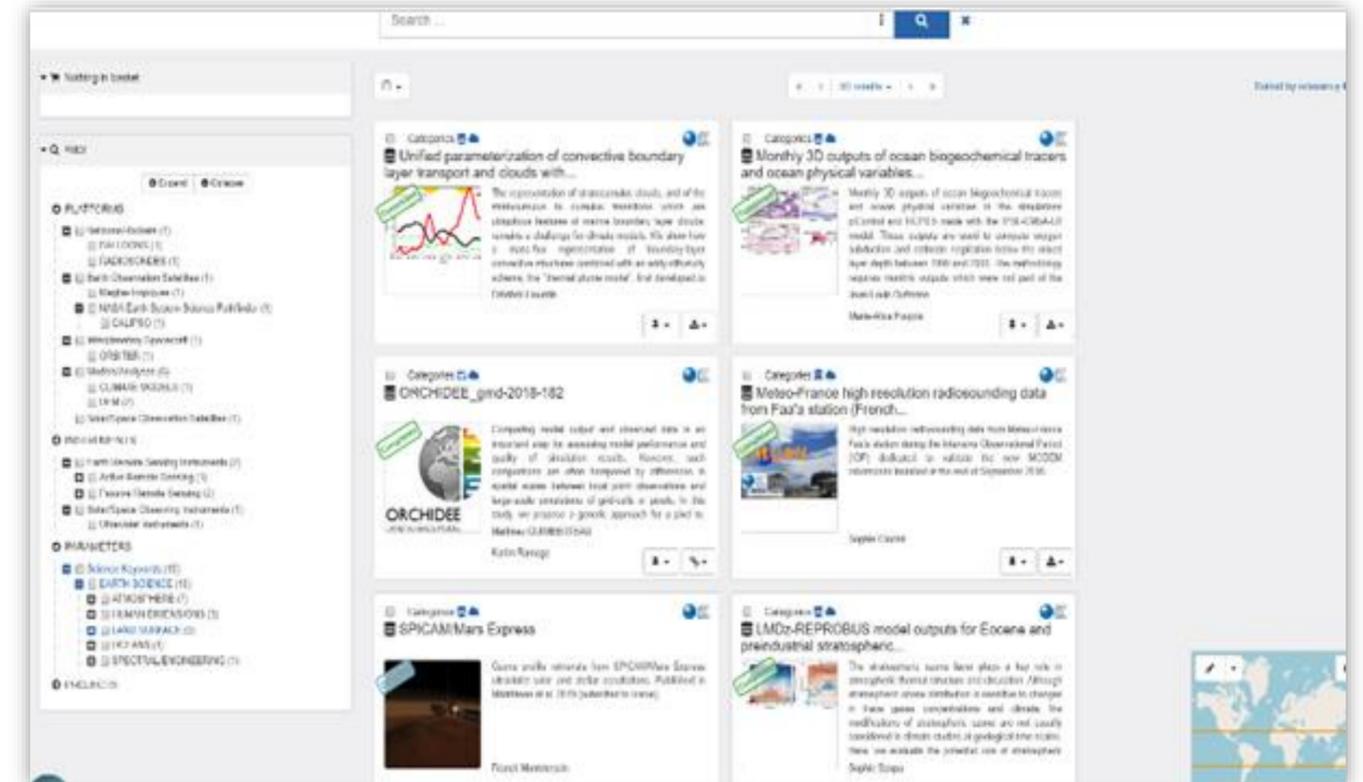
ESGF



## Catalogue (atmosphère, océan, planeto, etc.):

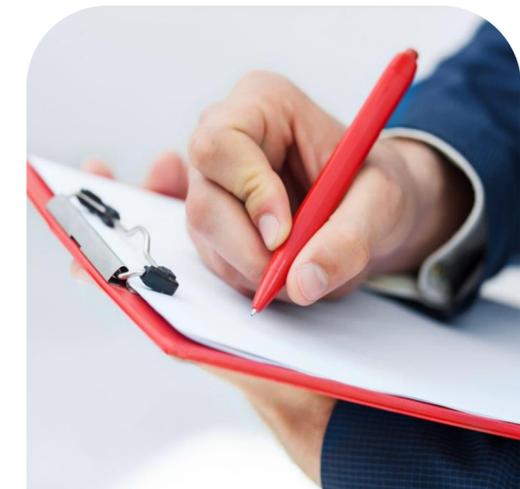
<https://data.ipsl.fr/catalog>

- ✓ Accès pérenne aux données et métadonnées
- ✓ Vient en support des demandes de DOI
- ✓ Accompagnement dans la rédaction de la fiche de métadonnées



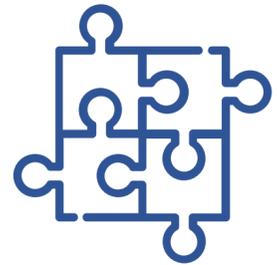
ESPRI est habilité auprès de l'INIST/CNRS pour générer des identifiants d'objets numériques (DOI) sur des ensembles de données ou des codes numériques, à la demande.





**04**

**Pour finir**



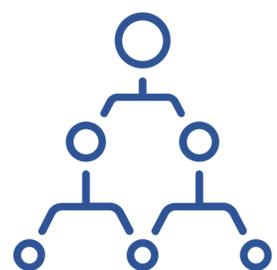
L'interdisciplinarité et l'hétérogénéité croissantes dans les jeux de données.

Les analyses au plus près des volumes de données toujours croissants.

