



DATA  
TERRA

# Data Terra & GAIA DATA

Un assistant de découverte des données pour accompagner les utilisateurs de données de la recherche – 05 Juin 2024

Joël SUDRE, Coor. Tech. GAIA DATA, Data Terra



# DATA TERRA, une e-Infrastructure de Recherche dédiée au système Terre / Environnement



Développer un **dispositif global d'accès et de traitement de données, produits et services** pour adresser des enjeux scientifiques et des défis sociétaux interdisciplinaires

- Producteurs de données d'observation multi-sources
- 30 Centres de Données et de Services
- 32 Centres d'Expertise scientifiques
- 250 FTE / 500 scientifiques, ingénieurs et techniciens

- Coordination/articulation avec les OSU, SNO
- En France
- Outre-Mer
- Etranger

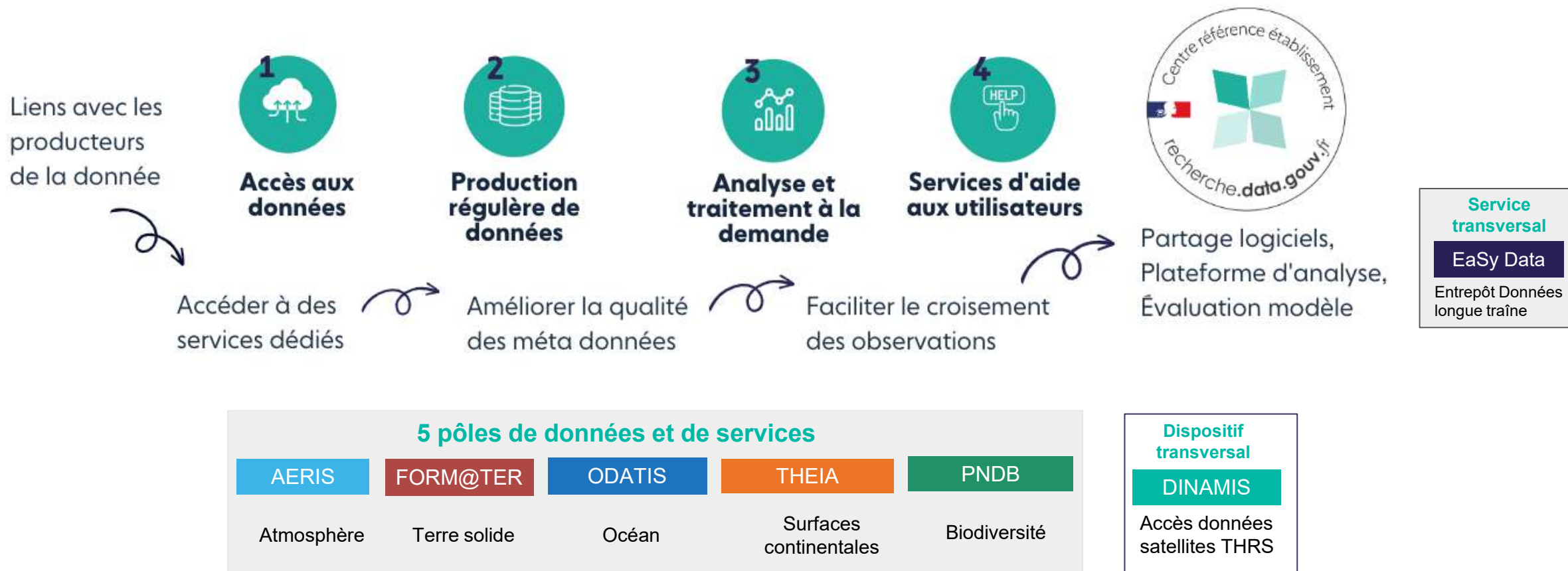


# Implantation de DATA TERRA et de ses Centres de Données et de Services



# DATA TERRA, une e-Infrastructure de Recherche dédiée au système Terre

L'IR DATA TERRA propose des **services** autour des données d'observation du **système Terre interopérables et interdisciplinaires** à tous les niveaux



# DINAMIS – un dispositif national institutionnel d'accès aux images satellites à Très Haute Résolution Spatiale

Pléiades (50 cm)



Spot 6-7 (1,5 m)



Pléiades Neo (30 cm)



## Caractéristiques

- 50 cm résolution spatiale (fusion)
- 5 bandes spectrales (panchromatique, bleu, vert, rouge, proche-infrarouge=)
- Traitements géométrique /radiométrique
- 1,5 m résolution spatiale (fusion)
- 5 bandes spectrales
- Traitements géométrique /radiométrique
- 30 cm résolution spatiale
- 7 bandes spectrales (Deepblue/Red Edge)
- Traitements géométrique /radiométrique

## Eligibilité<sup>1</sup> :

- Institutionnels publics français
- Scientifiques français (+européens/internationaux<sup>2</sup>)
- Prestataires français de commandes publiques
- Entités privées françaises pour des besoins de R&D

- ▶ Adhésion de l'Entité (signature Charte Dinamis)
- ▶ Désignation d'un référent
- ▶ Création de comptes d'accès individuel au sein de l'entité référente

