

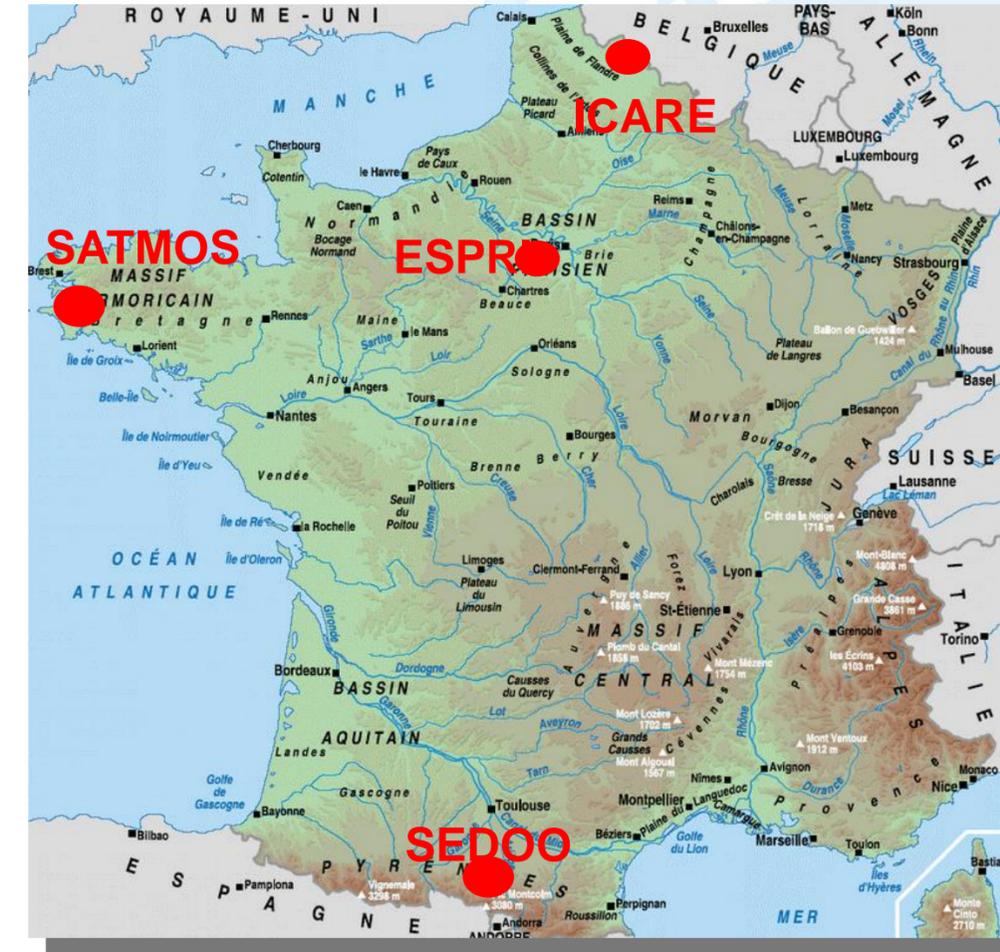
# INFRASTRUCTURE AERIS RENCONTRE IN<sub>2</sub>P<sub>3</sub>

10/04/2019

# Infrastructure distribuée sur 4 CDS

Une infrastructure distribuée et spécialisée entre les 4 CDS :

- Systèmes de stockage haute capacité, haute performance, sécurisés (ICARE/ESPRI/CMS)
  - 7,5 Po de stockage : Systèmes de fichiers GPFS et Lustre
  - Pour le stockage des données miroirs
  - Pour la production
  - Pour les espaces de travail des utilisateurs
- Moyens de calcul pour la production et pour les utilisateurs (ICARE / ESPRI)
  - 4500 coeurs de calcul
- Systèmes d'acquisition de données temps réel (SATMOS / ICARE / ESPRI)
  - Système de réception par antennes satellites (CMS Lannion, EumetCast)
  - Acquisition via serveurs/soft dédiés (ex. : Gosat à ESPRI)
  - Canaux d'acquisition privilégiés (Ex. : NOAA à ICARE)
- Système de préservation des données - sauvegarde et Archivage (tous les CDS)
- Infrastructure Réseau pour le calcul et pour la distribution des données :
  - Réseaux faibles latences – large bande passante en interne de chaque CDS
  - Accès internet via Renater de 1 Gbps à 10 Gbps
- Systèmes de virtualisation et de micro-services (ICARE/ESPRI/SEDOO)
  - Facilite le déploiement et la résilience des applications (Portails Web, services de distributions de données)
  - Hébergement de projets



