

# Introduction

Contexte, Objectifs, Programme, Logistique



# CONSTATS

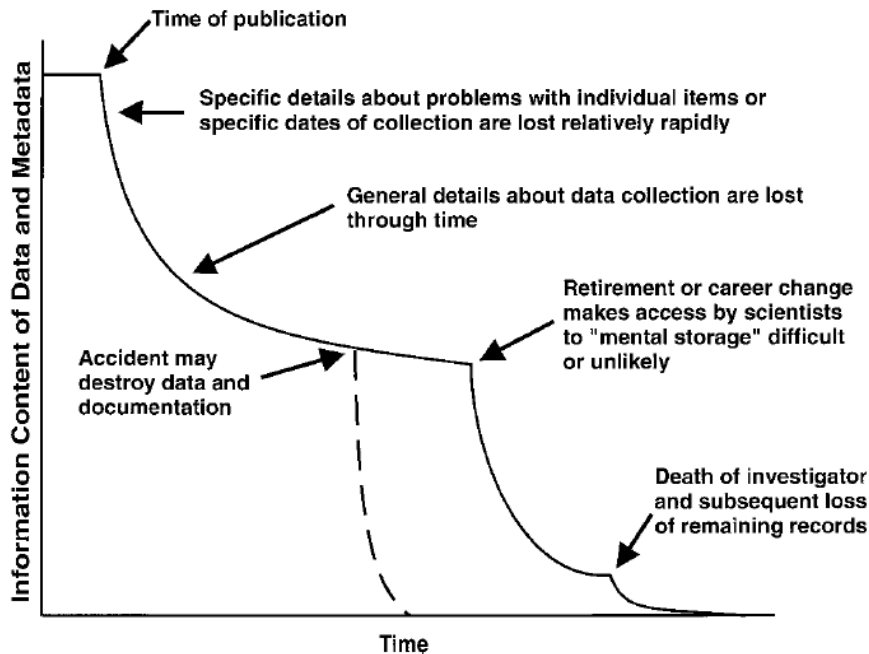
## 1. La perte des données



A data management **horror story** by Karen Hanson, Alisa Surkis and Karen Yacobucci

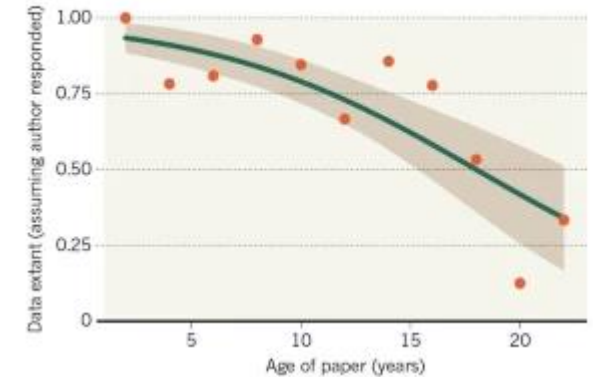
<https://www.youtube.com/watch?v=N2zK3sAtr-4>

516 Ecology papers publiés entre 1991-2011  
17% des données perdues par an !  
80% au bout de 20 ans



### MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



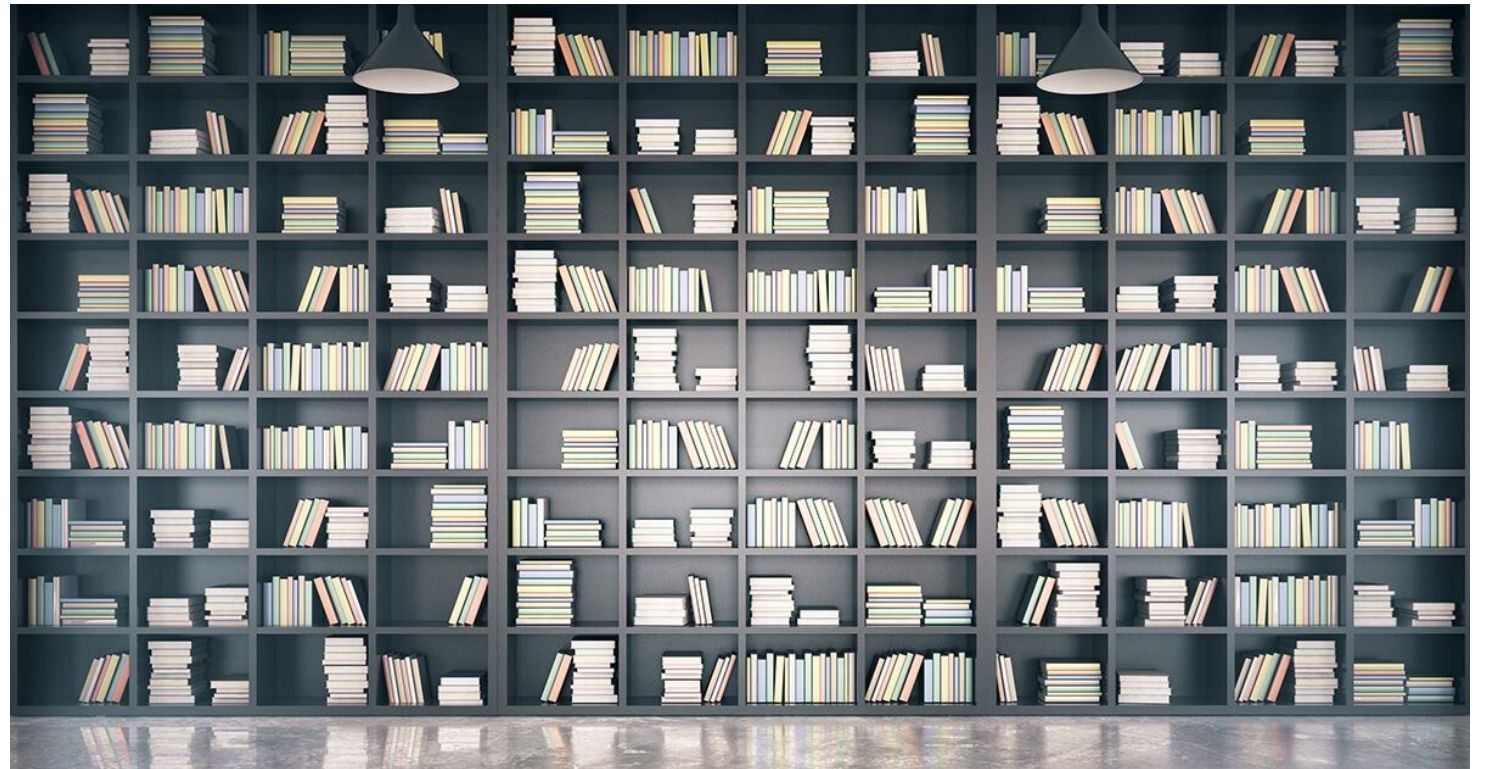
Gibney E. & Van Noorden R. 2013

FIG. 1. Example of the normal degradation in information content associated with data and metadata over time (“information entropy”). Accidents or changes in storage technology (dashed line) may eliminate access to remaining raw data and metadata at any time.

Michener et al. 1997 (non geospatial ecological data)

# CONSTATS

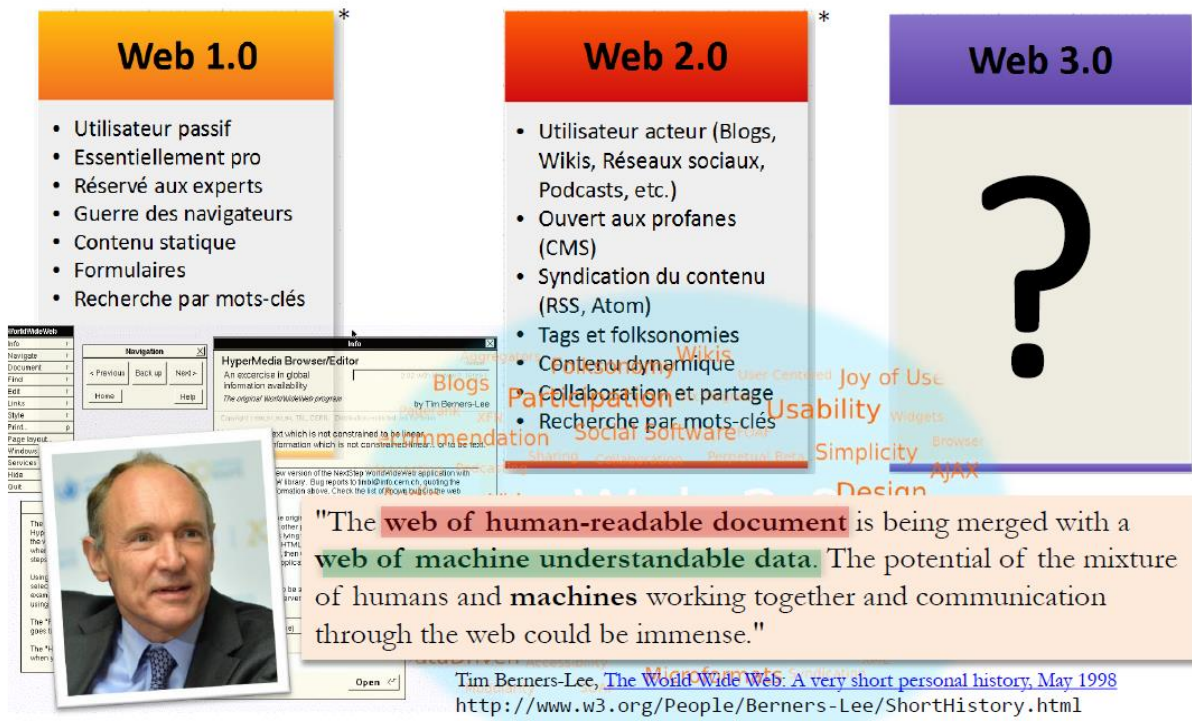
2. Le droit de ressaisir la n-ième fois les informations liées à la données dès lors qu'on les partage



