



**DATA  
TERRA**



**ODATIS**

# Atelier VRE - Plateformes d'analyse

Assemblée Générale ODATIS

12-13 décembre 2023

Observatoire Aquitain des Sciences de l'Univers, Pessac



13/12/2023

[contact@odatis-ocean.fr](mailto:contact@odatis-ocean.fr) | [www.odatis-ocean.fr](http://www.odatis-ocean.fr)

# Contexte

**VRE** = Environnement Virtuel de Recherche / Plateformes d'analyse / Cloud pour la science

A l'intersection de plusieurs concepts et besoins autour d'une science plus ouverte et d'une exploitation plus optimale des ressources, avec différentes implications techniques et organisationnelles :

**Accès aux données**

**Partage d'outils et de codes**

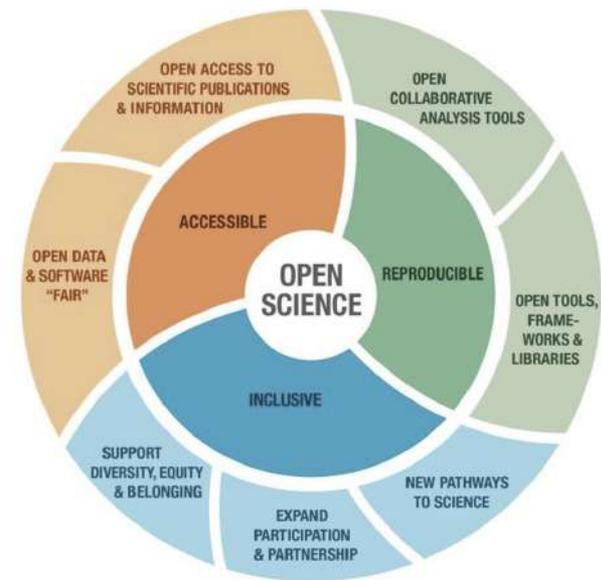
**Traitement local vs distant**

**Data intensive science**

**Interactivité**

**Intégration des centres de calcul et de données (Data Terra, Europe,...)**

**Données + Infrastructure + Software + Communauté**



Source : NASA/ESDS

# Défis

Difficultés d'usage et accès aux données  
(FAIRitude)

Volume croissant et non gérable localement par  
les utilisateurs

Changer le modèle classique de travail :  
déporter codes et traitements au plus près des  
données

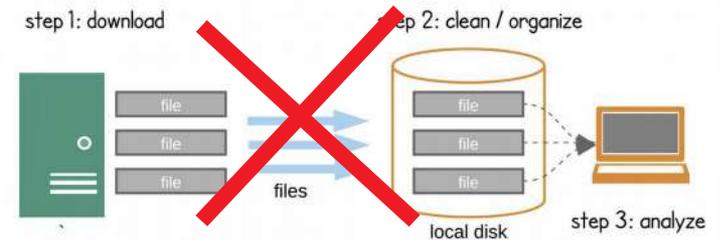
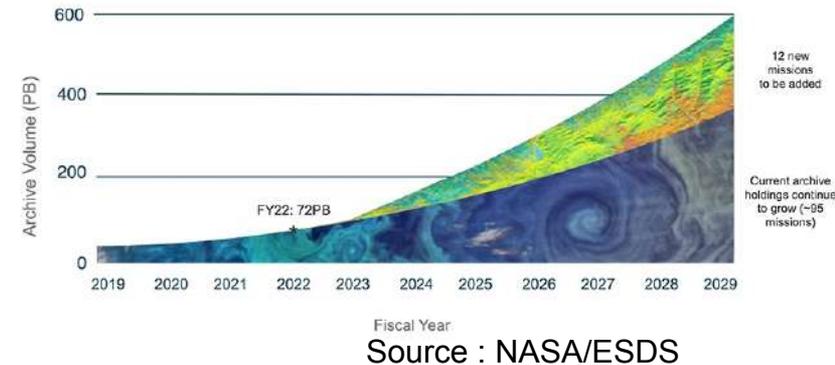
Proposer des moyens de traitement interactifs  
mais aussi de traitement de lot