# AERIS – Perspectives d'évolution



**Vers une infrastructure distribuée mais plus intégrée entre les CDS :**

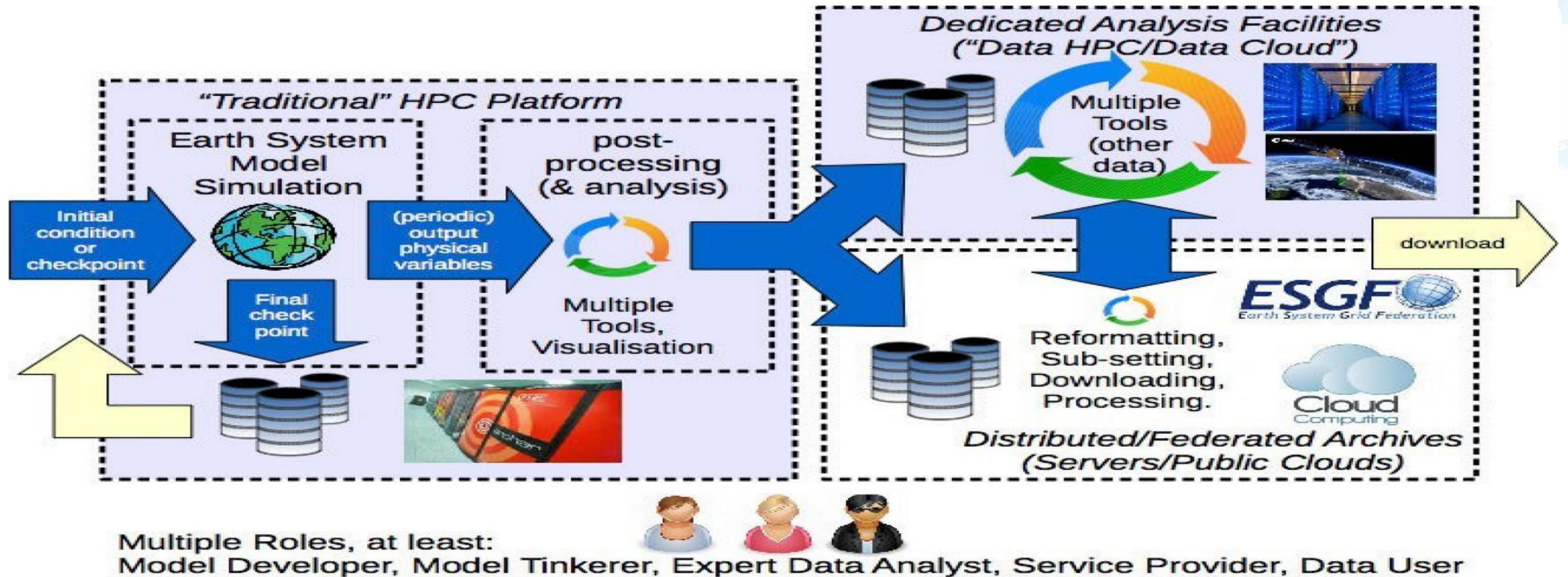
⇒ Rapprocher les données du calcul et/ou le calcul des données

⇒ Maintien de l'implantation locale des infrastructures des CDS

- Rapprochement avec les mésocentres/datacentres régionaux (Infranum)
- Interconnexion des infrastructures et des applications (ex. : intégration / déploiement continu)
- Interconnexion haut débit (> 10Gbps) dédiées entre les CDS
- Montage croisés des systèmes de fichiers / Système de fichiers distribués (Cephfs, Scality, ...)
  - Partage des données sans « copie »
  - Redondance des données multisites
  - Systèmes de stockage objet (ICARE et ESPRI) pour nouvelles applications (fouille de données)
- Système d'orchestration de containers multi-sites
  - distribution des applications « atomiques » au plus près des données
  - scalabilité des applications de traitement

