



## Infrastructure de Recherche DATA TERRA Pôles de données et services pour le système Terre

**Frédéric Huynh (IRD), directeur IR Data TERRA**

Michel Diamant (IPGP), directeur pôle Form@Ter ; Patrice Henry (CNES), directeur pôle AERIS ; Gilbert Maudire (IFREMER), directeur pôle ODATIS ; Nicolas Baghdadi (INRAE), directeur pôle THEIA ; Richard Moreno (CNES), directeur technique  
Nicole Papineau (CNES), chargée de mission ; Jean-François Faure (IRD), secrétaire exécutif DINAMIS

# Contexte et Enjeux

**La Terre**, un système complexe dynamique :

- processus géophysiques et environnementaux,
- différentes échelles spatiales et temporelles,
- interactions permanentes *entre les compartiments Terre solide, surfaces continentales, océan, atmosphère et l'anthroposphère.*

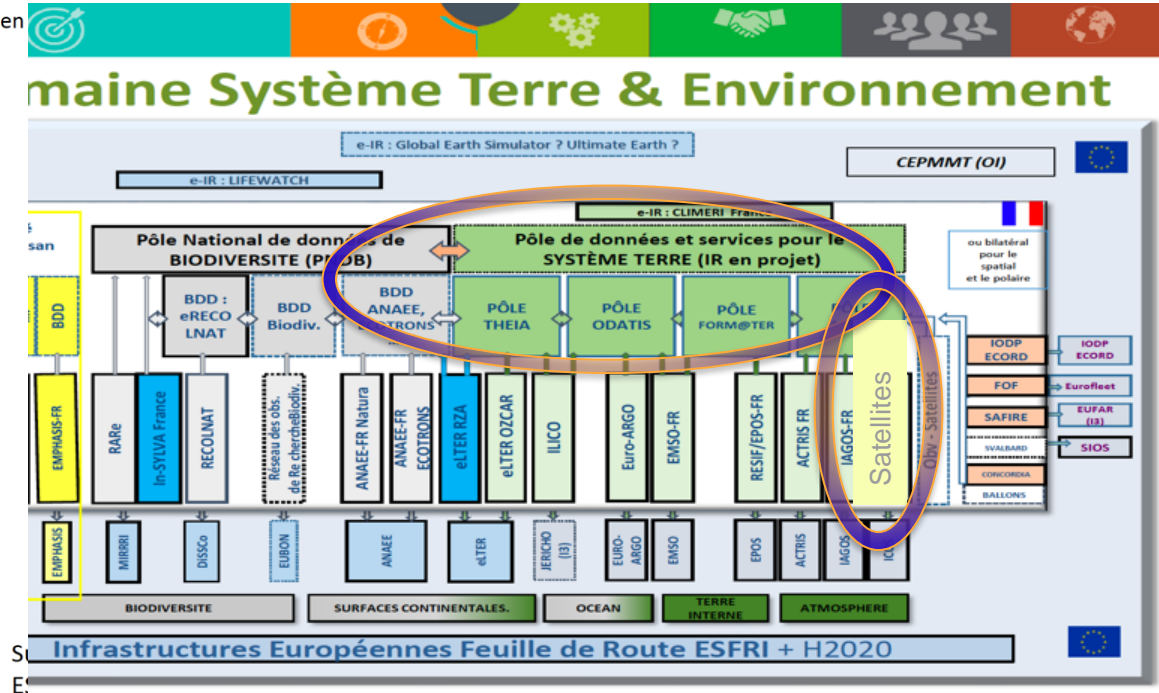
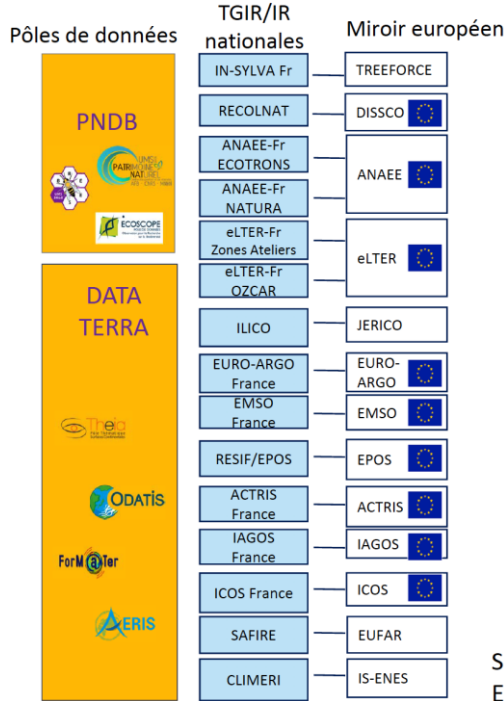