# SCIENCE OUVERTE

Initiée il y a plus de 30 ans à l'échelle internationale grâce au rapide développement du web

## Web evolution



## Web 3.0

- Contenus sémantiques
- Agents intelligents
- Objets communicants
- Intelligence ambiante

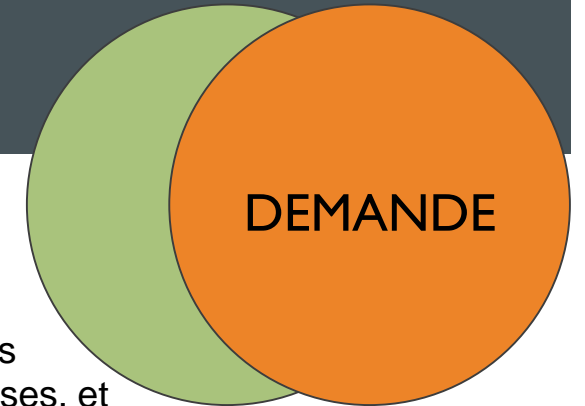
*Assisté par les intelligences artificielles*

FredCavazza.net



# SCIENCE OUVERTE

« ouvert autant que possible, fermé autant que nécessaire »



« les résultats de la recherche scientifiques soient ouverts à tous, chercheurs, entreprises, et citoyens sans entrave, sans délai, sans paiement » F. Vidal 2018

2014-2016

Findable Accessible Interoperable Reusable

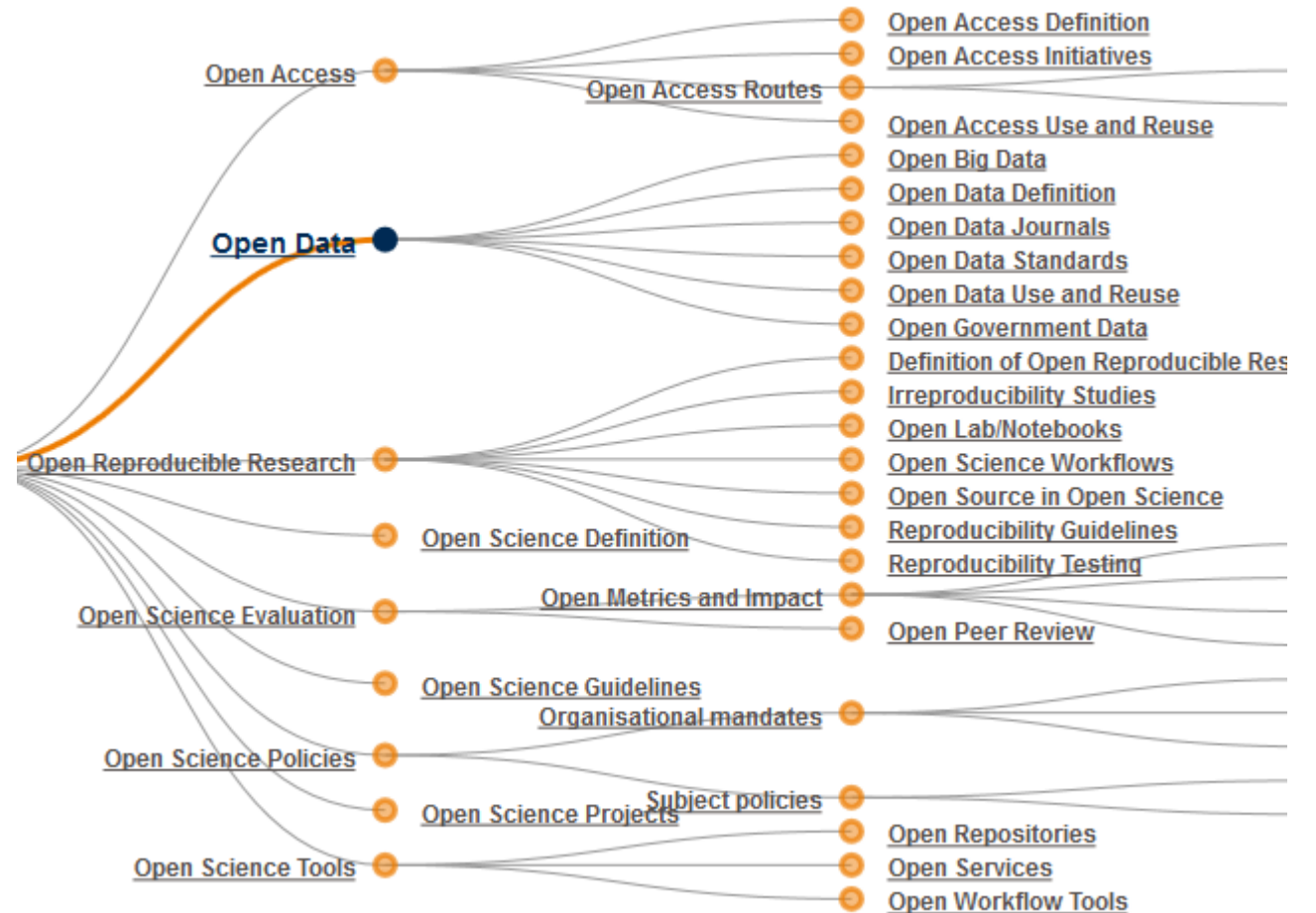
FAIR DATA PRINCIPLES

# SCIENCE OUVERTE

L' Open Data n'est qu'une partie de la science ouverte

Pourquoi des données de la recherche ouverte?

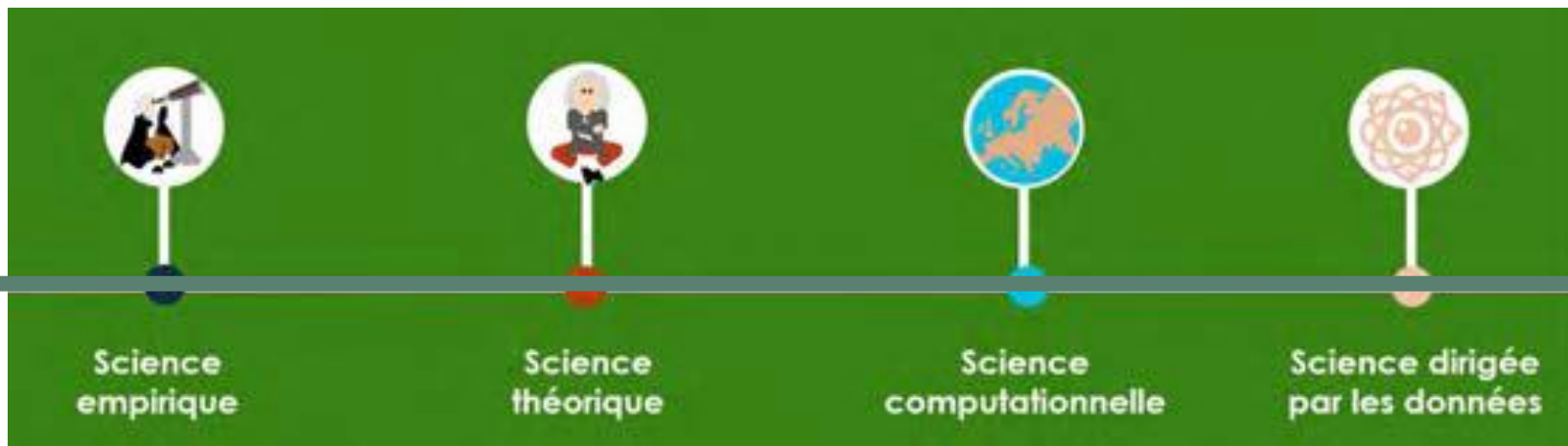
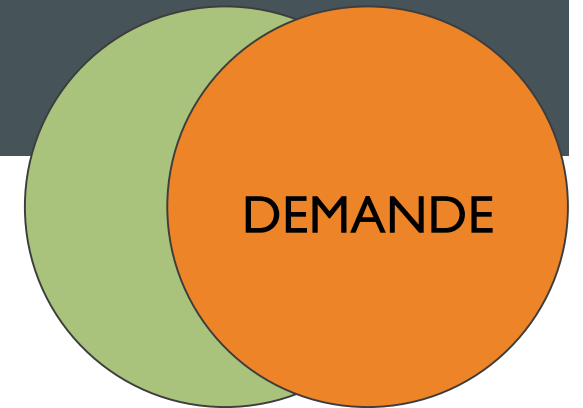
- Capitaliser sur les résultats précédents (qualité)
- Encourager la collaboration, éviter la dispersion des efforts (efficacité)
- Accélérer l'innovation (transfert)
- Impliquer les citoyens et la société (transparence)