Workflows scientifiques

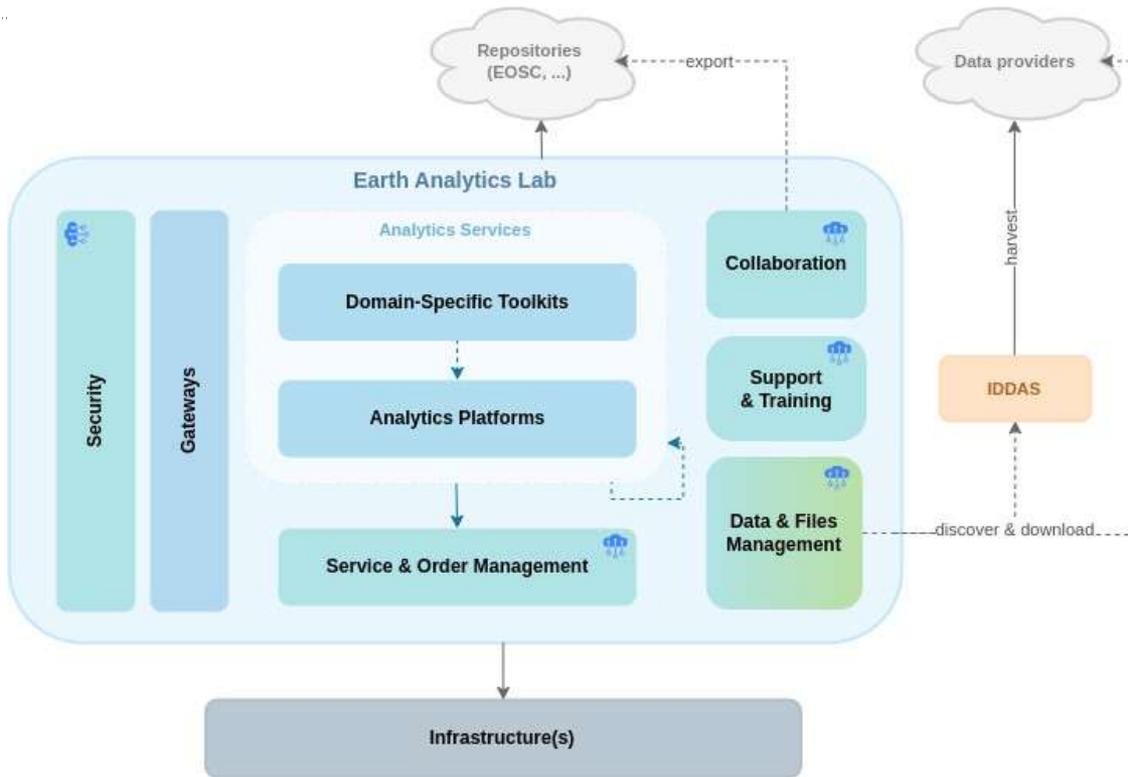
Moyens et outils de partage, transfert et  
réplication de données entre plateformes et avec  
d'autres sources

Offres (semi)-commerciales déjà disponibles  
(WeKEO, CREODIAS, AWS,...)

## Earth Science Data Archive Growth Projection



# Plateformes VRE - architecture



## Interdisciplinary Data Discovery and Access Service

Discover data collections and retrieve data fragments.

Use semantic capabilities (assets catalogue/selector)

## Earth Analytics Lab

Integrated, Collaborative and FAIR web platform (not just a set of softwares).

Abstract the technical complexity presented to the end user as far as possible

Based on FAIR principles, for data, software, and middleware.

## Core services

Middleware (special software) that connects and facilitates communication between different components (hard or soft). Some of these can be supplied by EOSC's core/federated services.

## Example of data flow



Source : E. Bodéré, FAIR-EASE project

# OPEN SCIENCE PLATFORM

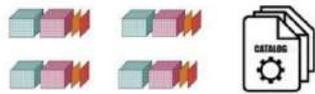
## "Analysis Ready, Cloud Optimized Data"

Cleaned, curated open-access datasets available via high-performance globally available storage system

## "Data Proximate Computing"

Bring analysis to the data using any open-source data science language.

### Data Library



Generic cloud object storage

### Compute Environment



Generic cloud computing

## "Elastic Scaling"

Automatically provision many computers on demand to accelerate big data processing.



**Expert Analyst**  
Direct Access via Jupyter



Web front ends



**Non-Technical User**  
Access via apps / dashboards / etc.

Downstream third-party services / applications



# Contraintes sur la découverte et l'accès aux données

Référencement des données disponibles dans ODATIS (Data Terra) – association avec les services (dont VRE) disponibles.

Données référencées directement accessibles à partir des VRE pour lecture, analyse, visualisation. Intégration des CDS auprès des infrastructures de traitement nécessaire.

Services de sélection de fragments de données (ex : zone et période d'intérêt pour des données satellite) – Opendsearch, STAC, Intake

Services de subsetting (THREDDS/OpenDAP,...)

Accès réseau local, accès cloud / stockage objet (S3), accès externe (FTP, HTTPS), services OGC WMS, WCS, WMTS,...

Transformation Data Cubes, formats ARCO (Analysis Ready, Cloud Optimized) → jeux de référence, optimisation pour le traitement interactif, jeux de données multi-variables (pour l'apprentissage),...