After O. de Viron



Comprendre ces processus nécessite l'accès et l'analyse de nombreuses, complexes et volumineuses sources de données (*satellites, in situ, campagnes, observations à long terme ainsi que des résultats d'expérimentation, modèles etc ...*)

Elisabeth Vergès, MENESR-DGRI—SSRI-A1

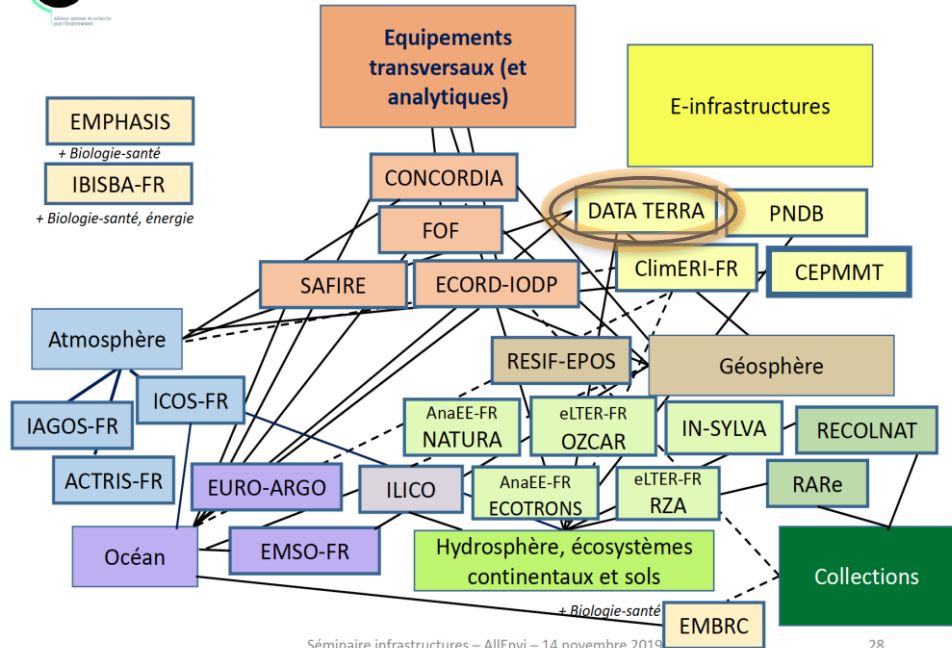


« Terre Vivante »

« Terre solide et fluide »

# Infrastructures de recherche du domaine Système Terre et Environnement

AllEnvi



Séminaire infrastructures – AllEnvi – 14 novembre 2019

28

## Grands types d'infrastructures dans le domaine des Sc. du Système Terre & Environnement (hors OI) de la feuille de route nationale

- 4 INFRASTRUCTURES LOGISTIQUES transversales : flottes bateaux, avions, plate forme forage, base antarctique
- 2 INFRASTRUCTURES COLLECTIONS : biologiques et géologiques
- 15 INFRASTRUCTURES OBSERVATION & EXPERIMENTATION sur les différents compartiments du système Terre
- 3 E-INFRASTRUCTURES Pôles de données compartiments système Terre, biodiversité modélisation simulation



# Contexte de la création de l'IR Data TERRA

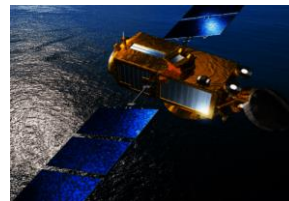
- **Groupe de réflexion** « pôles thématiques en Observation de la Terre » 2012-2014 – CNES-CNRS/INSU, ... -  
Création de pôles de données et d'une IR système Terre
- **Feuilles de route nationale IR/TGIR – MESRI** contribution domaine « système Terre et Env. » 2016
- **Open science / open data**
- **Européanisation** des activités : **infrastructures de recherche** (ESFRI...), **services d'accès** aux données spatiales et in-situ (EOSC, COPERNICUS, ...)
- **Internationalisation** des dispositifs de partage des données et services (GEO, GO FAIR, ...)



