



RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

anr
agence nationale
de la recherche



DATA
TERRA



Présentation générale de l'infrastructure technique intégrée de données et de services

Richard Moreno (CNES)
Coordinateur technique du projet GAIA DATA



KICK-OFF 12 AVRIL 2022



SOMMAIRE



01

OBJECTIFS ET ENJEUX

02

SOLUTION TECHNIQUE

03

MISE EN OEUVRE



01

OBJECTIFS ET ENJEUX



Porté par 3 Infrastructures de Recherche



OBJECTIF

Développer et mettre en œuvre **une infrastructure/plate-forme intégrée de données FAIR** et de services distribués pour **l'observation, la modélisation et la compréhension du Système Terre, de la Biodiversité et de l'Environnement**

- **sur l'ensemble du cycle de la donnée**, de son **acquisition** (spatiale, sols, in-situ) jusqu'à ses **multi-usages** (qualification/validation, stockage, accès, traitements/croisements de données multi-sources/extraction de connaissances, produits/services)
- **pour la communauté scientifique** contribuant à la connaissance du système Terre, de la biodiversité et de l'environnement ; **acteurs publics et privés**

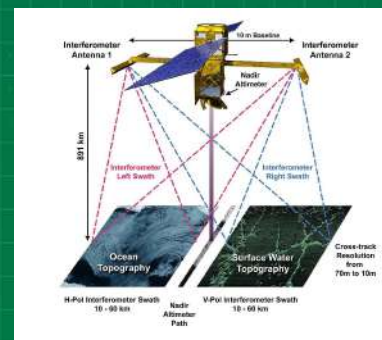
Enjeux sur les données

Explosion des flux de données et de la **diversité des sources de données** du système Terre.

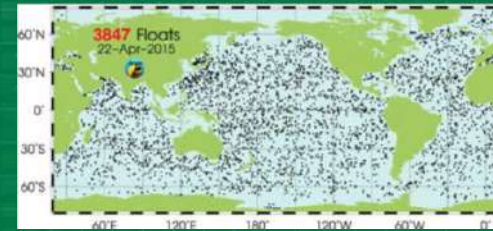
- Données spatiales, in-situ, aéroportées, simulations numériques, réanalyses combinant modélisation et observations,
- Très gros volumes et même pour les données non spatiales
- **Ensemble des compartiments** du système Terre, **biodiversité, modèles climatiques**

Le développement de nouvelles recherches repose sur notre capacité à **gérer l'ensemble du cycle de la donnée** :

- De l'acquisition à la mise à disposition des données à travers des **services FAIR** en passant par la gestion de l'information sur la donnée et son archivage long terme
- En fournissant des moyens de **facilement croiser ces données**



Satellite SWOT CNES/NASA



ARGO



Ballons



Biodiversité



Modèle climatique

Findable Magnifying glass icon	Persistent Identifiers (PID)s ID icon	Rich metadata Document icon	Indexed data repositories Database icon	PIDs in metadata ID icon
Accessible Hand pointing icon	Standard communications protocol Network icon	Open, free protocol Open lock icon	Authentication, where necessary User icon	Metadata in every domain Infinity icon
Interoperable Gears icon	Vocabularies Network icon	Vocabularies are FAIR Gears icon	Linked metadata Network icon	
Reusable Recycling icon	Metadata have multiple attributes Network icon	Usage license Open lock icon	Provenance Network icon	Community standards Group icon

Source Australian National Data Service

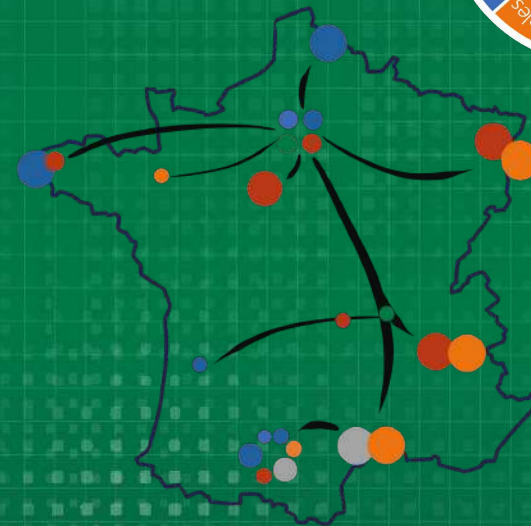
Contexte : distribué géographiquement

Dans une **infrastructure distribuée** :