Le croisement complexe des données et l'expérience utilisateur.

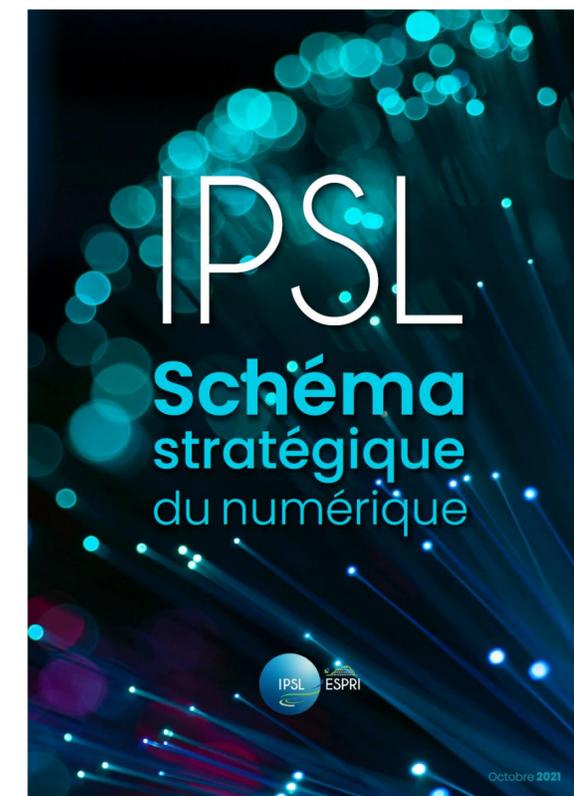
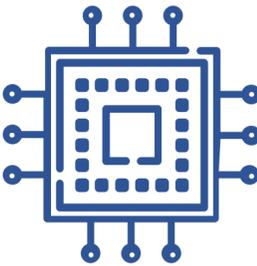
La montée en puissance des techniques de l'IA.

Le besoin de traçabilité et de transparence des analyses (FAIR).



Open Science

Technologies "Cloud"





MERCI – THANK YOU

Institut Pierre-Simon Laplace (IPSL)



## Services mutualisés pour le calcul et la gestion de données autour d'une infrastructure distribuée

### Au service :

- ✓ Des laboratoires et structures fédératives de l'IPSL
- ✓ Des EUR Climate Graduate School IPSL et Energy 4 Climate de l'Institut Polytechnique de Paris

### Composante :

- ✓ Des infrastructures de Recherche Nationales Data Terra et ClimERI
- ✓ Des Infrastructures Européennes/Internationales (ISENES, ESGF, Copernicus, ACTRIS, ..)