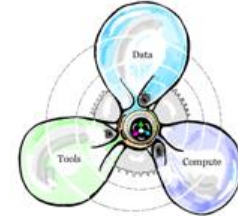
# Objectif général

Développer un dispositif global d'accès à des **données, produits et services** permettant **d'observer, comprendre et prévoir** de manière **intégrée** l'**histoire, fonctionnement et évolution** du système Terre soumis aux changements globaux

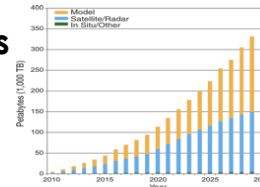
- **Faciliter l'accès et l'utilisation** des **données et produits** de qualité sur l'ensemble des **compartiments du système Terre**  
=> **Données spatiales, aéroportées, sols, in-situ**
- **Développer des services de visualisation et de traitements adaptés aux besoins, à l'accroissement de la volumétrie et aux avancées technologiques**
- **Favoriser la mutualisation, interopérabilité, émergence d'approches multi- et inter-disciplinaires**
- **Servir les communautés scientifiques, les acteurs de l'action publique et de l'innovation**
- **Mettre en œuvre une stratégie nationale, européenne et internationale**



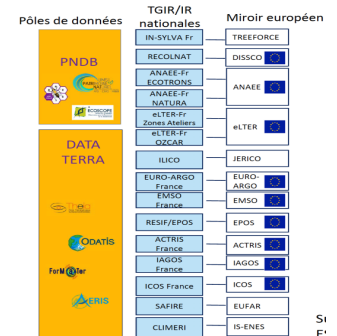
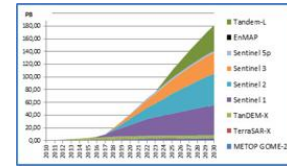
# Besoins et verrous



- **Evolution des besoins des communautés scientifiques**  
données **multi-sources**, **multi-capteurs**, services d'accès aux données, traitements, analyse/modélisation, IA, approches intégrées des interactions et complexité
- **Augmentation exponentielle du nombre de données, diversité des sources, complexités, ...**  
Spatiales, in-situ, modèles => besoins d'analyse/réanalyse, traitements intelligents
- **Co-construire une infrastructure distribuées de services spécifiques du domaine système Terre et Environnement**  
=> enjeux scientifiques et données transverses
- **Concilier structuration nationale transversale à l'Europe**  
=> modèle d'IR Data Terra en Europe et à l'international
- **Concilier recherche d'excellence** et développement de partenariats avec les acteurs publics et économiques



Projected increase in global climate data for climate models, remotely sensed data, and in situ instrumental/proxy data. From Overpeck et al. Science, 2011



# Défis scientifiques transversaux - Verroux scientifiques et techniques associés aux données

CONTINUUM ET INTERFACES TERRE – MER



CONTINUUM ET INTERFACES TERRE – MER

APPROCHE INTÉGRÉE DE LA ZONE CRITIQUE

OBSERVATION ET MODÉLISATION DES IMPACTS DES CHANGEMENTS GLOBAUX

→ développement d'une approche intégrée

DYNAMIQUE DES MILIEUX/ENVIRONNEMENTS URBAINS

DYNAMIQUE DES ÉCOSYSTÈMES MARINS

Echelles spatiales fines (mésos- et submésoséchelle)

→ accéder aux données in situ et à des produits spatiaux (ex. couleur de l'eau sur des zones géographiques méso-échelles)

CLIMAT ET CYCLE DE L'EAU

GESTION DURABLE DES RESSOURCES

ALEAS ET IMPACTS SUR LA SOCIÉTÉ

CLIMATOLOGIE DE RÉFÉRENCE DU MILIEU CÔTIER

ÉVOLUTION DU NIVEAU DE LA MER

GT Science en lien avec les CS des pôles – priorités et prospectives CNRS/INSU, CNES (CPS, TOSCA, ..), IRD, INRAE, ...