# SCIENCE OUVERTE

Il en découle de nouvelles opportunités

- pour les chercheurs, avec le gain en visibilité des données et leur réutilisation;
- pour la science : exploration des données avec de nouvelles méthodes et renforcement de la validation et qualité des données, et de leur reproductibilité
- pour la société : une réutilisation pour un meilleur usage de l'argent public, et impliquer avec plus de transparence les citoyens (ex. sc. participatives)



# SCIENCE OUVERTE pour les IRs

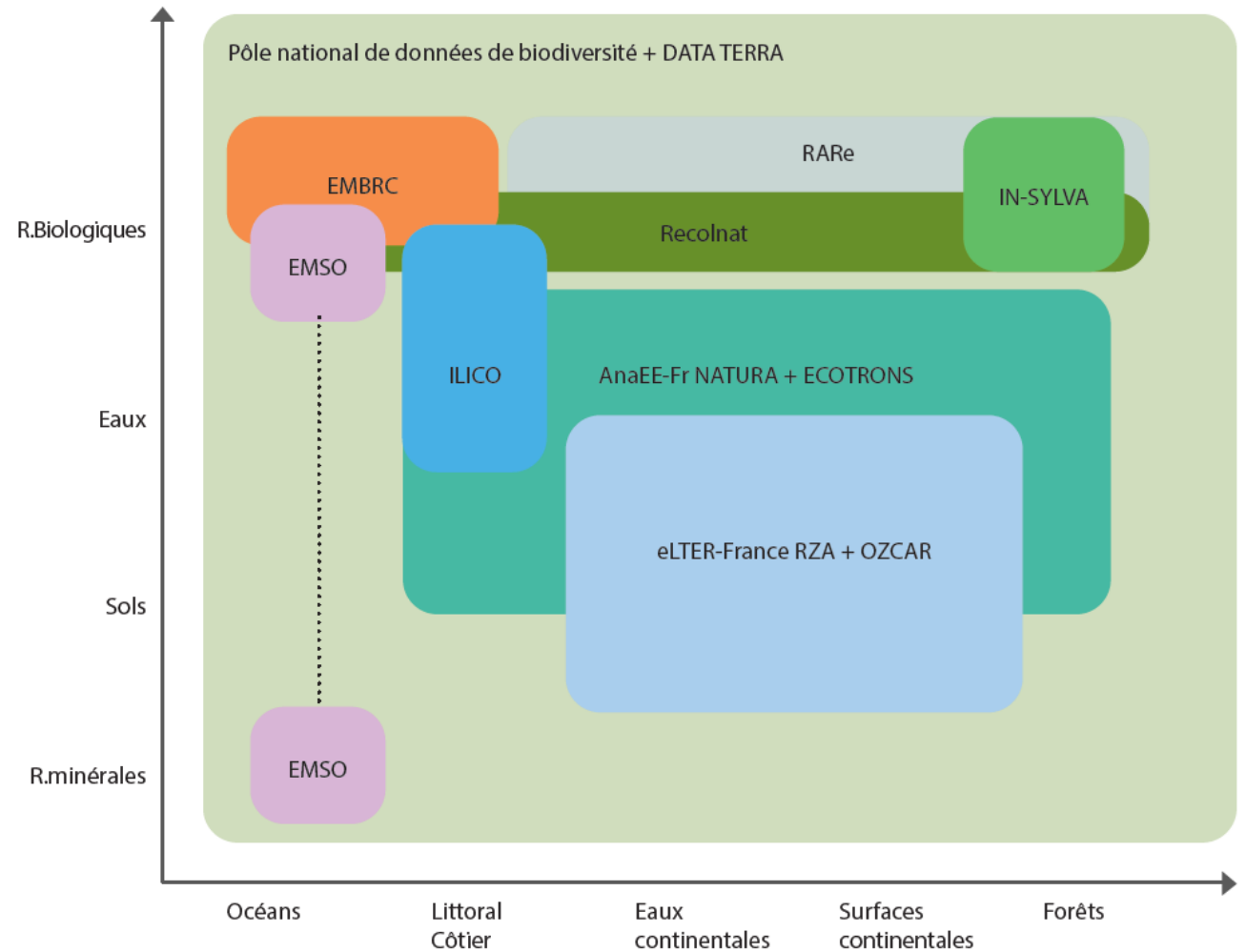
## Feuille de route du CNRS 2019

**Action 3 : soutenir et accompagner les infrastructures de recherche, productrices de données, dans la définition et la mise en œuvre de politiques de données.**

Le CNRS est largement engagé avec ses partenaires dans les Infrastructures de Recherche (IR) nationales et internationales, qui représentent les lieux où se créent et s'analysent les données de la recherche: instruments analytiques, infrastructures de calcul, infrastructures de données, observatoires, etc. Pour généraliser l'application des principes FAIR à toutes les disciplines, le CNRS publiera une charte des infrastructures, engageant celles-ci à respecter les pratiques FAIR et des standards de qualité, en affichant des politiques de données concertées avec les communautés scientifiques utilisatrices des infrastructures concernées. Certaines infrastructures (telles que Progedo et Humanum à l'institut des SHS (INSHS)) sont déjà bien engagées dans

FIGURE 7

PAYSAGE DES INFRASTRUCTURES DE RECHERCHE FRANÇAISES DANS LE DOMAINE DES RESSOURCES NATURELLES. L'ENCADRÉ ORANGE INDIQUE UNE INFRASTRUCTURE COMMUNE AVEC LE DOMAINE BIOLOGIE-SANTÉ. NE SONT PAS MENTIONNÉES LES INFRASTRUCTURES LOGISTIQUES (FOF, ECORD-IODP).



## Domaine Atmosphere

TGIR	ICOS-FRANCE	Système Intégré d'Observation du Carbone/Integrated Carbon Observation System	ICOS (2006)
------	-------------	---	-------------



# SCIENCE OUVERTE pour les IRs

## Livre Blanc Allenvi 2020 sur les IRs - 2020-2030

### Statut de la donnée et articulation entre la recherche scientifique et le monde socio-économique

Les que la valorisation/gestion des bases de données et celles des collections d'échantillons. Concernant la reconnaissance et la prise en compte par la sphère publique des bases de données produites par les infrastructures de recherche, un objectif pourrait être une **accroissement de la production de produits référencés**. Ceci impliquerait néanmoins d'une part, la