### Les services proposés:

- ✓ Moyens de calculs et de stockage distribués sur 4 sites (SU, X, UVSQ, IDRIS)
- ✓ Accès simple à un ensemble de données proches des moyens de calcul avec les outils d'analyse nécessaires et environnements logiciels adaptés
- ✓ Différents services "FAIR" autour des données

### Les données:

- ✓ Issues de la communauté scientifique IPSL
- ✓ En provenance d'autres centres de données ou communautés
- ✓ Données variées : observations in-situ, satellites, ré-analyses, simulations, ...



**1000 utilisateurs inscrits (150 nouveaux inscrits en 2024)**

**En moyenne annuelle :**

**300 utilisateurs ayant soumis des jobs sur les clusters du mésocentre :**

**1,6 millions de jobs soumis**

**12 millions d'heures de calcul CPU cumulées en 2024**

**1600 heures de TP numériques d'enseignement s'appuyant sur la plateforme (durée des TP x nombre d'étudiants)**

**Utilisateurs de la plateforme de calcul**

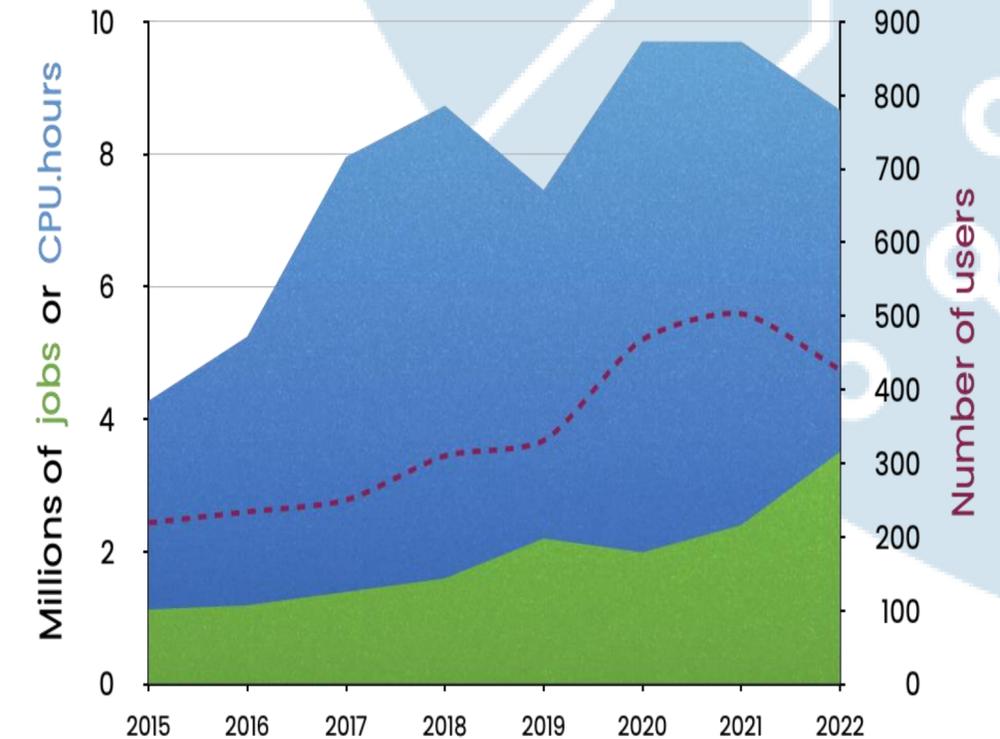
**65% laboratoires de l'IPSL**

**25% laboratoires partenaires français (OMP, CNRM, IGE,...)**

**10% partenaires étrangers**

**99% académiques**

**1% privé**



**Statistiques 2022**

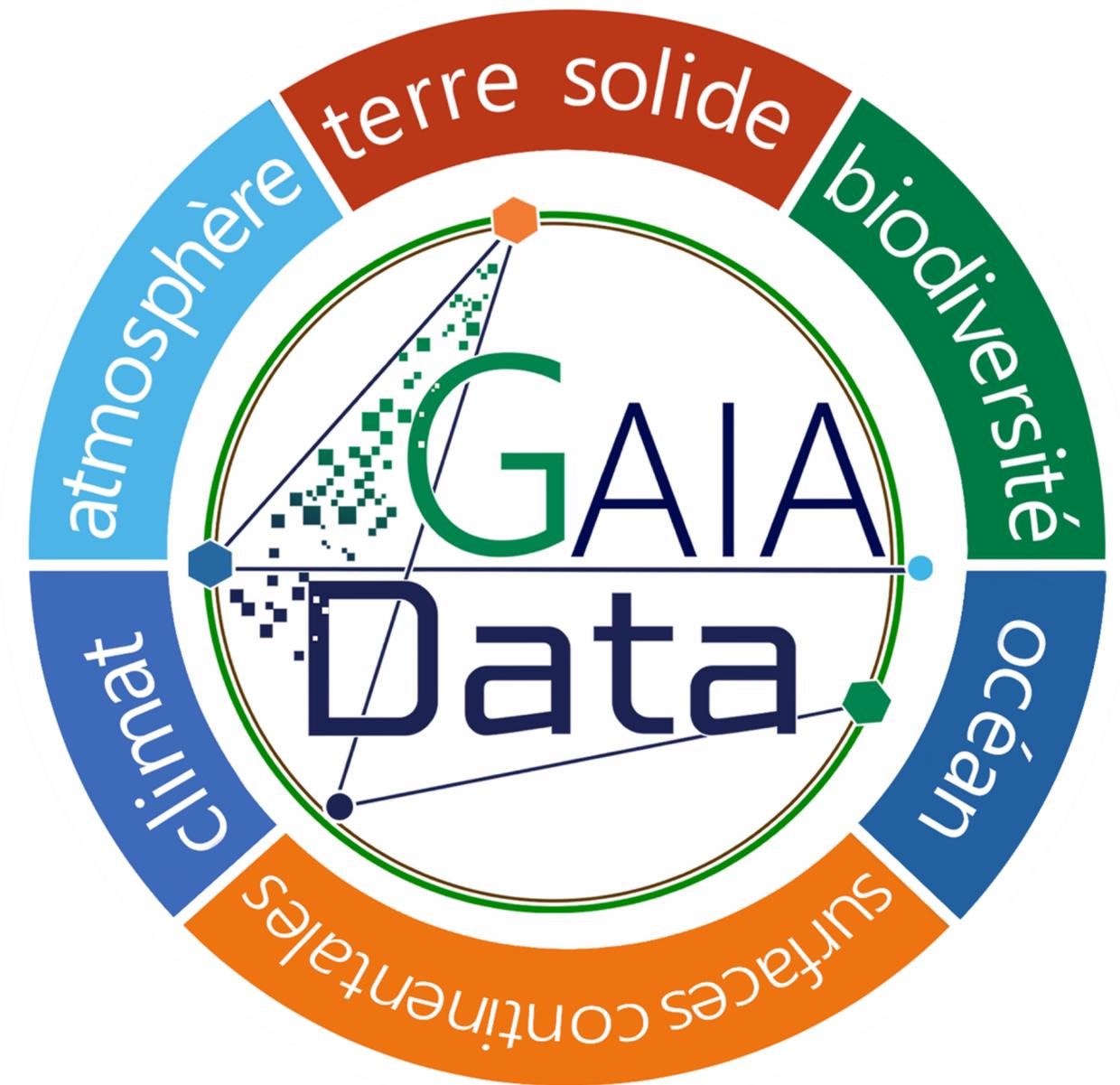


**GAIA-DATA** est un **projet PIA** (Programme d'Investissement pour l'Avenir) soutenu par **3 infrastructures de recherche nationales** (incluant 21 partenaires nationaux).

