

Données et calcul : quelles pratiques et quels besoins pour la communauté RESIF ?

La compréhension du système Terre et de son évolution nécessite d'analyser des quantités croissantes de données de provenance multiple.

L'objectif de l'atelier est (i) de participer au **recensement des pratiques** de la communauté RESIF dans ces analyses (accès aux données multi-composantes, moyens de calcul utilisés pour les traitements massifs) et (ii) d'**identifier des besoins** spécifiques à prendre en compte dans la construction de la future infrastructure nationale de recherche Data Terra.

La communauté des utilisateurs des données RESIF sont invités à présenter **des cas d'étude** pour lesquels une évolution des infrastructures de données et de calcul actuelles sont souhaitables.

Restitution

- **Participation** : 83 % des inscrits (5) !
- **Contexte** : **production-distribution** de données (centre de données+produits, **Résif**) / **analyse et traitement** (infra de calcul+services, **Data Terra** ou/et google ?).

Etat de pratique sismologie

- Mode opératoire dominant : download données pour traitement en « local » (laptop, labo, mésocentre).
- Quelques expérimentations réussies sur des offres de calcul communautaires (grille FG/EGI, vo earth science), limitées à des faibles quantité de données.
- Quelques initiatives de rapprocher calcul et (grandes quantités de) données (CIMENT, S-CAPAD, HPC UNISTRA), usage local, principalement montage disque.
- Pas de retour des analyses de données composites (sismo + GPS + météo + satellites + ...).
- Pas de retour des applications IA.
- Peu (ou pas?) de chaînes de traitement/environnement de calcul « communautaires ».

Limites / Besoins

- Inégalité de l'offre de calcul (expertise) locale.
- Limitations du modèle actuel pour grands volumes de données.
- Rapprocher les calculs des données : comment faire le pas vers un environnement de calcul dématérialisé (cloud) et communautaire ?

Cas d'étude

- Volumes de données issus des Large N arrays et des DAS : bons cobayes pour tester une solution communautaire.
- Principe : pas de distribution sans réduction, visualisation & traitement interactif à distance.
- Exemple PANGEO (communauté océano) : plateforme communautaire python, structure données spécifique, format stockage objet, ressources calcul élastiques, implémentations HPC et cloud privés (google).
- Pas encore d'état de pratique, développement from scratch d'une solution communautaire.