# Structuration IR Data Terra et pôles : dates clés

AERIS  
THEIA  
ODATIS  
Form@ter

SATMOS (1986)  
Pôle thématique ETHER (1996)

Rapport  
« Walteufel »

Pôle thématique ICARE  
MEDIAS/SEDOO  
Postel  
CADTS

Création du pôle  
thématique surfaces  
continentales

Rapport "Podaire"  
Mission de préfiguration  
des pôles

1996

2012

2013

CIO Pôle système Terre  
Nomination des 3  
chargés de mission et  
Directeur pour Theia

Création AERIS  
16/12/2014  
Comités Directeurs  
et BE

Création ODATIS  
Création du CS  
Premier portail  
Form@ter

Assemblée Générale  
Création de GTs  
Succès AO PIA3  
ETALAB

Signature 2<sup>ème</sup>  
convention Theia  
Lancement site web  
Equipe de direction  
Portail FORM@TER

REVEX AERIS  
Avis très positif du  
groupe de revue

Signature de  
convention  
ODATIS prévue  
Service Opérationnel FLATSIM  
Imagerie radar et optique

2013

2014

2015

2016

DINAMIS

2017

2019

2020

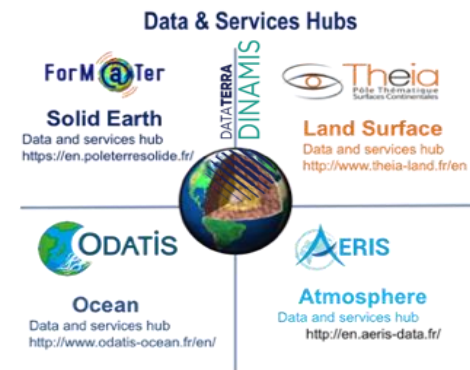
Création des différentes :  
instances : CD, BE, CS, CES, ART

2015 – 2018 Nouvelles tutelles : ONERA Cerema  
AgroParisTech

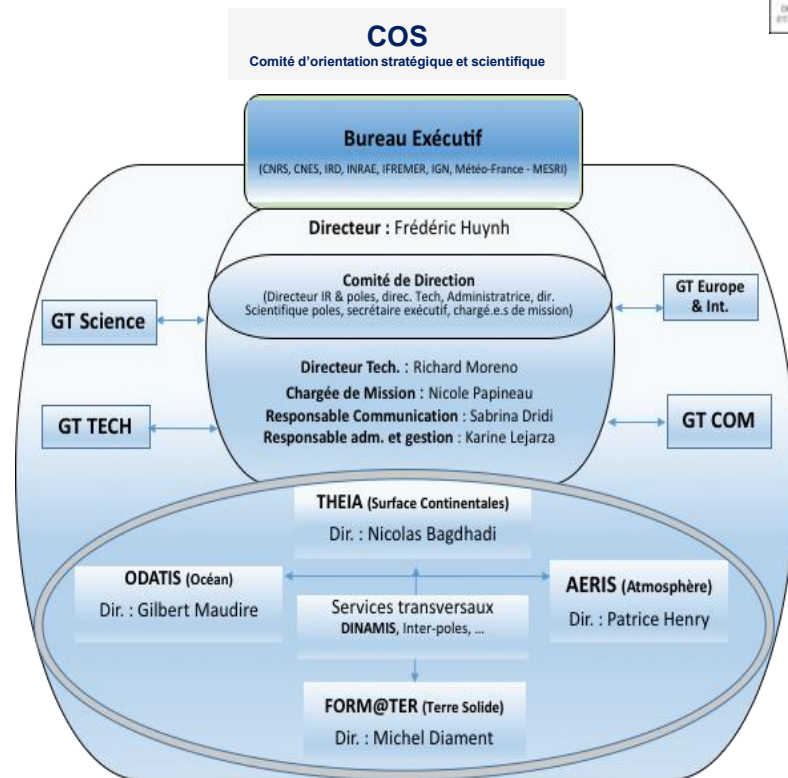
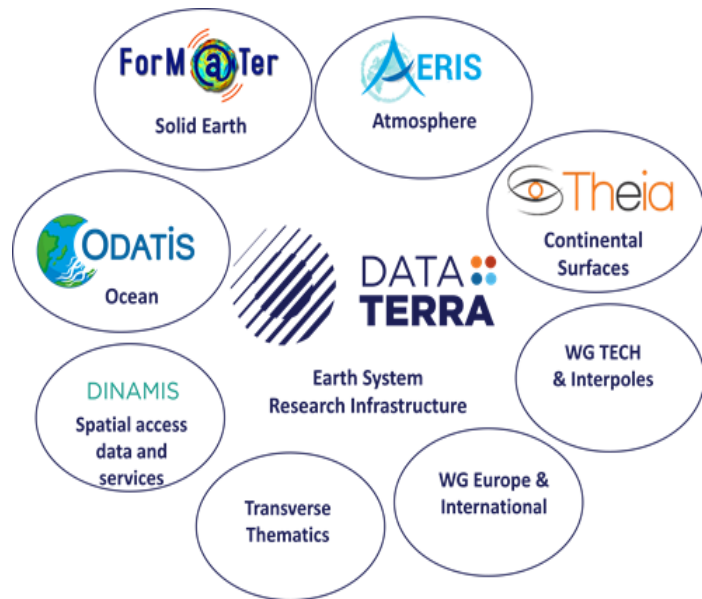
Portail OZCAR - THEIA