- 30 Centres de Données et de services
- Dynamique des IR **régionale**
- Besoin de **rationaliser les infrastructures** pour améliorer l'efficacité énergétique et mieux utiliser les ressources humaines expertes
- Sur un continuum d'infrastructures (y compris centres HPC nationaux, EuroHPC, cloud commerciaux)
- Reliés par un **réseau performant**
- En rapprochant autant que possible les traitements au plus proche de la donnée

Favoriser les échanges :

- De données (entre les sites et transdisciplinaire)
- De traitements (interopérabilité entre les sites et avec l'extérieur)



8 sites principaux
30 sites existants



02

La solution technique

INFRASTRUCTURE GAIA DATA

Grille de données et de services : 8 principaux centres en réseau



Projets Equipex+ ou PIA4 infra

- FITS
- MesoNet
- Clusster

Projets Equipex+ ou PEPR thématiques

- Obs4Clim
- TerraForma
- Marmor
- OneWater

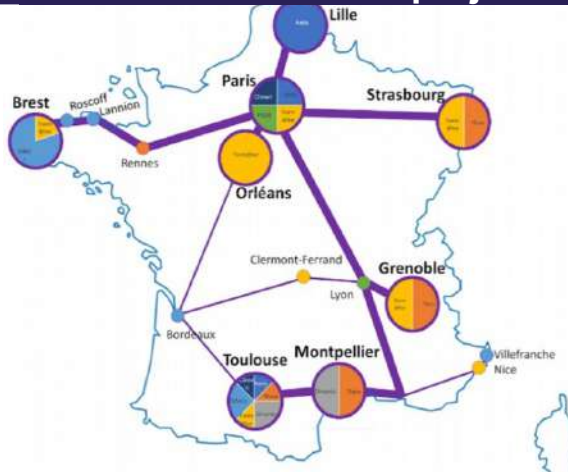
Projets H2020 – Horizon Europe

- IS-ENES
- PHIDIAS
- EOSC-Pillar
- FAIR EASE
- FAIR IMPACT

Projets CPER en région



En relation avec des projets connexes



Intégré dans le paysage international / Européen

- Mise en place d'un réseau dédié haut-débit et sécurisé
- Déploiement d'une grille de données et de datalakes pour permettre un accès distant aux données et le transfert rapide et automatique de grands ensembles de données d'un centre vers un autre
- Interopérabilité des traitements entre les 8 centres de Gaia Data, avec les centres HPC en France et avec les clouds commerciaux (DIAS – OVH – Orange, ...)



Services découverte, Accès et Gestion de données

Catalogue (métadonnées, vocabulaires, ontologies), systèmes évolués d'accès et de recherche

Consultation et accès aux données via **web services interopérables** (INSPIRE, Opensearch, STAC, ...)

Services avancés de visualisation

Accompagnement des communautés pour la **FAIRisation**



Services transversaux pour faciliter les travaux transdisciplinaires

Grille de données,

Cloud GAIA Data,

Portail connaissances,

Authentification unique

Support utilisateurs & formation – animation communautés

Support aux campagnes

Analysis Ready Data
Datacubes, ...



Earth Analytics Lab Exploration de la donnée, Bac à sable, Virtual Research Environment