### Recommandation 15

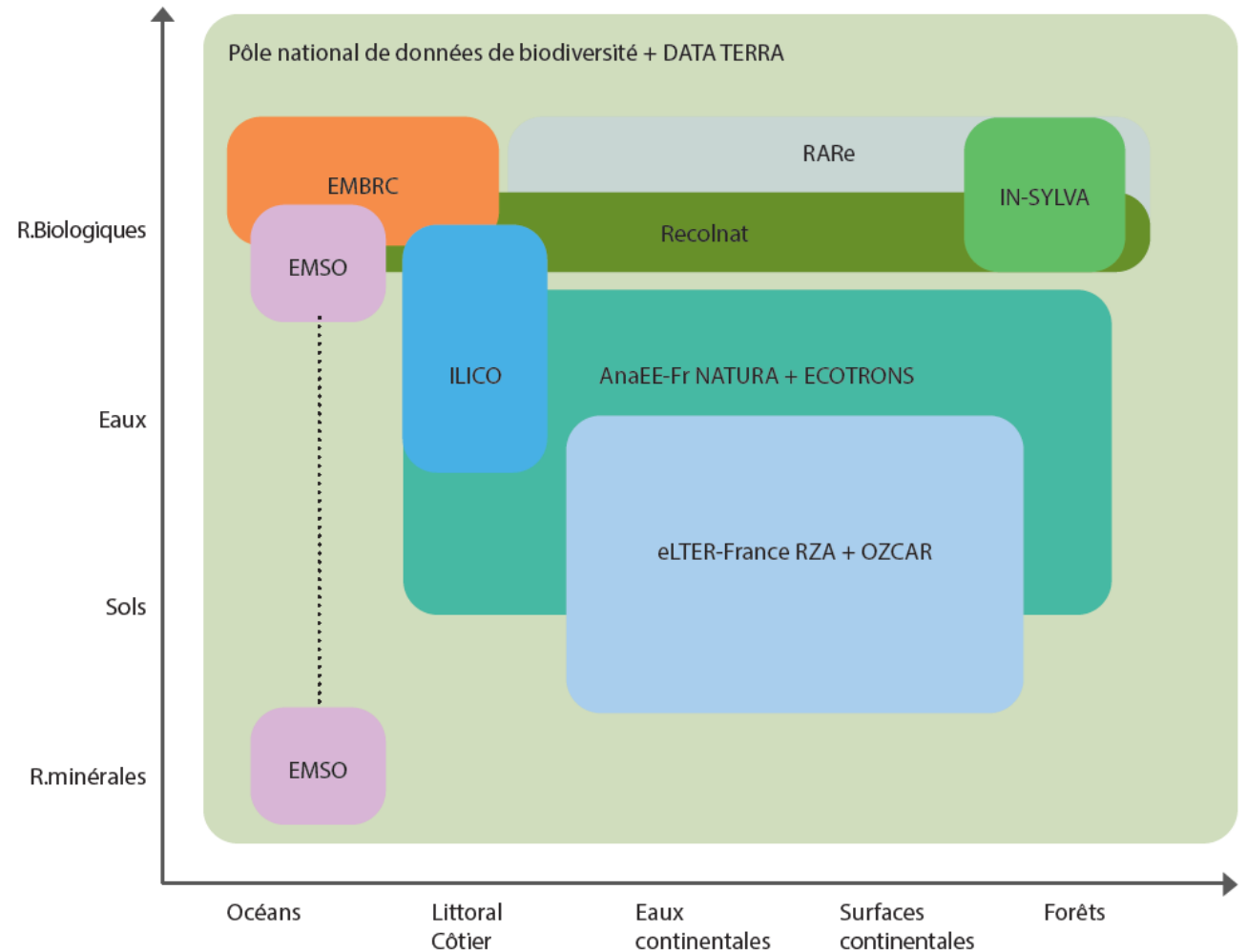


**Conduire une réflexion nationale sur le modèle économique de la donnée de recherche dans le domaine sciences du système Terre et de l'environnement.**

ET E-ENVIR21 - 2 AU 5 NOVEMBRE 2021

FIGURE 7

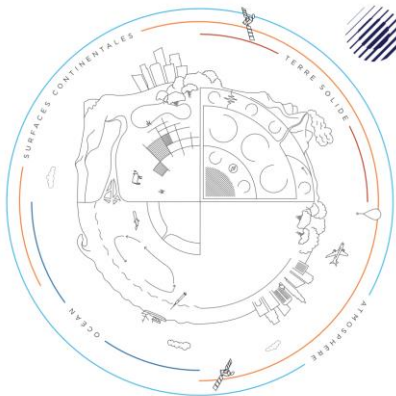
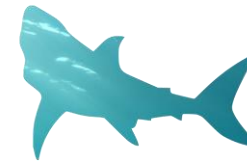
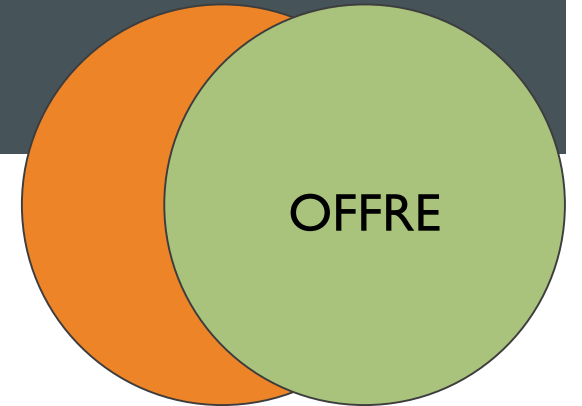
PAYSAGE DES INFRASTRUCTURES DE RECHERCHE FRANÇAISES DANS LE DOMAINE DES RESSOURCES NATURELLES. L'ENCADRÉ ORANGE INDIQUE UNE INFRASTRUCTURE COMMUNE AVEC LE DOMAINE BIOLOGIE-SANTÉ. NE SONT PAS MENTIONNÉES LES INFRASTRUCTURES LOGISTIQUES (FOF, ECORD-IODP).



### Domaine Atmosphere

TGIR	ICOS-FRANCE	Système Intégré d'Observation du Carbone/Integrated Carbon Observation System	ICOS (2006)
------	-------------	---	-------------

# MULTIPLICATION DES OUTILS, DES ACTEURS (1/2)



LICENCE OUVERTE  
OPEN LICENCE



Callisto

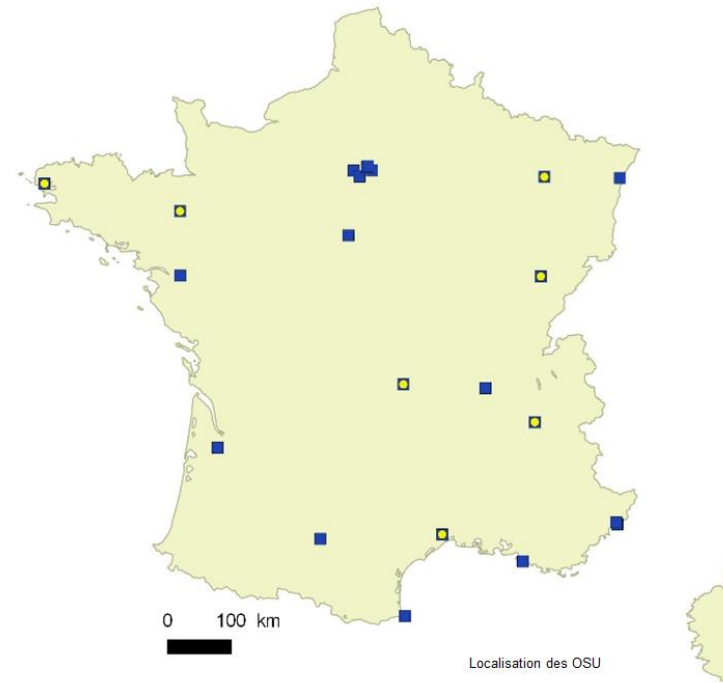
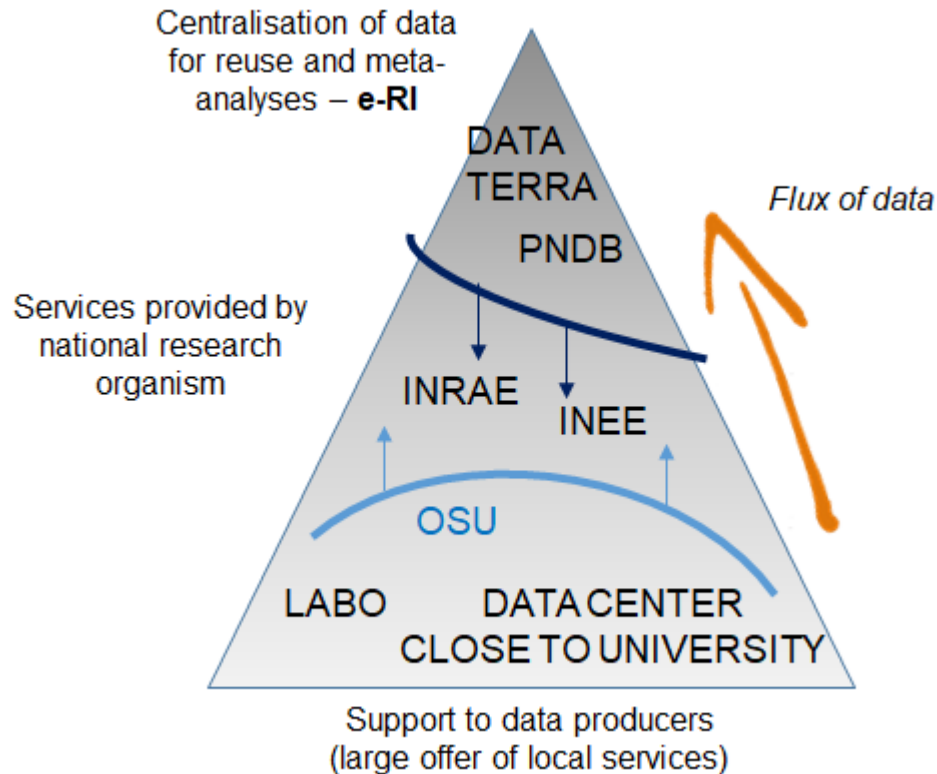
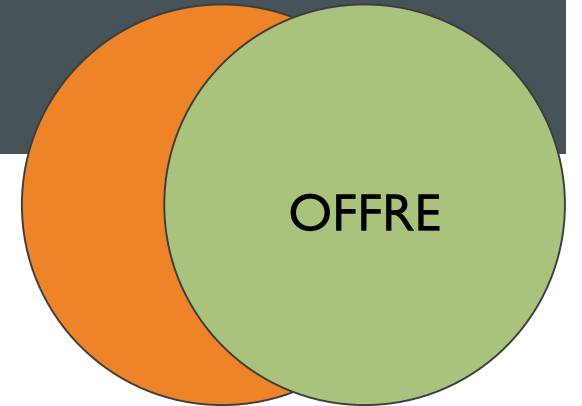
Proposed by CALMIP



PANGAEA.