<sup>1</sup>Entité gouvernementale, ministérielle, régionale, départementale, territoriale, communale, université, laboratoire scientifique, établissement d'enseignement,...), organismes associatifs ayant une mission d'intérêt public,

<sup>2</sup> Sous conditions

# DATA TERRA, accès unifiés pour accéder aux données et services d'observation du Système Terre et de l'Environnement



Atmosphère



Terre solide



Océan



Surf.continentales

## ✓ Catalogue de données et de services



## ✓ Services associés

Support aux campagnes, services aux producteurs de données, cluster de calculs, logiciels et produits thématiques, ateliers thématiques, appel à projets, newsletter, adhésion en ligne, etc



Service transversal



## ✓ Entrepôt


Données "orphelines"  
longue traîne

# ATMOSPHERE

## Physique et chimie de l'atmosphère, changement climatique, qualité de l'air, etc

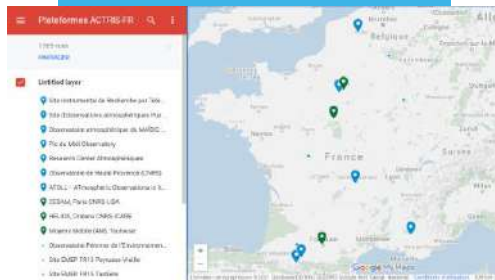
Le pôle de données et services AERIS regroupe les activités de **gestion des données et l'expertise scientifique de l'atmosphère.**

Les données proviennent de **plateformes, de réseaux d'observation, d'instruments à bord de satellites, de ballons, d'avions, de mesures en laboratoire, d'inventaires, de modèles, de données de campagnes scientifiques.**



### ETUDE DE CAS : QUALITÉ DE L'AIR

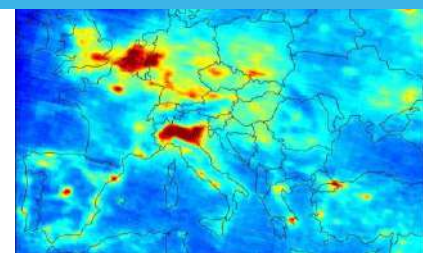
#### SURVEILLANCE QUALITÉ DE L'AIR



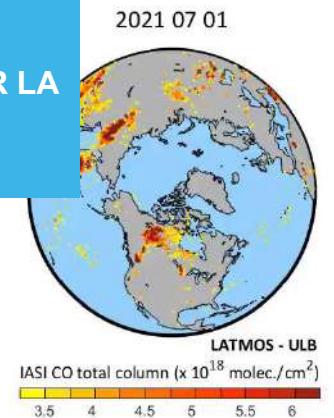
#### ESTIMATIONS DES ÉMISSIONS DE CARBONE



#### RÉSEAU EUROPÉEN DE LA QUALITÉ DE L'AIR



#### EFFETS DES FEUX DE FORÊTS SUR LA QUALITÉ DE L'AIR



## Connaissance de la structure et de la composition de la Terre, risques et ressources telluriques

Le pôle de données et services FORMATER regroupe les activités de **gestion de données** et **l'expertise scientifique de la terre solide**.

Les données proviennent de **missions satellitaires**, **d'instruments in situ** et **des résultats d'expériences**, de **modélisations** et de **simulations**.  
Des **services de calcul** sont également disponibles.

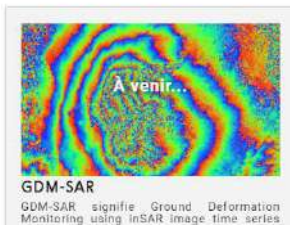
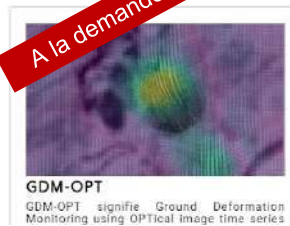


### ETUDE DE CAS

**CALCUL DE LA DEFORMATION DU SOL PAR IMAGERIE SPATIALE**

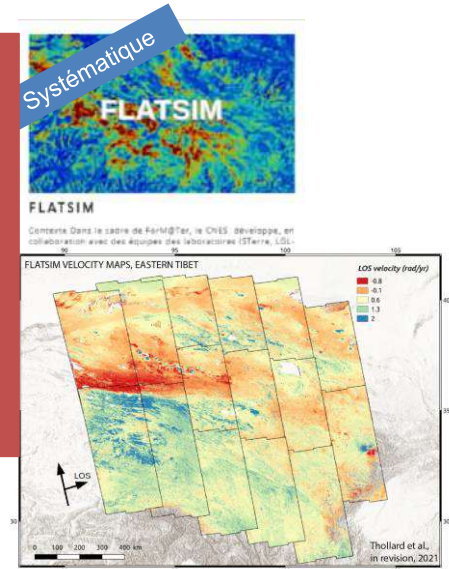
**SURVEILLANCE SPATIALE TEMPORELLE DES RÉGIONS CRITIQUES**

A la demande



**APPLICATIONS**

Zones de failles actives et systèmes magmatiques, glissements de terrain et déformations lithosphériques à l'échelle des grands massifs.