Capacité à se connecter directement sur les centres

Traitements à la demande

Notebook/PANGEO/STAC

Low code / NoCode : Galaxy-E, FG/VIP



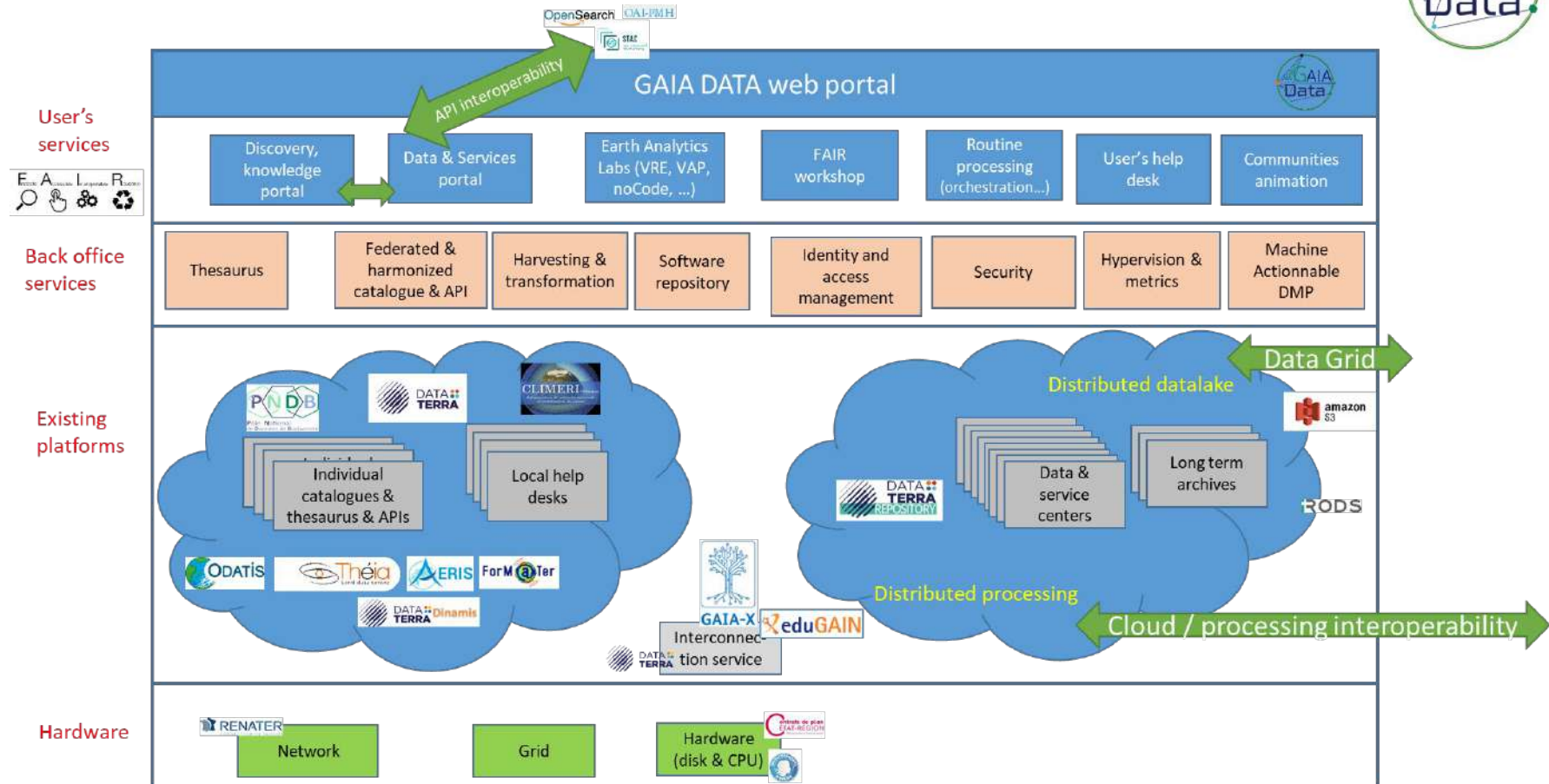
Services de production réguliers

Optimisation des traitements (outils orchestration) et formats de données (Zarr, CoG, ...)