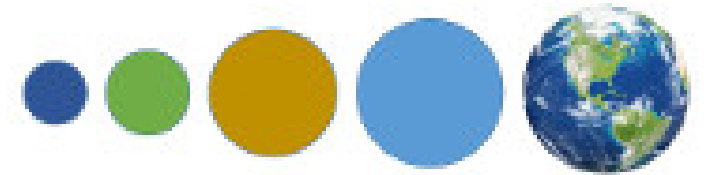
# MULTIPLICATION DES OUTILS, DES ACTEURS (2/2)



Distribution of local data center (blue square) ; concomitant use by CZO and LTSER (yellow and blue square)

# E-ENVIR 2019 - Gif-sur-Yvette

E-Envir 2019 Ecole thématique CNRS



<https://e-envir.sciencesconf.org/>

Coordonnée par ANAEE & UMS BBEES

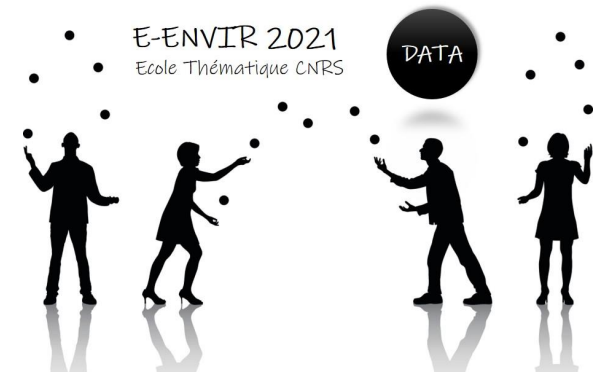
Orienté sur la technique pour un public qui travaille sur les portails de données

# E-ENVIR 2021 - Strasbourg

<https://e-envir-21.sciencesconf.org/>

Coordonnée par RZA, OZCAR/THEIA, PNDB, UMS BBEES

Orienté sur la pratique pour un public qui souhaite partager ses données dans des portails de données



# EQUIPE ORGANISATRICE

Wilfried



Emilie



David



Isabelle



Charly



Isabelle



Eric



Floriane



Maison Interuniversitaire des  
Sciences de l'Homme  
- Alsace -



Virginie



Véronique



Regis



Sylvie



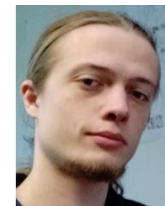
Françoise



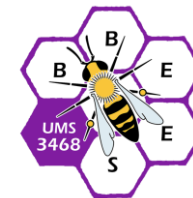
Cécile



Yvan



Elie



Amandine

# LES OBJECTIFS DE E-ENVIR 2 I



Vous situer dans le paysage des offres



## Producteurs des données

- Acquisition des données
- Qualification des données
- Bancarisation des données / SI local

## SI des IRs (OZCAR, RZA) et eIR (PNDB, Data Terra)

- Centralisation les données
- Homogénéisation et standardisation des données
- Outils de visualisation, voire d'analyse
- Proposition de vocabulaires contrôlés
- Fourniture de statistiques sur l'utilisation des données
- Aide à l'attribution de DOI

## Utilisateurs des données et services (humains ou machines) via

- mail (personne contact) (humains)
- portail de données (humains)
- web services (machines)

# LES OBJECTIFS DE E-ENVIR 2 I

Vous situer dans le paysage des offres

Vous initier aux principes FAIR et aux niveaux de FAIRisation

Vous guider par la pratique avec les recommandations actuelles des IRs OZCAR, RZA, PNDB et de l'UMS BBEES

Vous traduire les éléments techniques clés en piochant dans nos expériences et les nombreuses ressources existantes

En revanche, l'idée n'est pas de proposer des solutions individuelles ou d'étudier des **cas particuliers**.

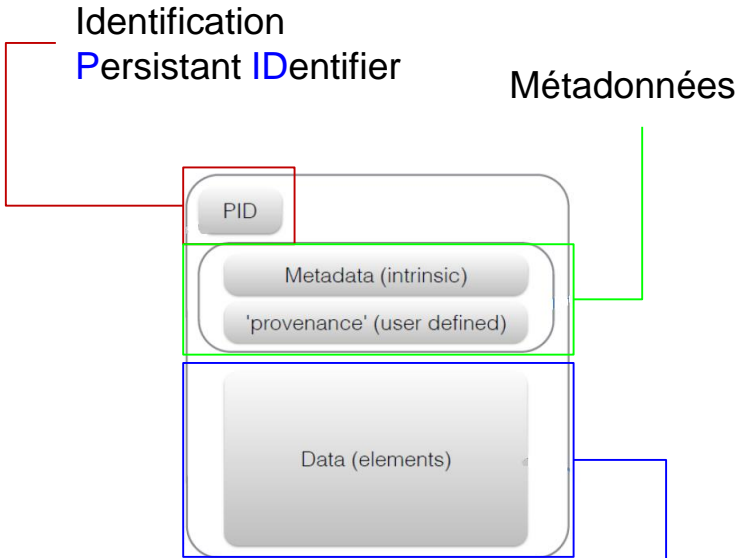


<https://www.ands.org.au/working-with-data/fairdata/training>



# VOTRE PROGRESSION

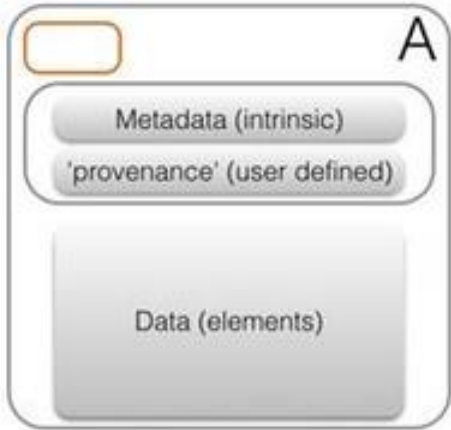
## Niveaux de FAIRisation



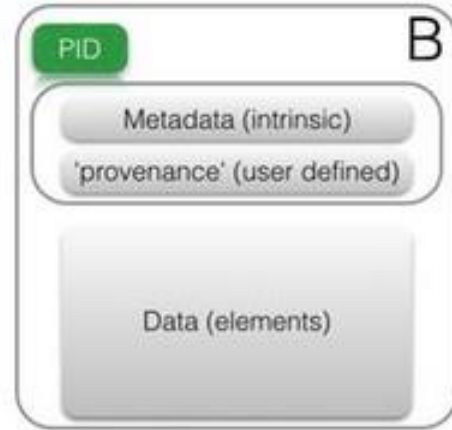
Produits de la recherche : article, jeux de données, vidéo, carte logiciel

## Data as increasingly FAIR Digital Objects

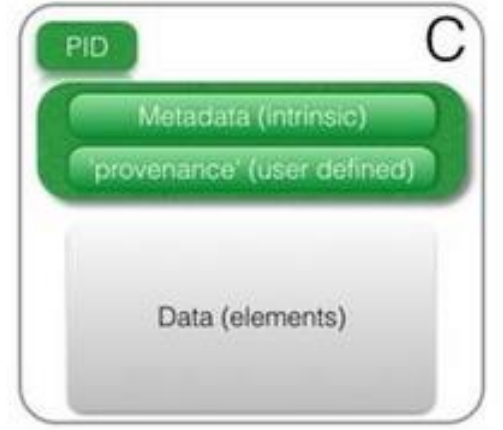
Re-useless data (80%)