## Observations et compréhension des océans au service du développement durable

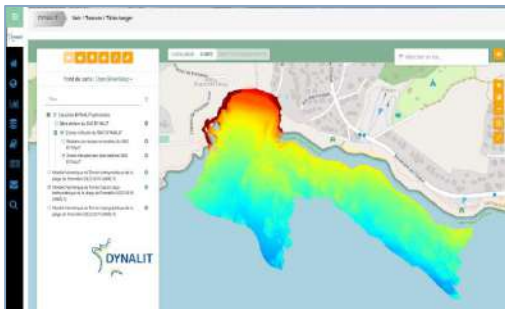
Le pôle de données et services OCEAN regroupe les activités de **gestion de données** et de **produits pour le bénéfice des sciences océaniques**

Les données proviennent de **missions satellitaires**, **d'instruments in situ** (flotteurs, planeurs, radars, campagnes en mer), d'**expérimentations** et de **mesures en laboratoire**.



### DIVERSITÉ DES DONNÉES : de la physique à la biologie

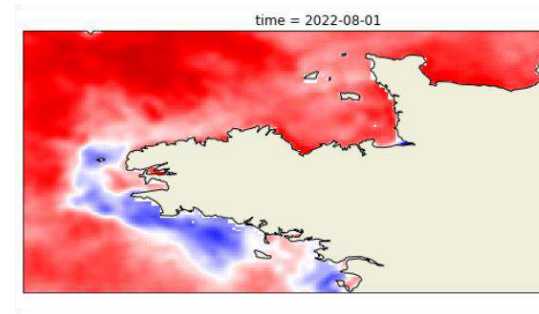
SURVEILLANCE DU LITTORAL



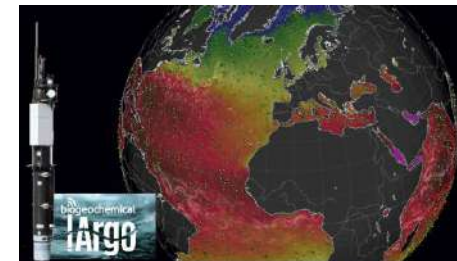
QUALITE DES EAUX / HYPOXIE



TEMPÉRATURE, VAGUE DE CHALEUR MARINE



BIOGÉOCHIMIE



[www.odatis-ocean.fr](http://www.odatis-ocean.fr)

# SURFACES CONTINENTALES

## Observation et compréhension des dynamiques des surfaces terrestres et des ressources naturelles

Le pôle de données et services THEIA regroupe les activités de **gestion de données** et **l'expertise scientifique des surfaces continentales**

Les données proviennent de **missions satellitaires, de réseaux d'observations et d'instruments in situ**. Il s'agit d'accompagner le développement de produits à valeur ajoutée et de services de traitement sur différentes thématiques : eau, forêt, neige & glace, risques naturels, santé, occupation des sols, urbain, agriculture, biodiversité, littoral, etc.