Supporté sur un continuum d'infrastructures partagées

Les services GAIA Data : Réutilisation, développement, intégration

~PUZZLE : briques de base réutilisables pour des projets, les IRs ou les pôles



Portail de découverte des données : un point d'accès unique aux données

Source d'inspiration / Modèle earthdata.nasa.gov

Les vocabulaires et ontologies pour aider à **filtrer, trier, réduire le nombre de jeux de données** à un nombre raisonnable

Discipline/Océan : Pointeur vers page sur le thème « Océan » Informations sur les missions et données associées

- Pointeur vers WebMap ne présentant que les données relatives à l'Océan
- Ensemble de services liés à l'Océan – ex animation de la communauté si nécessaire

WebMap permettant d'accéder à l'ensemble des données GAIA DATA



WebMap utilisateur configurée pour n'afficher qu'un ensemble de jeux de données configuré par l'utilisateur

Pointeurs vers les portails des pôles pour les utilisateurs de ces domaines ou vers des projets ciblés

Possibilité d'habiller la WebMap avec l'identité visuelle du projet, ...

Cette WebMap peut être configurée (vue STAC) pour n'afficher qu'un sous-ensemble du catalogue GAIA DATA (sous-projet, vue utilisateur)

Possibilité d'ajouter cette collection à une Webmap personnalisée



Ici les critères génériques du STEP 1

Une recherche en 2 temps :
- STEP 1 : pour trouver des jeux de données à partir des critères/facettes ci-dessus
- STEP 2 : pour chaque jeu de données, des critères spécifiques (ex couverture nuageuse pour S2 et polarisation pour S1, ...)

The screenshot displays the GAIA Data web interface. At the top, there is a navigation bar with the GAIA Data logo on the left and menu items: ACCUEIL, USE CASE, ACTUALITES, A PROPOS, RESSOURCES ET FAQ, and MON PROFIL. Below the navigation bar, there are two tabs: 'Recherche par critères' (selected) and 'Recherche sémantique'. The 'Recherche par critères' section contains several filter fields: 'Période d'acquisition' with 'Début' and 'Fin' date pickers, 'Collection' (highlighted with a green arrow), 'Plateforme', 'Instrument', 'Niveau de traitement', 'Type de produit', 'Mode du capteur', and 'Fraîcheur du produit'. To the right of the filters, it shows '50 RÉSULTATS PARMIS 3200 COLLECTIONS' and 'Showing 20 of 254 matching collections'. Three search results are visible, each with a thumbnail, title, description, and metadata. The first result is 'VIIRS (S-NPP) | Band 375 m Active Fire Product NRT (Vector data)'. The second is 'MetOp-B ASCAT Level 2 25.0km Ocean Surface Wind Vectors in Full Orbit Swath'. The third is 'MetOp-C ASCAT Level 2 25.0km Ocean Surface Wind Vectors in Full Orbit Swath'. On the right side of the interface, there is a map view showing a satellite image of the Middle East region. A red arrow points from the text 'Possibilité d'ajouter cette collection à une Webmap personnalisée' to a green plus icon in the search results. A purple arrow points from the text 'Cette WebMap peut être configurée...' to the GAIA Data logo. A green arrow points from the text 'Possibilité d'habiller la WebMap...' to the GAIA Data logo. A black arrow points from the text 'Ici les critères génériques du STEP 1' to the 'Collection' filter dropdown.

Step 2 : recherche sur chaque jeu de données avec des critères spécifiques

Des critères spécifiques au jeu de données pour raffiner la recherche : Ici Cloud cover pour des données MODIS

The screenshot shows the EarthData Search web application. On the left, there is a sidebar with search filters. The 'Cloud Cover' filter is highlighted, with a red arrow pointing to it. The filter is set to 'Find granules by cloud cover percentage'. Below this, there are input fields for 'Minimum' (Example: 10) and 'Maximum' (Example: 50). Other filters include 'Orbit Number' and 'Equatorial Crossing Longitude'. The main area displays search results for 'MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061'. It shows a grid of 12 granules, each with a thumbnail image, a title, and a table of 'START' and 'END' dates. A yellow arrow points from the text 'Possible de ne pas passer par un panier' to the granule grid. At the bottom, there is a timeline for the month of April 2022.

L'utilisateur peut sélectionner des fichiers dans son panier pour y associer des services : VRE, traitements, téléchargement.