- Nomination de chargés de mission - mise en place des pôles (2014/2015)
- Lancement du GT "Inter-poles" (2015)
- Inscriptions sur la feuille de route nationale IR/TGIR du MESRI en 2016 et 2018
- Directeur de l'IR nommé en mai 2017
- Création de l'UMS CPST : CNRS en 2019, IRD en 2020, ...
- Lancement de DINAMIS (2019)
- Mise en place d'une Direction Technique (2019)
- Groupes de travail : GT TECH (2019), GT COM (2019), GT Science (2020), ...
- Projets nationaux (ANR, CPER), Européens (2018, 2020)
- Dépôt projet GAIA Data – AMI EQUIPEX+ (2020-2029)



# Organisation de l'IR Data Terra



# Data Terra en quelques chiffres

- 26 organismes et universités
- 4 pôles de données
- 6 services (DINAMIS) et groupes de travail transversaux
- 30 Centres de Données et de Services (CDS) et Infrastructures de données spatiales (IDS)
- 25 Consortium d'Expertise Scientifique
- 170 ETPT / 400 scientifiques, ingénieurs et techniciens appartenant à plus de
- 33 M€ (2016), 39 M€ (2017), > 40 M€ (2019)
- Plus de 500 produits et services, plus de 15000 utilisateurs
- 50 000 To (2018) ; 100 000 To (2022/2023) ; 150 Peta (2025)

# Participations actives à des projets nationaux, européens et internationaux

## ENVRI – FAIR : FAIRisation multi-domaines

Implication au travers des IR/TGIR ESFRI : ACTRIS, IAGOS, EPSO, EURO-ARGO, ANAEE ; 37 part., 19M€ - 16% Fr (CNRS, Ifremer, BRGM, INRA)

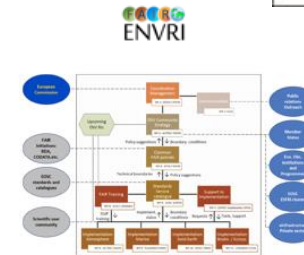


Figure 5. Work package structure and work flow of ENVRI-FAIR.

## CEF – PHIDIAS (CEF – DG Connect, 3 M€) : service à la demande

« données spatiales et environnementales » et HPC : CINES, CNRS, CSC, IFREMER, IRD, MARIS, Univ. Louvain, ...



## EOSC-Pillar : gouvernance nationale et Eur. EOSC (INFRA-EOSC, 12 M€), coordination WP6 : 10 études de cas (FAIR Data and Services)



## Seadatanet - SeadataCloud, projet d'appui à l'Infra (10 M€)

Blue Cloud : EOSC ocean (10 M€) Ifremer/Odatis, CNRS, Sorbonne Univ. Mercator-Ocean, ...



## Copernicus : services C3S, CAMS, Land, Marine, ...



ANR Flash (open science) : 4 projets : FAIRisation, certification



GO-FAIR IN : GAIA Data In (global integrated Data Impl. Net.)



GEO : programme Board, Working Group, initiatives, ...