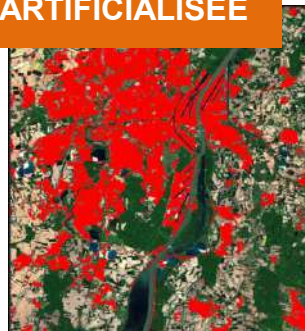


### UNE DIVERSITÉ DE PRODUITS THÉMATIQUES : de l'occupation des sols à la modélisation de paramètres biophysiques

OCCUPATION DES SOLS



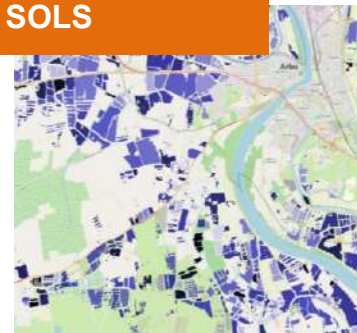
TACHE ARTIFICIALISEE



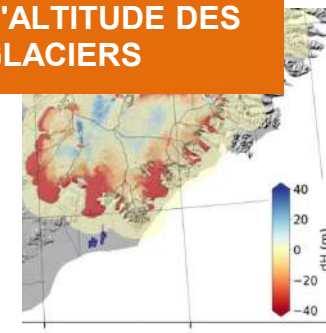
PRODUCTION DE LA BIOMASSE



HUMIDITÉ DES SOLS

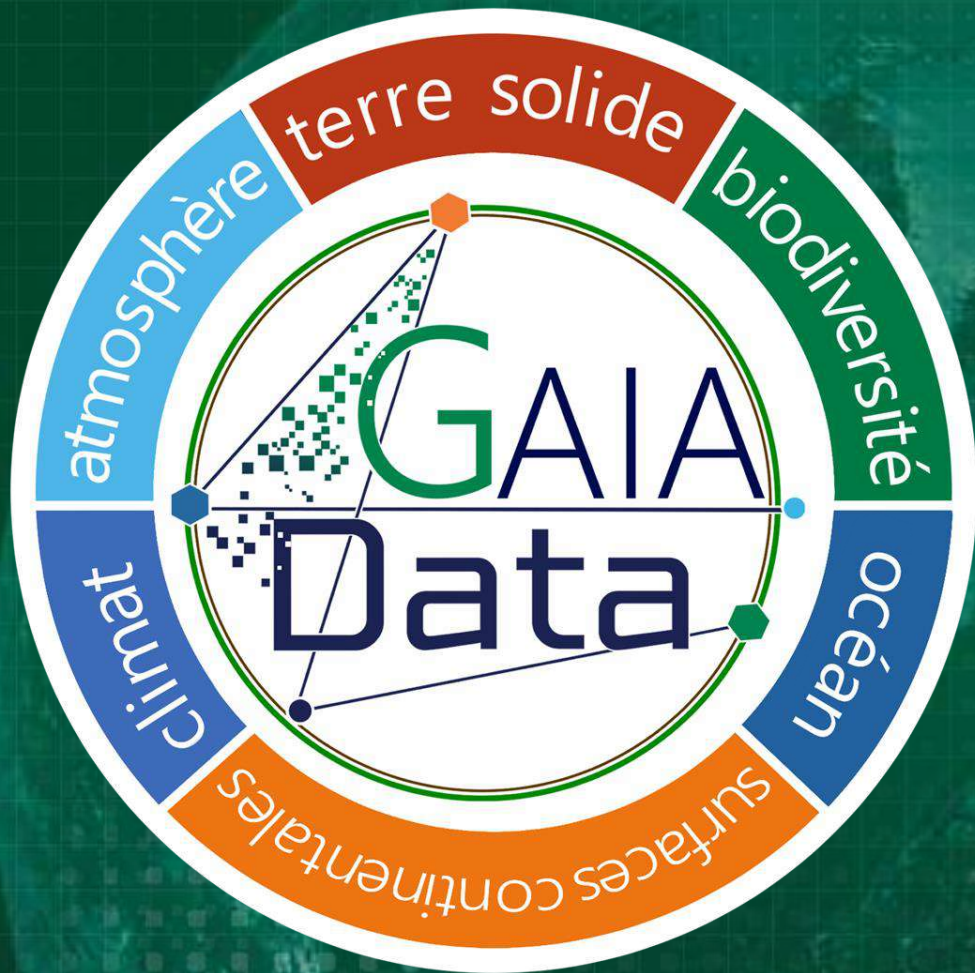


CHANGEMENT D'ALTITUDE DES GLACIERS



COUVERTURE DE NEIGE





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**

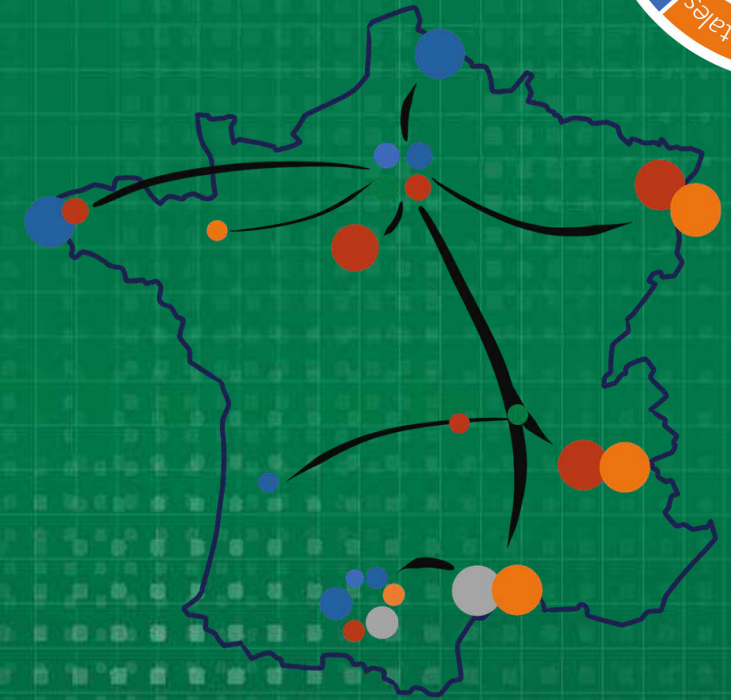
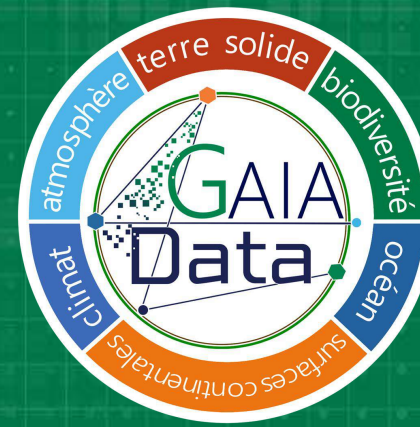
# Contexte : distribué géographiquement