Possible de ne pas passer par un panier : la structuration en fichier est masquée

Ces données et services sont distribués géographiquement

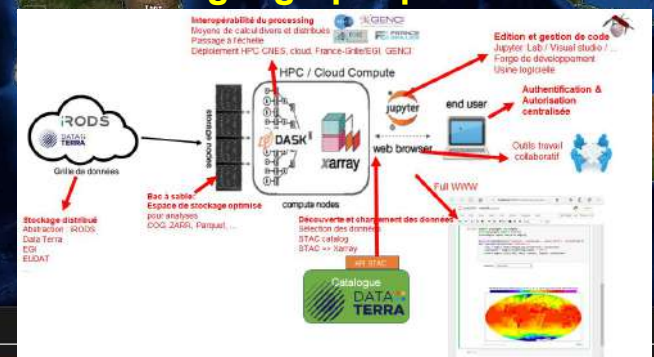


Illustration : Page projet ou thème dans un cadre GAIA DATA



ACCUEIL

USE CASE

ACTUALITES

A PROPOS

RESSOURCES ET FAQ

MON PROFIL

PROJET

LOGO

BIENVENUE SUR LA PLATEFORME D'EXPLOITATION DES PRODUITS PROJET EST LA PLATEFORME FRANÇAISE QUI OFFRE UN ACCÈS AUX PRODUITS DES SATELLITES SENTINEL S1A ET S1B, S2A ET S2B, S3A ET S3B DU PROGRAMME DE L'UNION EUROPÉENNE POUR L'OBSERVATION ET LA SURVEILLANCE DE LA TERRE COPERNICUS. LE PROGRAMME COPERNICUS EST COORDONNÉ PAR LA COMMISSION EUROPÉENNE. LES PRODUITS SATELLITES SONT GÉNÉRÉS PAR L'ESA QUI ASSURE LA COORDINATION TECHNIQUE DE LA COMPOSANTE SPATIALE, ET SONT REDISTRIBUÉS GRATUITEMENT PAR LA PLATEFORME PEPS.

DERINIÈRES ACQUISITIONS



CONTACT

BIENVENUE@PROJET.COM



COLLECTIONS

EARTH ANALYTICS LAB

WebMap permettant d'accéder à l'ensemble des données d'un projet GAIA DATA

Possibilité de fournir des informations générales sur le Projet mené dans le cadre GAIA Data

Ainsi que des news ! Des exemples d'application, ...