GAIA-DATA vise à **développer** et à **mettre en œuvre** une **infrastructure intégrée de données FAIR** et de **services distribués** pour **l'observation**, la **modélisation** et la **compréhension du système terrestre**, de la **biodiversité** et de **l'environnement**.



**DATA  
TERRA**



# Engagement d'ESPRI dans le projet GAIA-DATA



Cette infrastructure distribuée s'appuiera sur :

## 30 centres de données et de services,

Une dynamique régionale de IRs,

Une rationalisation de l'efficacité énergétique et une meilleure utilisation des ressources humaines expertes,

Un **continuum de services** (incluant les centres HPC nationaux, EuroHPC, cloud commercial),

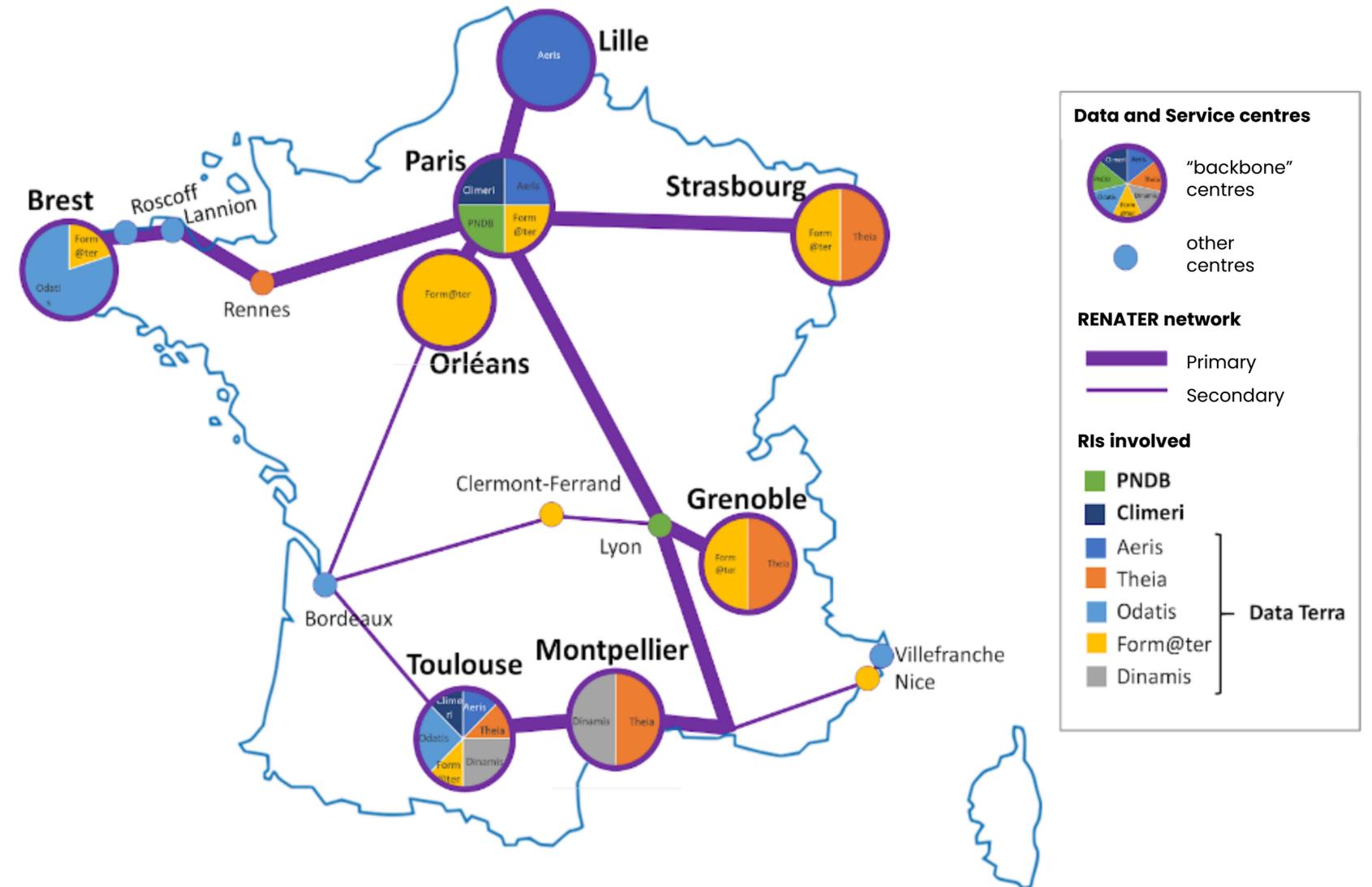
Un **traitement au plus près des données.**