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**

# 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

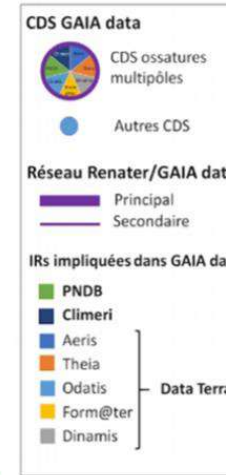
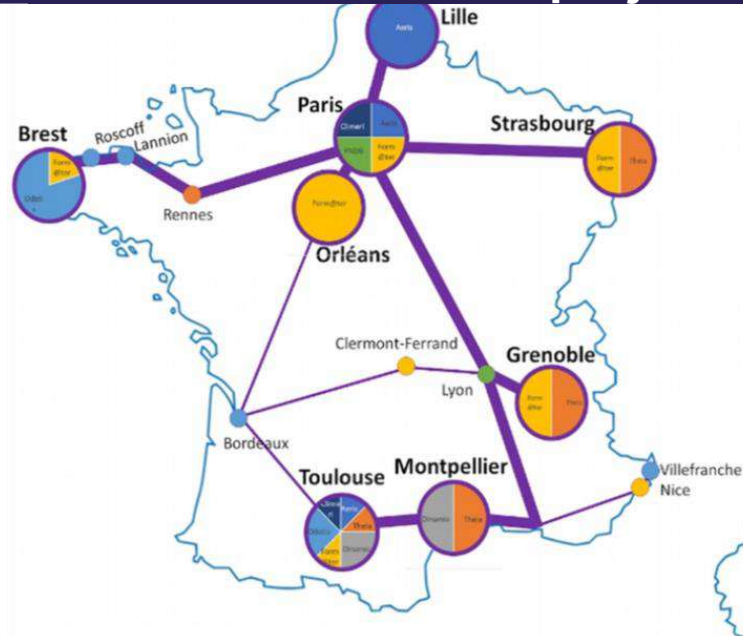
### Projets H2020 – Horizon Europe

- 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, Opendata, 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**



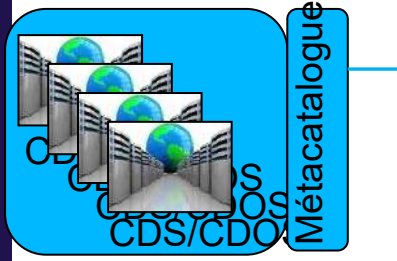
## Services de production réguliers

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

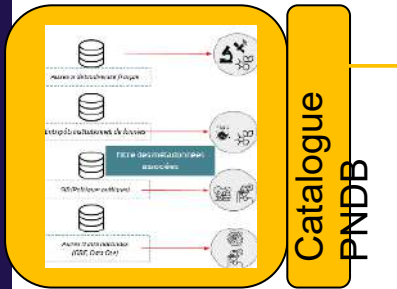
Supporté sur un continuum d'infrastructures partagées



Pôle X



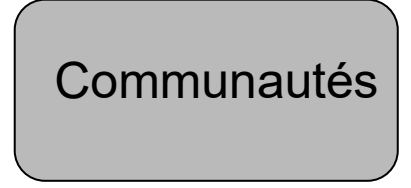
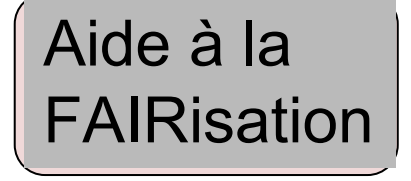
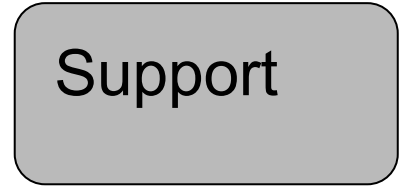
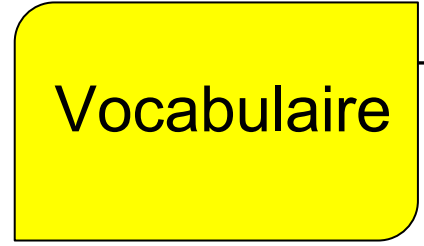
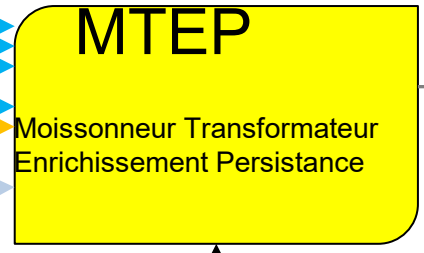
DINAMIS