Services sur les données

METADONNÉES

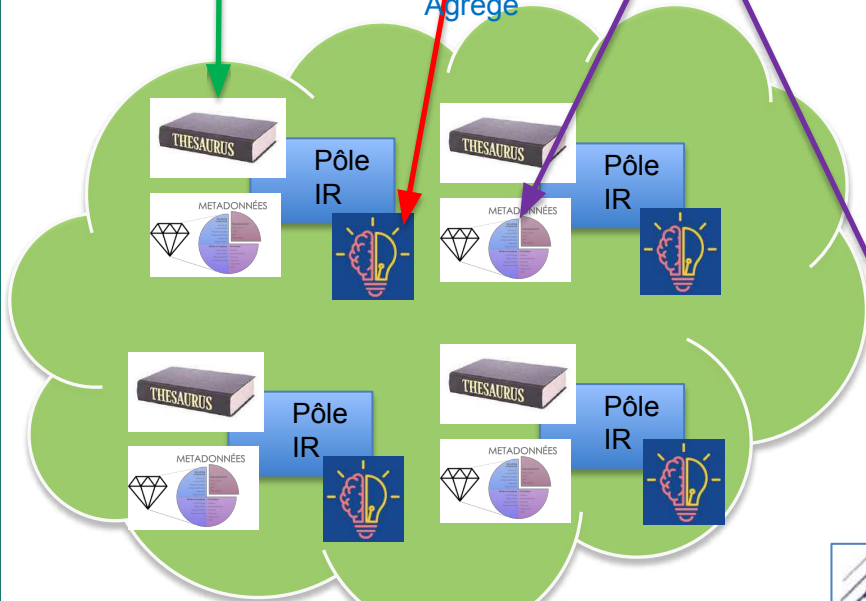
Moissonne

Agrège

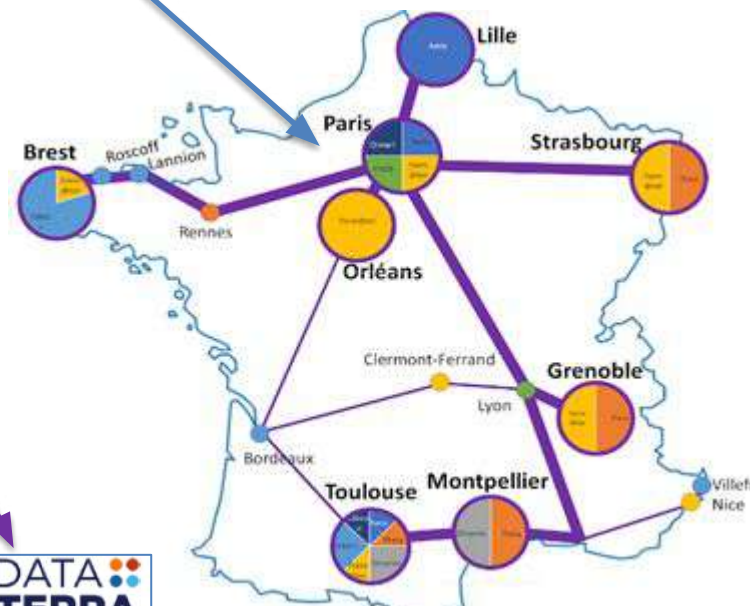
Met en correspondance

Redirige

Les services GAIA DATA s'appuient sur ceux proposés par les différents sites ou CDS