Une telle infrastructure favorise les échanges de :

**Données entre sites et entre disciplines** grâce au déploiement d'une grille de données et de lacs de données avec transfert automatique de grands ensembles de données d'un centre à l'autre,

**Des traitements interopérables entre les 8 sites principaux et avec l'extérieur** en reliant les centres HPC en France et les clouds commerciaux.

Développé en lien avec les projets nationaux d'infrastructure numérique MesoNet, FITS et Clusster



## Soutenu par :

Une "ossature" de 8 centres

Un réseau dédié à haut débit et sécurisé



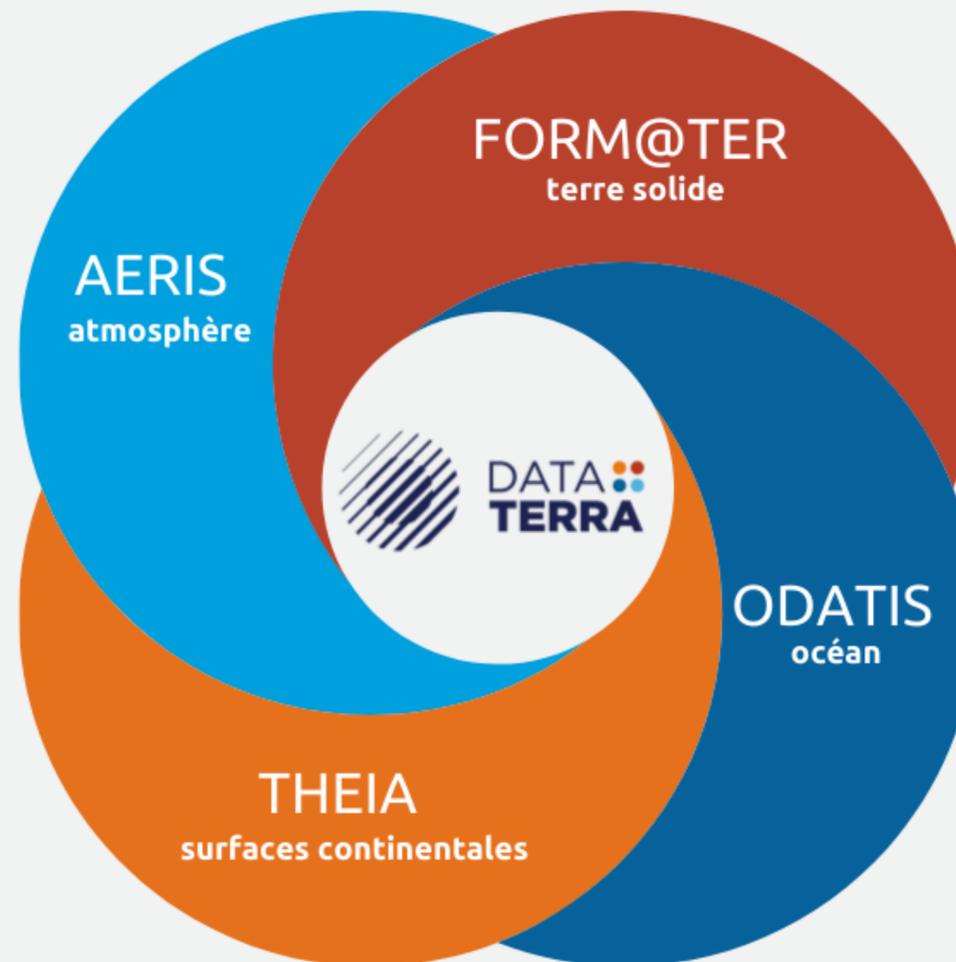
## L'IR DATA TERRA ses pôles de données & dispositifs transversaux