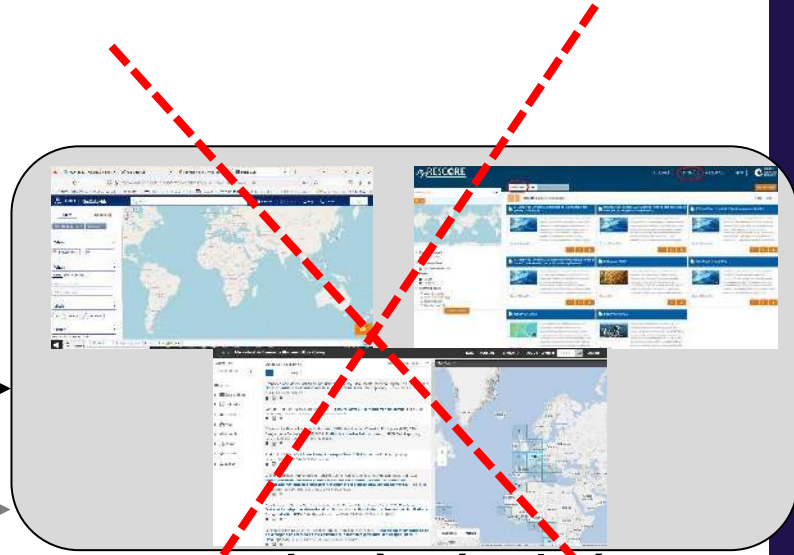
PNDB



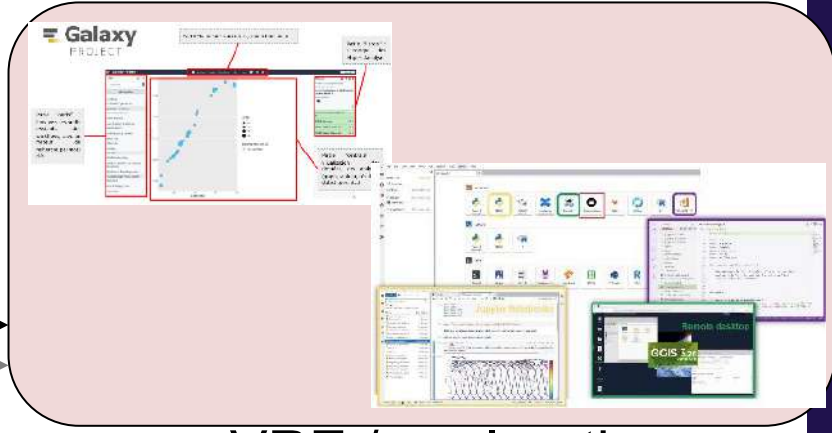
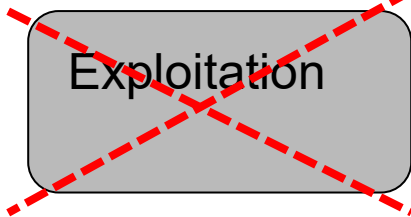
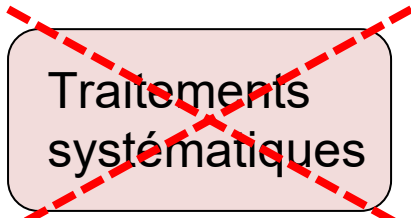
Catalogue CLIMERI



Portail de découverte



~~Accès évolué~~



VRE / exploration



En tant qu'utilisateur, je souhaite découvrir des jeux de données et les ressources (ex services, publications, codes, ...) associées pour répondre à mes objectifs scientifiques et techniques

J'ouvre la page du portail de découverte de GAIA DATA.  
Je peux choisir la langue du portail : français ou anglais.

Je peux m'authentifier tout de suite pour pouvoir accéder à mes collections personnelles et à des services

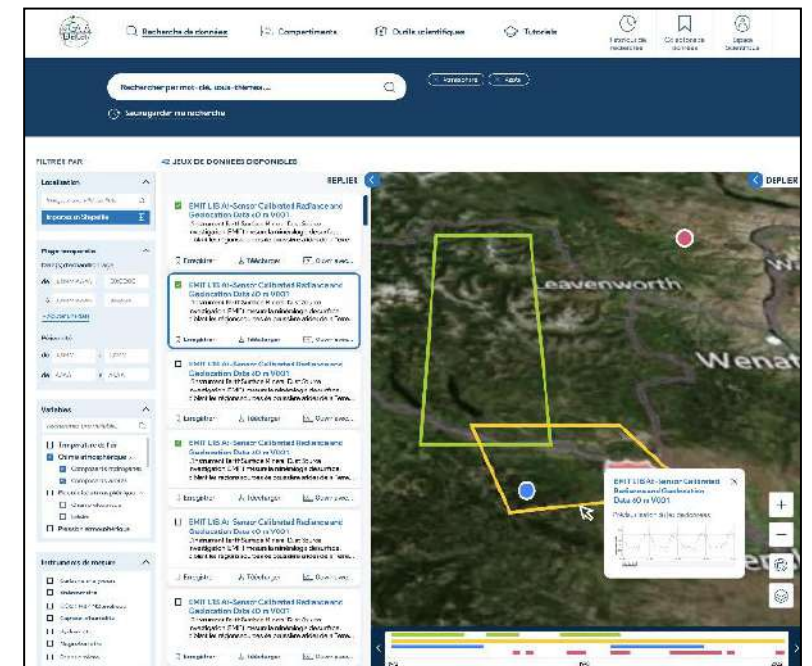
Je peux aussi découvrir des données sans avoir à m'authentifier. Je dispose pour cela de plusieurs canaux de découverte

Page d'accueil du portail de découverte  Point d'accès central à GAIA DATA





Un premier 'A la Google' : je peux saisir du texte dans une boîte de texte. Le système me propose des suggestions par auto-complétion basées notamment sur le thesaurus de GAIA DATA.