**En cohérence avec les évolutions d'infrastructure prévues pour l'IR Système Terre**

## Multiple type of storage & data interaction



# Intra ESPRI pour AERIS et ClimERI : Partage de Systèmes de Fichiers

- **Objectif :**

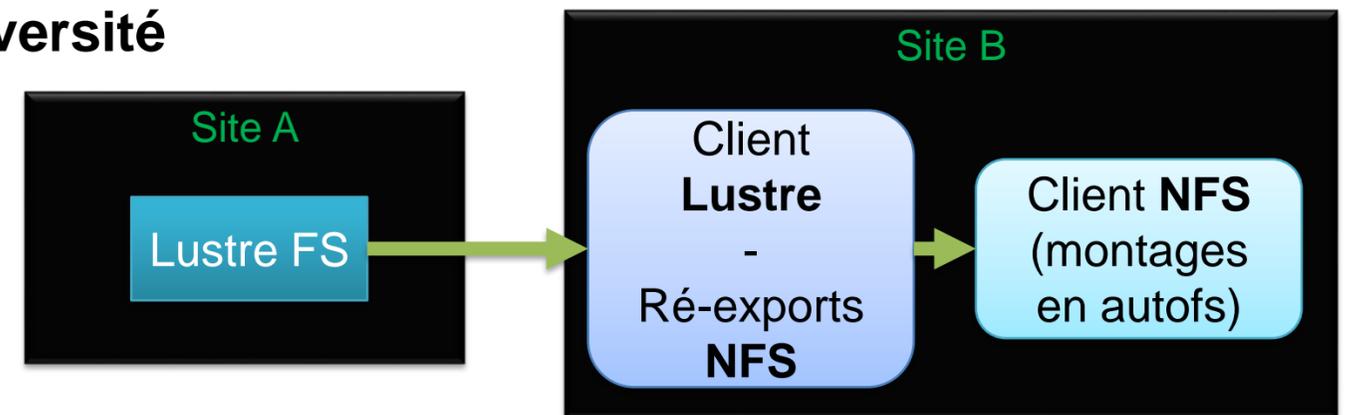
- Permettre le traitement de données d'un site A sur les ressources de calcul d'un site B
- Favoriser la synergie de données distribuées en évitant la copie

- **Contraintes :**

- La bande passante réseau
- La tolérance aux pannes (coupures réseau, limiter l'impact d'une panne d'un site sur le second)
- Les droits d'accès (partage des répertoires utilisateurs)

- **Montages croisés entre les Sites Polytechnique et Sorbonne Université**

- 2 systèmes de fichiers Lustre indépendants
- Visibilité croisée des FS entre les deux sites du mésocentre ESPRI
- Montages des FS Lustre/NFS via autofs en Read-Only
- Nécessite un annuaire commun pour partage UID/GID
- Mise en oeuvre simple, performant (3 / 4Gbps), mais réseau dédié
- Cas d'utilisation :
  - Sauvegarde des données du SIRTAs
  - Etudes croisées IASI/ECMWF



# Infra ESPRI pour ClimERI : Stockage sur les CC Genci / Traitement et diffusion via ESPRI

## • Montages des Systèmes de Fichiers du TGCC/IDRIS sur le mésocentre ESPRI :

- Cadre de ClimERI, mais démonstrateur pour une solution compatible InfraNum
- Données produites au TGCC (> 10 Po) et hébergées à l'IDRIS (4 Po) visibles sur la plateforme d'analyse IPSL pour la distribution et le post-traitement
- Export des FS en SSHFS (TGCC) et SAMBA (IDRIS) vers deux serveurs dédiés
  - à IPSL-X et IPSL-SU
- Réexport Samba des FS vers l'ensemble des serveurs IPSL
  - Mise en oeuvre simple, performances limitées par la bande passante et SSH
  - (2/3 Gbps pour SSH, 3 /4 Gbps pour Samba)
  - Réseau dédié : L3VPN
    - discussions Renater/IDRIS/TGCC/Paris-Saclay/IPSL
- Traitement et distribution des données via Plateforme d'analyse ESPRI :
  - Environnement logiciel adapté à la communauté climat (ouverture internationale dans le cadre d'IS-ENES)
  - Outils de distribution spécifiques ESGf (solution alliant contraintes des CC Nationaux – sécurité - et besoins de la communauté thématique)