Cette IR est fondée sur quatre pôles de données correspondant à chacun des grands compartiments du système Terre :

- **THEIA pour les données surfaces continentales** (agriculture, forêts, biodiversité ...)
- **AERIS pour les données atmosphère** (gaz, aérosols, nuages ...)
- **ODATIS pour les données océan** (niveau moyen des mers, hydrologie, risques littoraux...)
- **ForM@Ter pour les données terre solide** (volcanologie, érosion des sols, sismologie ...)

Elle regroupe aussi deux dispositifs transversaux :

- **DINAMIS** (Pour accéder aux données spatiales haute résolution)
- **INTER-PÔLES** (Fédérer et animer des communautés d'experts en données)



Dispositifs transversaux



**RAMAGE, K. : Directeur technique adjoint**

Participation/pilotage différents groupes de travail transverses AERIS et DATA TERRA.

**ESPRI-Obs est un des 4 Centre de Données et Services (CDS) d'AERIS.**





**Support opérationnel et BDD campagne** : PANAME, RALi-Thinice, MAGIC; EUREC4A, STRATEOLE2.

**Gestion/traitement des données satellite** :

Archive nationale: IASI, GOME2, GOSAT

Archive primaire: BDD multi-satellites (nuages, précipitation, cadre GEWEX)



**Gestion/traitement des données sols** (Data center pour ACTRIS-FR) : GRUAN, E-Profile, lidar Température, O3, vent, H2O, UVVIS (saoz), données sites MF, données réseau GPS

**Traitement/homogénéisation des séries d'observations long-terme** : Re-Obs

**Banque de données**: GEISA, IUPAC

**Projets Européens**

Data center pour différents data streams ACTRIS-RI (Eurochamp, GRES)

HEMERA (BDD ballon Europe)

ENVIR-FAIR (Fairisation des données environnementales des RI).