Je peux aussi découvrir des données sans avoir à m'authentifier. Je dispose pour cela de plusieurs canaux de découverte

GAIA Data

Recherche de données Compartiments Outils scientifiques Tutoriels

Historique de recherches Collections de données Espace Scientifique

## Les compartiments Système-terre et environnement

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer sodales gravida porta. Aenean ac magna eu velit mattis blandit. Etiam vitae eros tellus. Etiam sed sagittis erat. Etiam quis nunc sem.

### Découvrez les compartiments

Terre solide Océans Atmosphère Climat Biodiversité Surfaces continentales Agro-environnement

Deuxième mode de découverte, via l'exploration de thesaurus.

GAIA Data

Recherche de données Compartiments Outils scientifiques Tutoriels

Historique de recherches Collections de données Espace Scientifique

## Découvrez le compartiment Atmosphère et ses variables

L'atmosphère est une enveloppe gazeuse qui entoure et protège notre planète du rayonnement intense du Soleil et sert d'interface clé entre les cycles terrestre et océanique.

Recherchez les données Atmosphère

### Introduction

L'atmosphère terrestre est un mince voile de gaz entourant la planète. Bien qu'il ne s'étende qu'à quelques centaines de kilomètres au-dessus de la surface, il contient un mélange de gaz, tels que l'oxygène et l'azote, qui sont essentiels à l'existence de la vie. Il distribue le rayonnement solaire entrant, protégeant la vie des rayonnements ultraviolets nocifs, mais régulant également la circulation atmosphérique et les conditions météorologiques.

L'atmosphère permet l'effet de serre, ce qui rend la Terre plus habitable. L'activité humaine, cependant, génère plus de gaz, ce qui a beaucoup d'impact négatif sur la nature protectrice de cette couche vitale. Les données de la NASA fournissent des mesures sur les phénomènes météorologiques ainsi que sur les gaz dans la troposphère (la couche la plus basse de l'atmosphère) et la stratosphère (la couche au-dessus de la troposphère) et leur effet sur la qualité de l'air.

### Sous-domaines et variables relatives

Recherchez directement des données en cliquant sur la variable de votre choix

Aérosols Qualité de l'air **Climat atmosphérique** Pression atmosphérique Rayonnement atmosphérique Température atmosphérique Vapeur d'eau atmosphérique Vents atmosphériques Nuages Précipitations Événements météorologiques

### Centres de données & infrastructures

Data Terra AERIS Satmos IAGOS

Ci-dessus la page qui apparaît après avoir cliqué sur 'atmosphère'

- Propose d'aller plus finement (ex aérosols)
- Ou d'afficher la Web SIG avec les données qui sont tagguées 'Atmosphère'

# Portail de découverte : Design Thinking

## Découverte #3 : 'Web-SIG sur jeux de données'

4/9

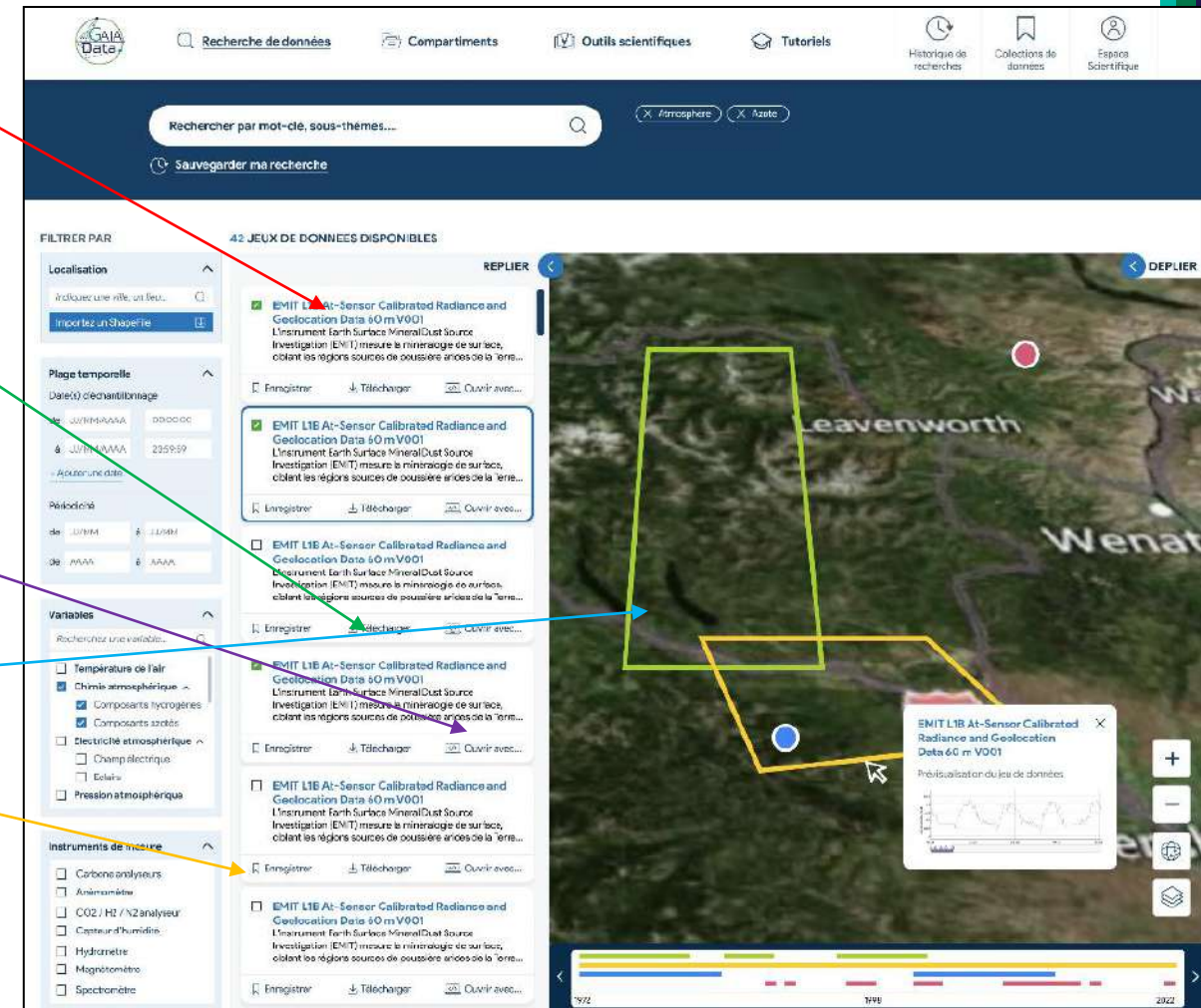
Affiche les informations détaillées sur le jeu de données et à une liste de ressources associées  
=> Doit permettre la réutilisation par un non spécialiste