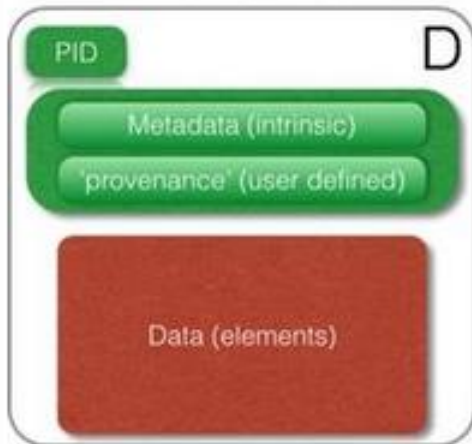
Findable



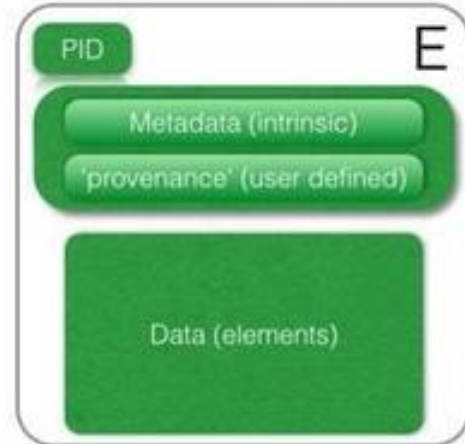
FAIR metadata



FAIR data-restricted access



FAIR data-Open Access



FAIR data-Open Access/Functionally Linked



# PROGRAMME

## Des sessions en plénières

1. présentant les outils
2. sur la préparation des jeux de données à la FAIRisation
3. sur F & A du FAIR
4. sur I & R du FAIR
5. sur les spécificités de certaines données (notamment SHS) et les services PUD
6. une vision stratégique en guise de conclusion

Des TD pour utiliser pas à pas les outils proposés, avec les jeux de données mis à disposition et les vôtres et partager les bonnes pratiques.

Ma	Ven. 05
08:00	
09:00	
10:00	
11:00	
12:00	
13:00	
14:00	<b>Ouverture et introduction générale à l'Ecole Thématique</b>
15:00	<b>TD - Présentation de vos jeux de données</b>
16:00	<b>Pause</b> <b>Animation</b>
17:00	<b>Introduction aux flux de données et aux outils</b>
18:00	<b>Debriefing</b>

# LOGISTIQUE

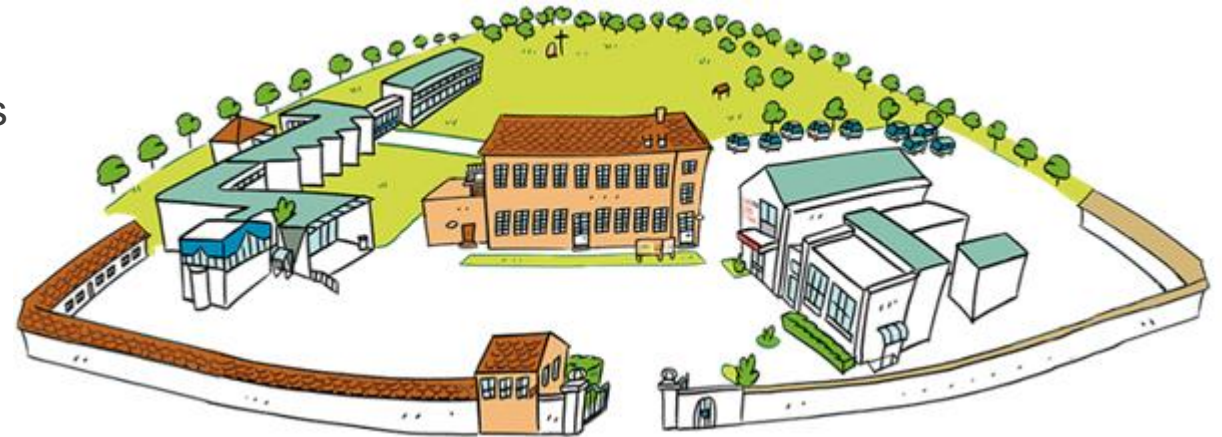
**Hébergements** : vous disposez de chambres personnelles accessibles à partir de 15h, et qui doivent être libérées vendredi matin à 8h30

**Repas** : !! respect des horaires

- petit déjeuner 7h45,
- déjeuner 12h15,
- dîner 18h30

pour vendredi, revoir qui souhaite un lunch box pour vendredi midi (j'ai compté 21 repas et 1 lunch)

**Notes de frais** : conserver vos billets de train/avion, ticket de repas...



mercredi et jeudi  
soir apéritif offert  
à 18h30 au  
préalable du  
repas



# POINTS ADMINISTRATIF & FINANCIER

Pour **les personnes du RZA**, merci de remplir et signer la **fiche état de frais**, même si le montant de vos dépenses est égal à 0 euros, et de la renvoyer par mail à Carole DIDIERJEAN ([carole.didierjean@univ-smb.fr](mailto:carole.didierjean@univ-smb.fr)). Ne pas oublier de joindre à cette fiche le scan des pièces justificatives (qui seront à envoyer par voie postale ultérieurement).

Pour **les invités**, idem mais à renvoyer à Peggy RUPP ([peggy.rupp@unistra.fr](mailto:peggy.rupp@unistra.fr)).

**ME METTRE EN COPIE!**

## Pour le bilan de l'ET

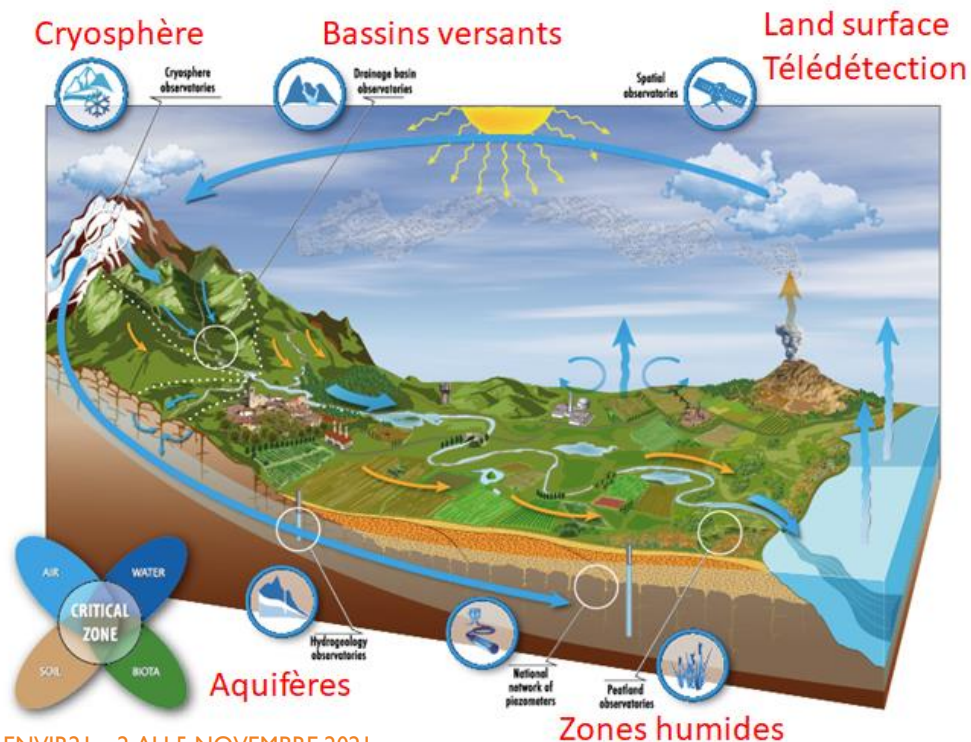
- A fournir dans un délai court (fin novembre)
  - Merci de vous assurer auprès de vos services administratifs que les BC ont été transmis puis, que le paiement ait été effectué
  - Merci de répondre au **Questionnaire d'évaluation** fourni à l'issue de l'école thématique, dont les résultats anonymisés sont impératifs pour le bilan.



# Présentation du réseau eLTER France

# FOCUS sur l'IR OZCAR : un réseau de sites de long terme densément instrumentés

« Un réseau d'observatoires de long terme pour l'observation, la compréhension et la simulation intégrée de l'évolution de la Zone Critique et de ses divers compartiments dans l'Anthropocène, ouvert à la communauté scientifique et au-delà »



**Mots clés:** Ressources en eau, sol (quantité et qualité); risques naturels (crues, sécheresses, avalanches, contamination); Territoires



# PROGRAMME

Des sessions de mise au point “CLAIR/PAS CLAIR” à la fin de chaque journée

*10' chacun note ce qui est clair/pas clair.*

*10' partage de ce qui est clair (une idée chacun; parole concise).*

*10' partage de ce qui n'est pas clair.*

*15' de séance retour (priorité des questions; report de certaines si on n'a pas de temps)*

Deux sessions ICEBREAKER pour faciliter les échanges entre vous = partage d'expérience !