Mécanismes de partage, transfert et réplication (EODAG, iRODS,...) entre plateformes et utilisateurs

# Outils et services

Des environnements scientifiques dédiés (Conda/dockers, Pangeo) pré-configurés et extensibles

Intégrés dans des frontaux d'analyse interactive (jupyter notebooks, R-studio, ...)

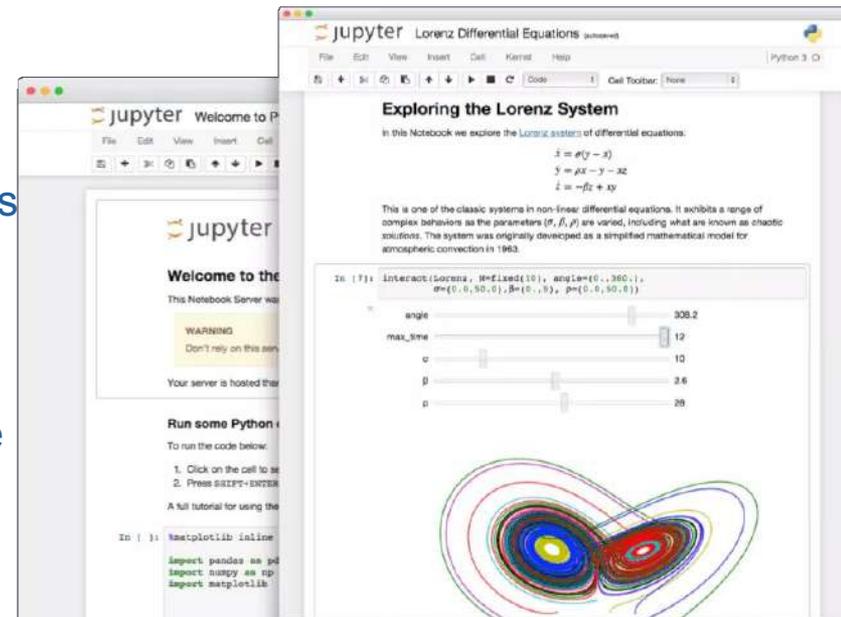
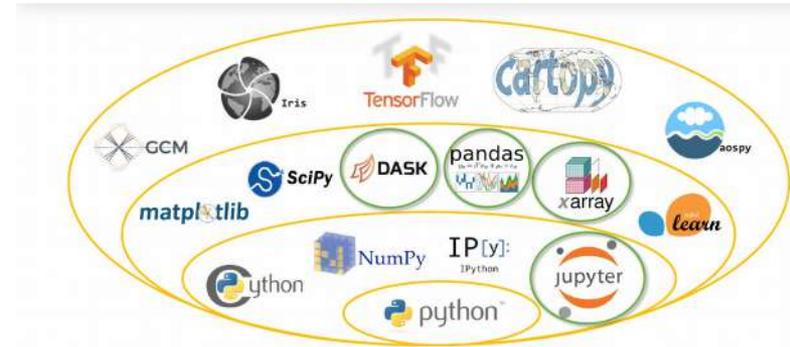
Gestion et partages de workflows scientifiques (Galaxy, Airflow,...)

Traitements et transformation de données « pré-cablés » (OpenEO,...)

Outils Desktop (ex : QGIS) et spécifiques à certains domaines ou communautés – déploiement à la demande de services et outils

Doivent pouvoir être assemblés et spécialisés sur un chantier / thématique / communauté particulière

Support aux utilisateurs



# Infrastructures

2 infrastructures principales : TREX (CNES) et Datarmor (Ifremer)

Intégration plus large dans Data Terra avec autres infrastructures nationales couvrant d'autres domaines

Moyens de calcul : HPC, Cloud, GPU

Accès élargi à terme (communauté eduGAIN ?) - question des ressources... => VRE dédiées

Intégration avec les bases de données

# Exemple d'une plateforme dédiée : CNES - SWOT

# Points de discussion

## Services VRE proposés – adéquation aux besoins

- Les concepts, services et outils mentionnés sont-ils familiers ?
- Quel est le niveau de maturité des utilisateurs par rapport à ce changement de paradigme ?
- Répondent-ils aux besoins ?
- Quels sont les manques ou besoins spécifiques ? Points de blocage actuels pour accès et traitement de données, ...

## Engager les utilisateurs – aider à l'adoption par la communauté

- Développement de cas d'usage, « beta testeurs »
- Présentations aux labos, formation, webinar,...
- Partage d'expérience, tutoriels, notebooks, outils d'échange/communication
- Support utilisateur à envisager ?
- Suggestions de démarche ?