Permet d'accéder à un système d'accès évolué pour ce jeu de données (système originel : pôles, IR, GDH, système GAIA DATA ... ?)

Permet de lancer une VRE / Earth analytics Labs sur cette donnée : lien avec T3.3

Carte affichant les emprises des jeux de données

Pour ajouter à ma/mes collection(s) de jeux de données



# Assistant virtuel Gaia Data (Prestation de Cap Gemini)

La prestation porte sur l'étude des possibilités offertes par les technologies 'Generative AI' et 'Large Language Model' pour GAIA DATA RAG sur les données du pôle ODATIS (génération augmentée de récupération)

## PLANNING



20/06

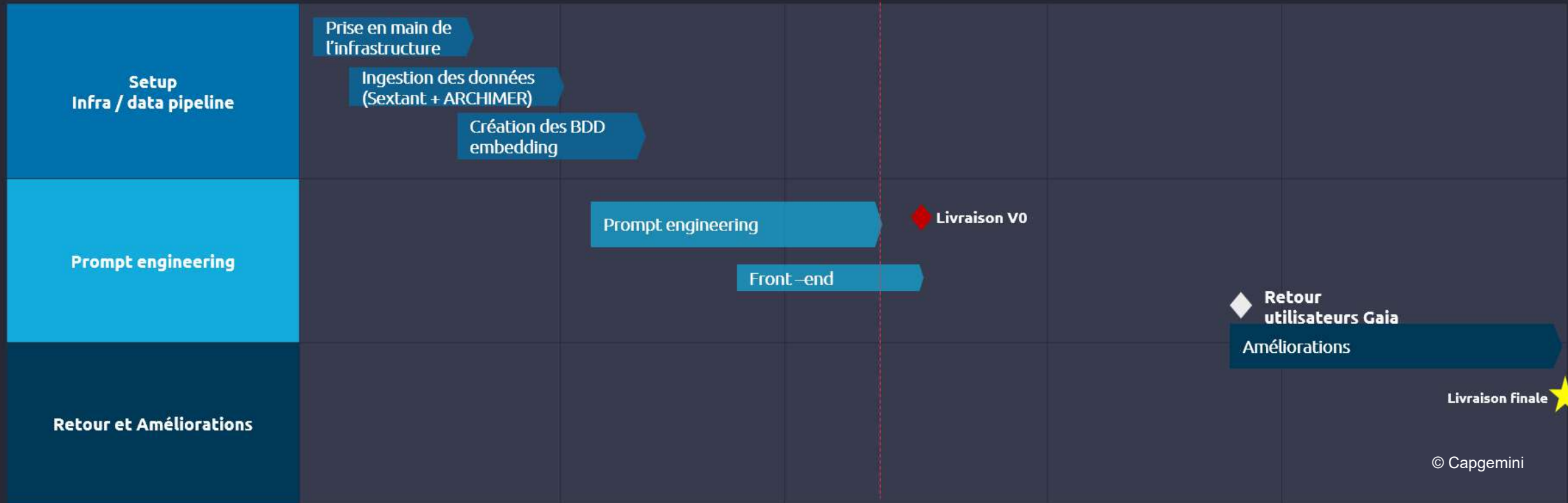
31/06

S1-S2

S3-S4

S5-S6

S7-S8





**DATA**TERRA



[contact@data-terra.org](mailto:contact@data-terra.org)

+33 (0)4 67 54 87 08

[www.data-terra.org](http://www.data-terra.org)