Pôles et IRs



8 Sites / 30 CDS



Les différentes couches



Exploitation layer

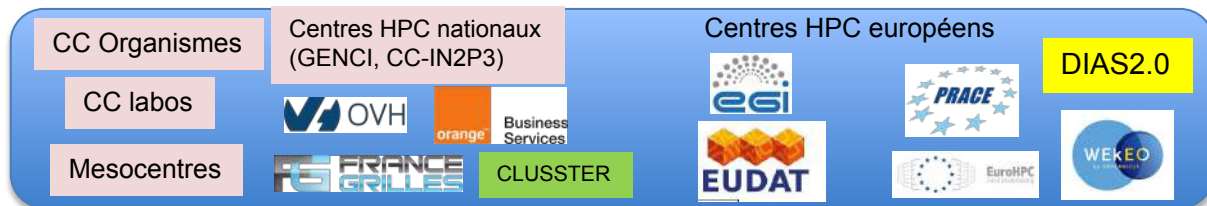


Nombre à limiter pour ne pas perdre les utilisateurs

Platform Service layer



Resource tier layer Data/ICT

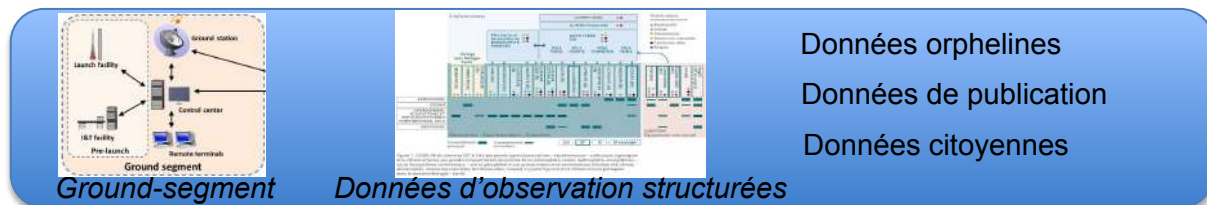


Enjeu de rationaliser les infrastructures de calcul



GAIA-X

Data Generation Layer



Données orphelines
Données de publication
Données citoyennes

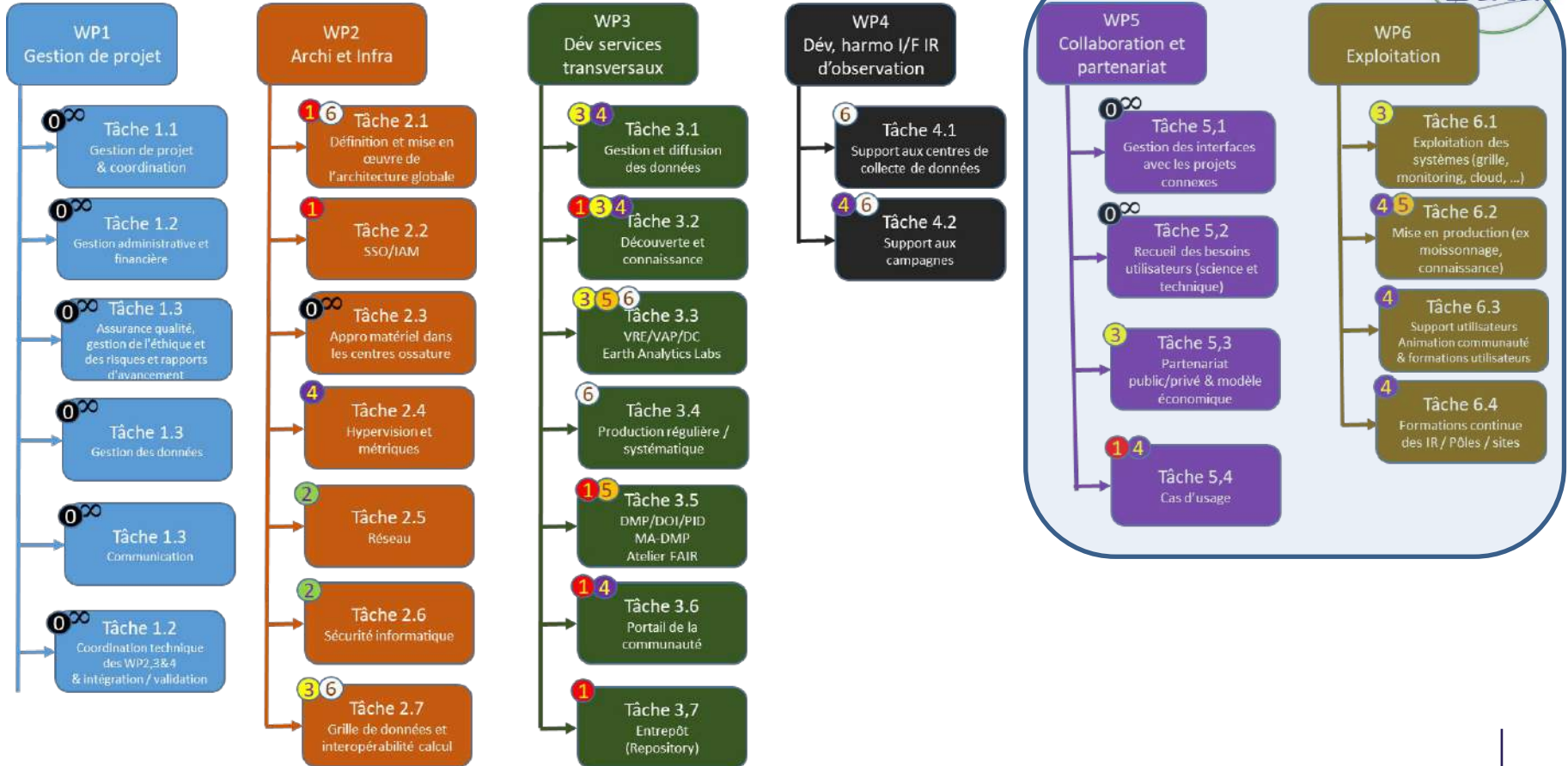
L'entrepôt et WP4 pour améliorer la situation