Votre participation :

- active !
- toutefois, réserver certaines questions pour les TD si focus sur un élément précis de votre jeu de données ou encore à la fin de journée pendant le temps de débriefing

Notre écoute :

- active ...et donc un prgm en partie modulable



# FOCUS sur l'IR OZCAR : un réseau de 21 observatoires



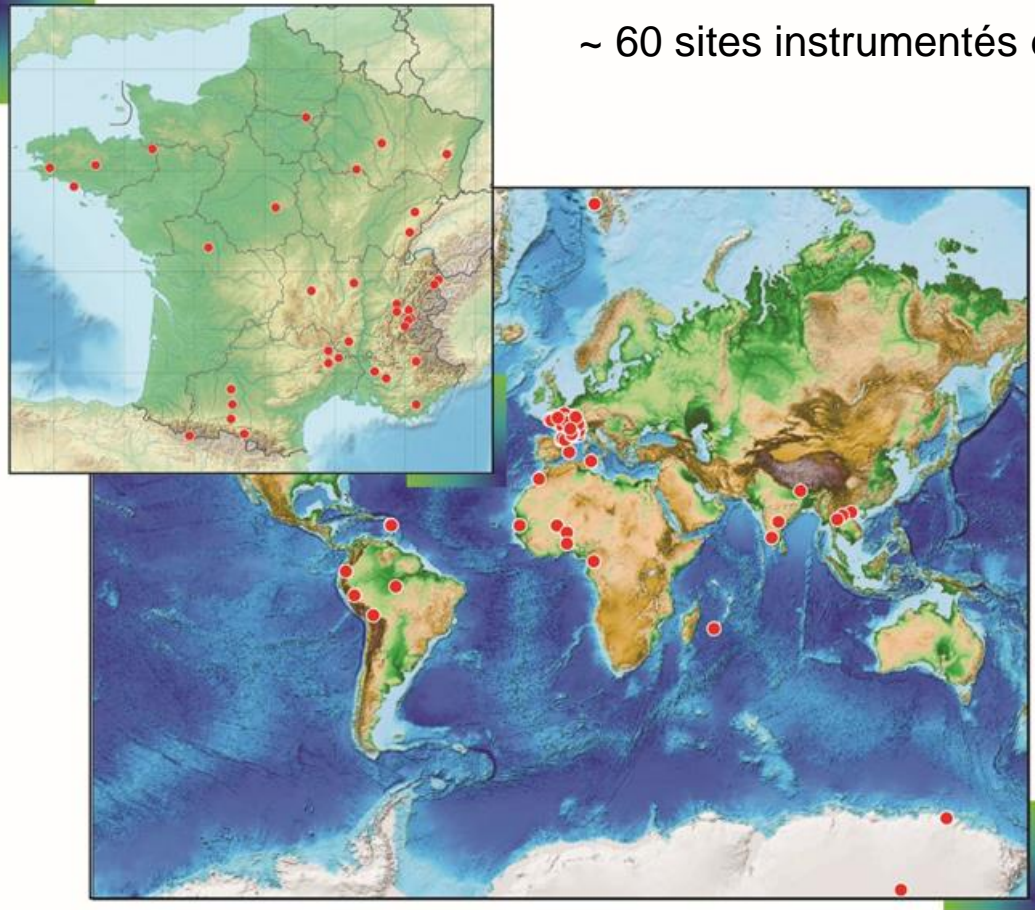
- Rassemblés autour d'un intérêt commun: la zone critique et l'établissement de bilans et le calcul de flux
- La promotion d'une approche systémique de la zone critique
- Des sites hautement instrumentés sur le long terme
- Une question scientifique fédératrice: comment monitorer, décrire simuler l'adaptation des zones critiques à l'heure de l'anthropocène



# FOCUS sur l'IR OZCAR :

## Les objets d'intérêt et les variables mesurées

~ 60 sites instrumentés en France et dans les pays du Sud



### Observations

- Transport de l'eau, des sédiments et solutés
- Bilans d'énergie de surface
- Exploration géophysique
- Occupation des sols

### Objets d'intérêt

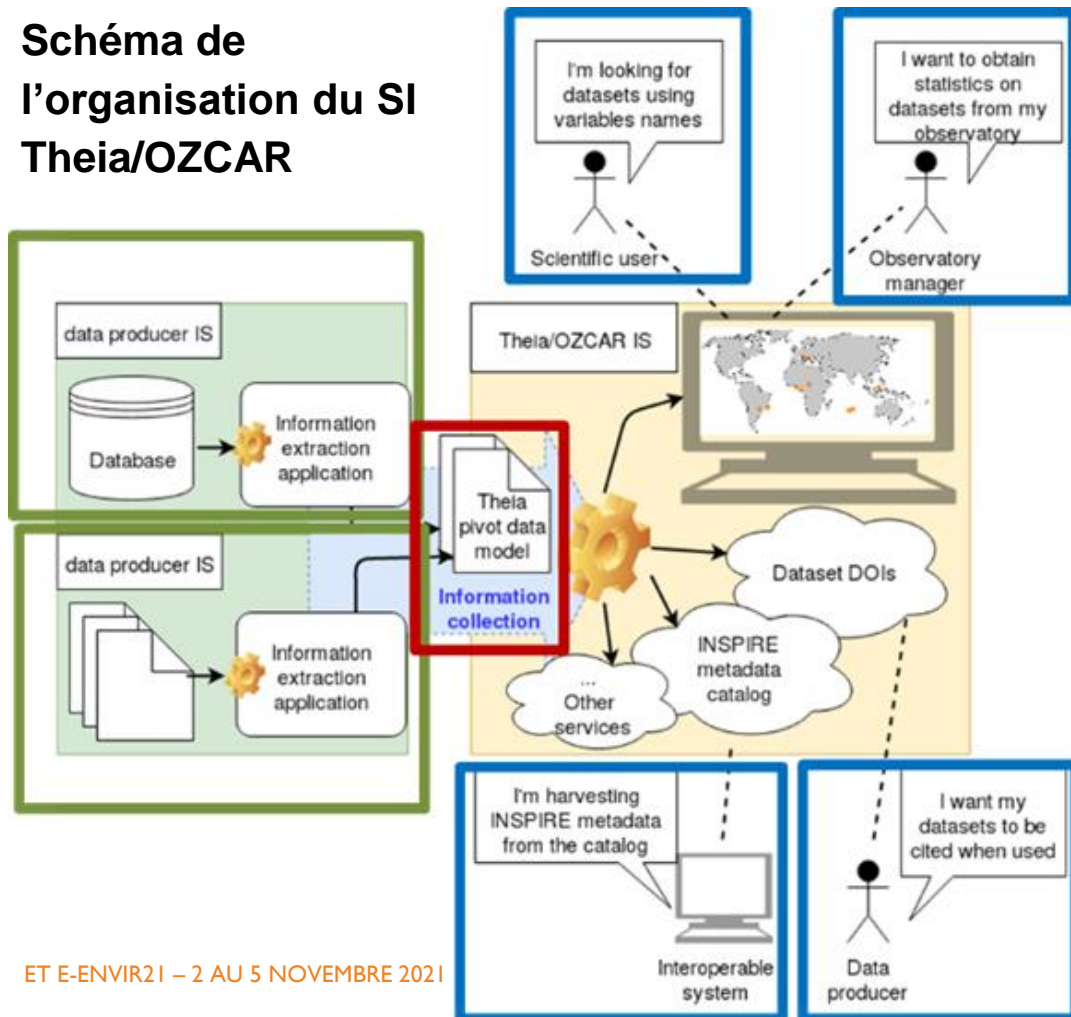
- Bassins versants, rivières
- Aquifères
- Glaciers
- Permafrost

**Une grande diversité de mesure in-situ (> 300 var. mesurées)**

- Séries temporelles (météo, hydrologie, glaciologie, chimie eau, ....)
- Carottes de sol
- Coupe géophysique 2D
- Cartes : occup. sol, MNT, param. géophysique sols