**Catalogue AERIS** : <https://www.aeris-data.fr/catalogue/>

# Engagement d'ESPRI dans les infrastructures de modélisation du climat

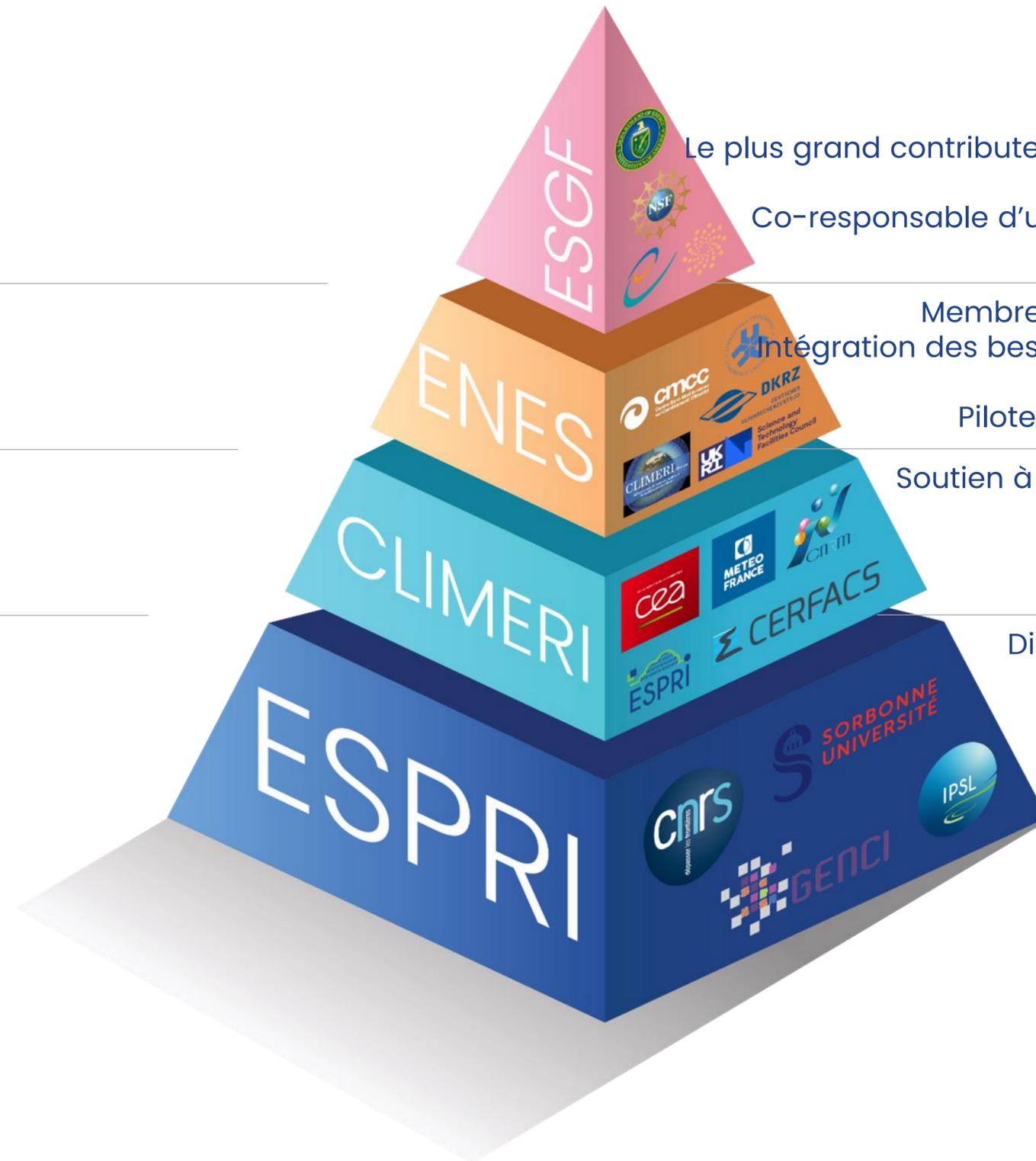


International

European

National (France)

Regional (Ile-de-France)



Le plus grand contributeur (> 1,4M de jeux de données  $\approx 1.7\text{Po} \approx 22\%.\text{ESGF}$ )  
Serveur ESGF de niveau 1  
Co-responsable d'un groupe de travail du WCRP Infrastructure Panel  
Pilote principal d'ES-DOC

Membre de la colonne vertébrale de l'infrastructure ENES  
Intégration des besoins nationaux dans les projets à grande échelle  
Maintenir et consolider ESGF  
Pilote de l'archive de données climatiques pour le C3S

Soutien à la communauté française (GAIA-DATA, TRACCS)  
Coordination avec les partenaires HPC  
Réplication et base de donnée multi-modèles  
Réponse aux services climatiques

Distribution des simulations climatiques françaises  
Environnements logiciels et d'analyse  
Post-traitements (correction des biais, etc.)  
Documentation et traçabilité des simulations