Porté par 3 Infrastructures de Recherche : Data Terra, CLIMERI, PNDB



IR PNDB



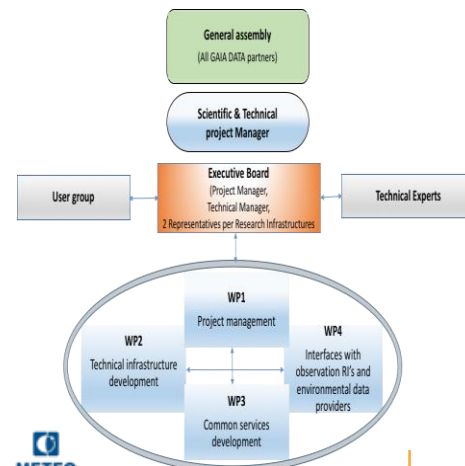
**21 Partenaires** : CNRS (coord.), CNES, IFREMER, IRD, BRGM, IGN, INRAE, Météo-France, MNHN, CEA, IGP, CINES, Sorbonne Univ., Univ. Grenoble-Alpes, Univ. Lille, Univ. F. Toulouse, UNISTRA, SHOM, OCA, FRB, CERFACS



**Objectif** : Développer et mettre en œuvre une infrastructure/plate-forme intégrée de données et de services distribuées 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 (observation, modélisation), de son **acquisition** (spatiale, sols, in-situ) jusqu'à ses **multi-usages** (qualification/validation, stockage, traitements/extraction de connaissances, produits, services, ...)
- pour la communauté scientifique contribuant à la connaissance du système Terre, de la biodiversité et de l'environnement

**Budget** : 65 M€ (couts complets) - Demande ANR-EQUIPEX+PIA3 : 19,6 M€

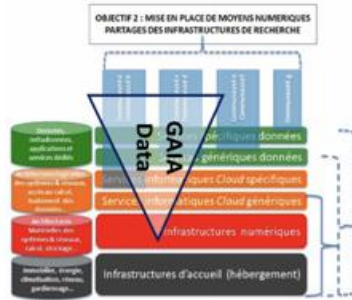
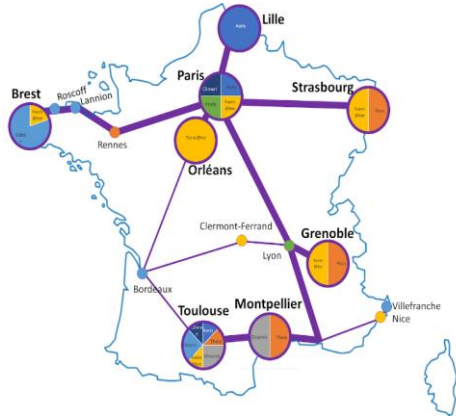


# GAIA DATA Caractéristiques du projet

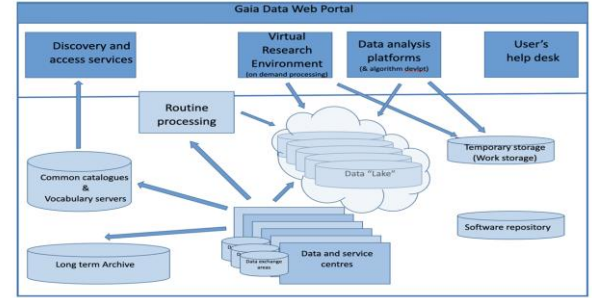
## Infrastructure distribuée de services

Articulation avec les infrastructures nationales (TGIR GENCI : CINES, IDRIS) et régionales : faciliter un accès intégré à un continuum de services distribués de stockage adaptés à la diversité et aux volumes de ces données, et de calcul

**Fédérer 8 sites principaux**  
Associer 30 sites existants



OBJECTIF 1 : MUTUALISATION DES INFRASTRUCTURES A L'ECHELLE DE CHAQUE COUCHE DE SERVICE\*



### Services Découverte, Accès et Gestion des données

- Catalogue (métadonnées, vocabulaires, ontologies),
- Archive long terme, entrepôts, DOI, Services avancés de visualisation
- Aide à la collecte des données

### Services transversaux => travaux transdisciplinaires

- Grille de données, cloud, standardisation de la production, portail connaissances, support utilisateurs & formation

### Services – Virtual Analysis Platform

- Travail collaboratif, bac à sable, développement et exécution d'algorithmes

### Services d'analyse des données à la demande & Virtual Research environnement

- Interface interactive
- Exécution par les utilisateurs
- VRE : définition et exécution de workflows de traitements spécifiques des domaines

### Services de production réguliers

=> Data Scientist  
=> Thématiciens  
=> Scientifiques non informaticiens