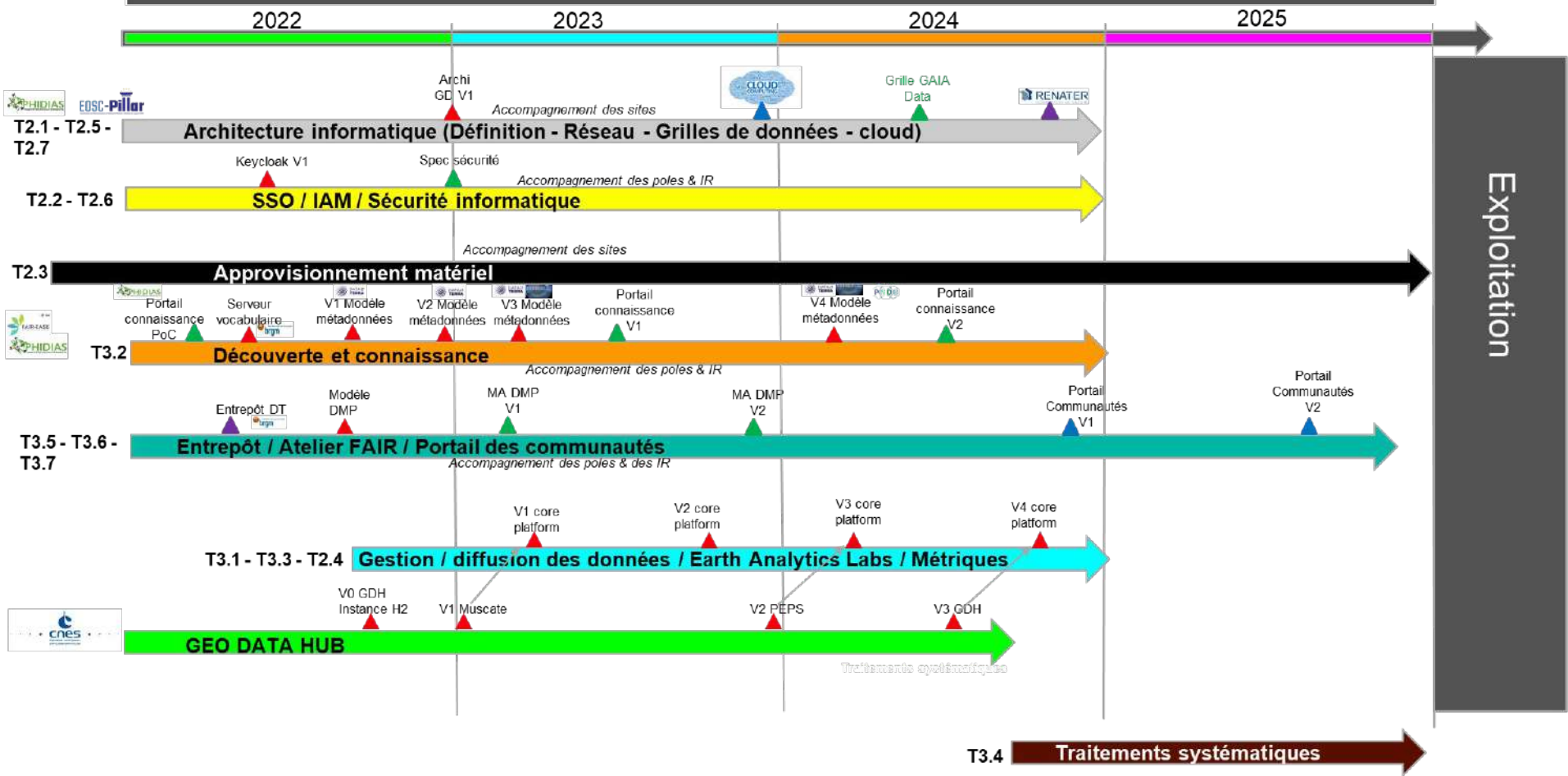
03

MISE EN ŒUVRE

Organigramme des tâches GAIA DATA



MISE EN PLACE





RÉPUBLIQUE
FRANÇAISE
*Liberté
Égalité
Fraternité*

anr[®]
agence nationale
de la recherche



DATA
TERRA



*Ce travail a bénéficié d'une aide de l'Etat
gérée par l'Agence Nationale de la Recherche
au titre du programme Investissements
d'Avenir Equipex+.*



contact@gaia-data.org

www.gaia-data.org