# FOCUS sur l'IR OZCAR : Le Système d'Information Theia/OZCAR

## Schéma de l'organisation du SI Theia/OZCAR

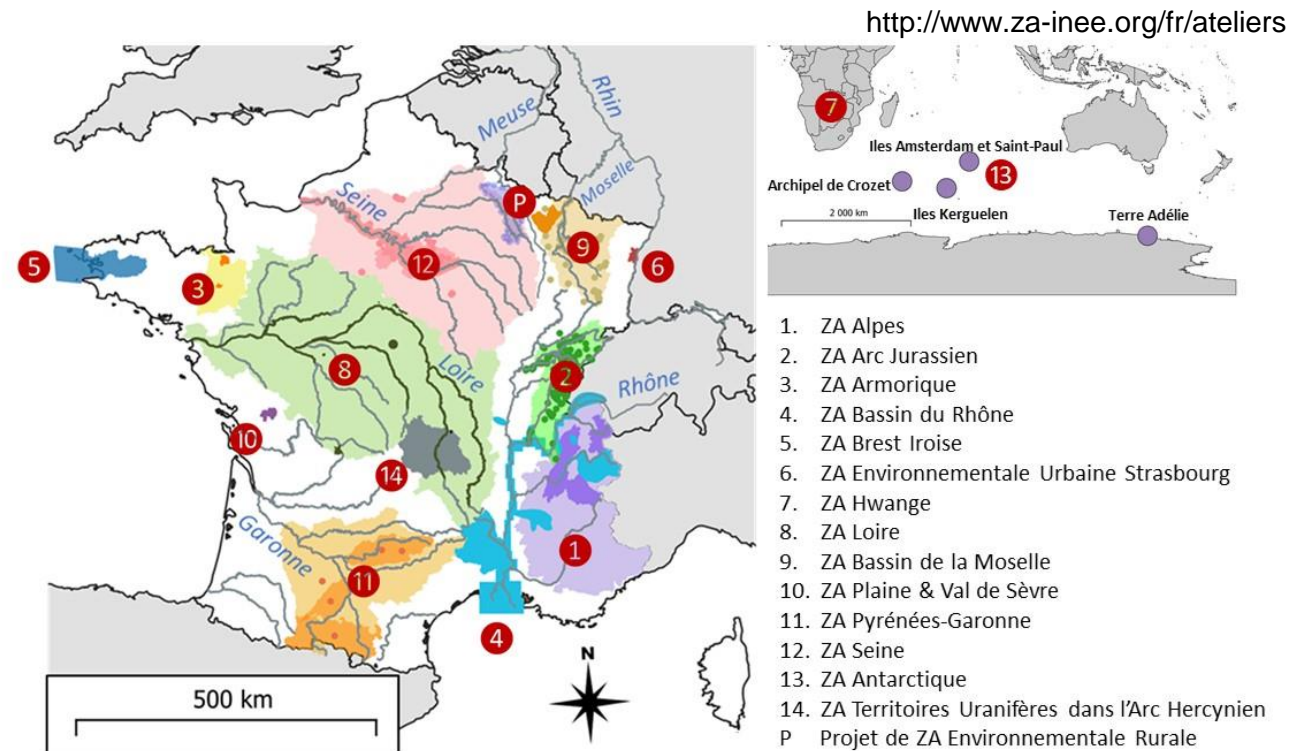
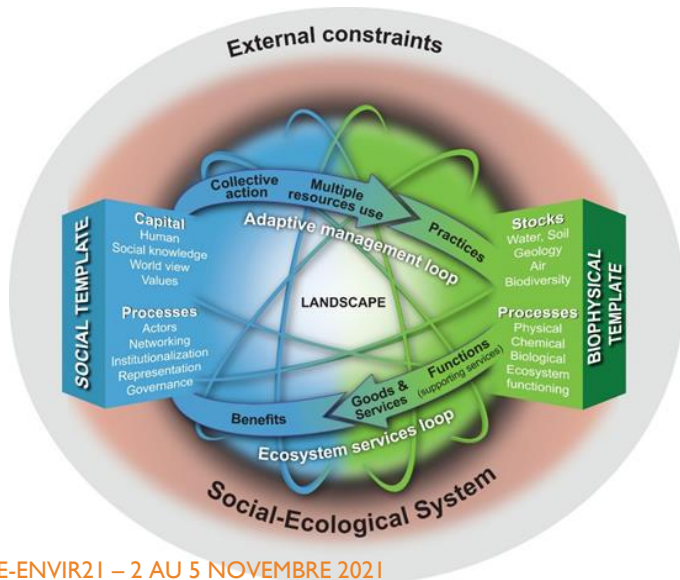


- + Ecriture du Data Management Plan
  - +/- Découverte des données via le portail OK mais uniquement 10 observatoires sur 21 visibles
  - + Un thésaurus FAIR et publié sur le web
  - + Un service de catalogage implémenté et les métadonnées sont visibles sur eLTER\_DIP
- => **Prochaine étape: téléchargement des données**

# FOCUS sur l'IR RZA : un réseau de sites de long terme inter- et trans-disciplinaire

Les Zones Ateliers sont des dispositifs de recherches sur le long terme

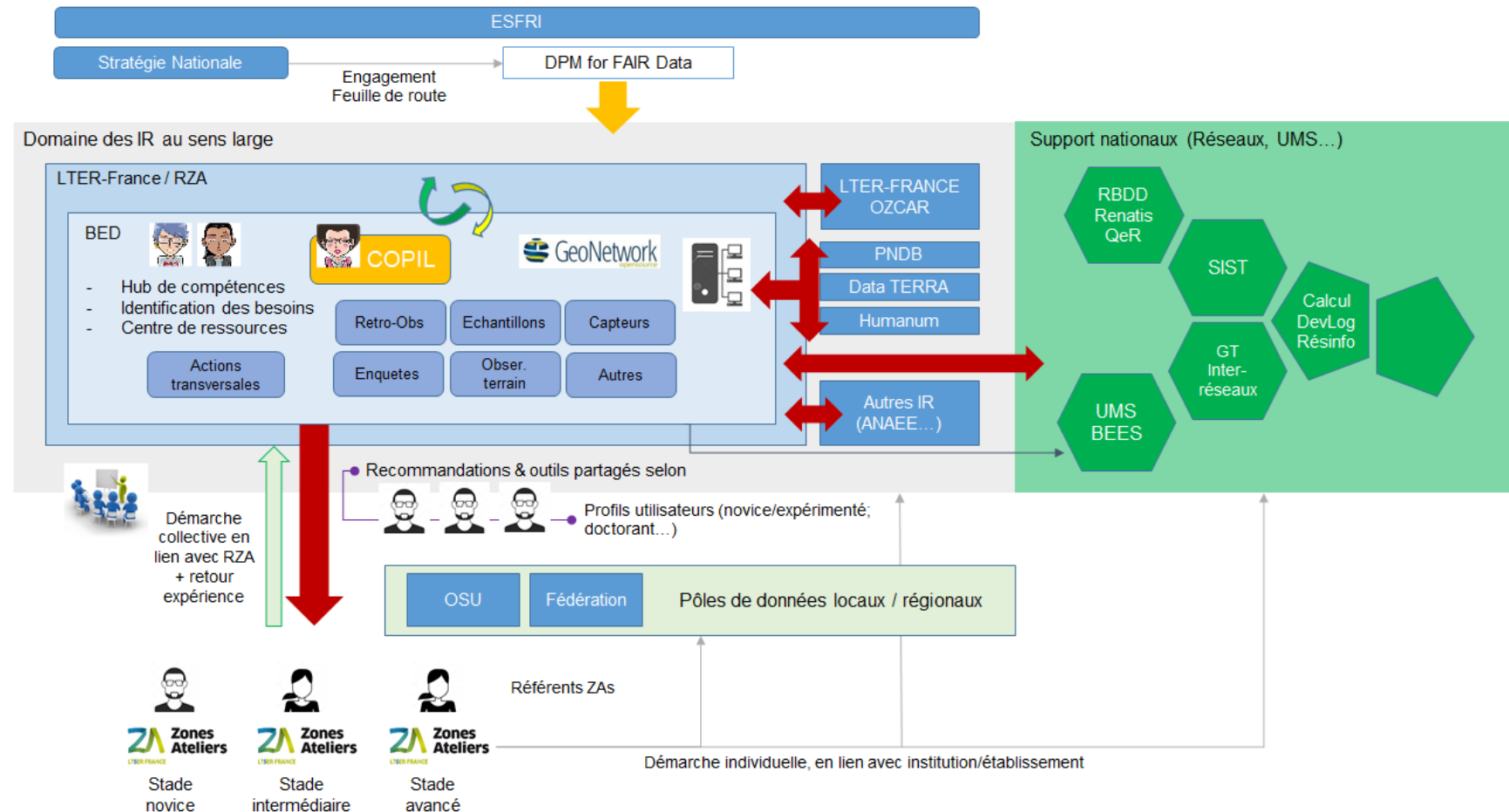
- inter et transdisciplinaire / ancré sur des territoires ;
- construit autour d'une **unité fonctionnelle** (un fleuve et son bassin versant, les paysages...)
- observations et expérimentations réalisées à l'échelle des sites ateliers



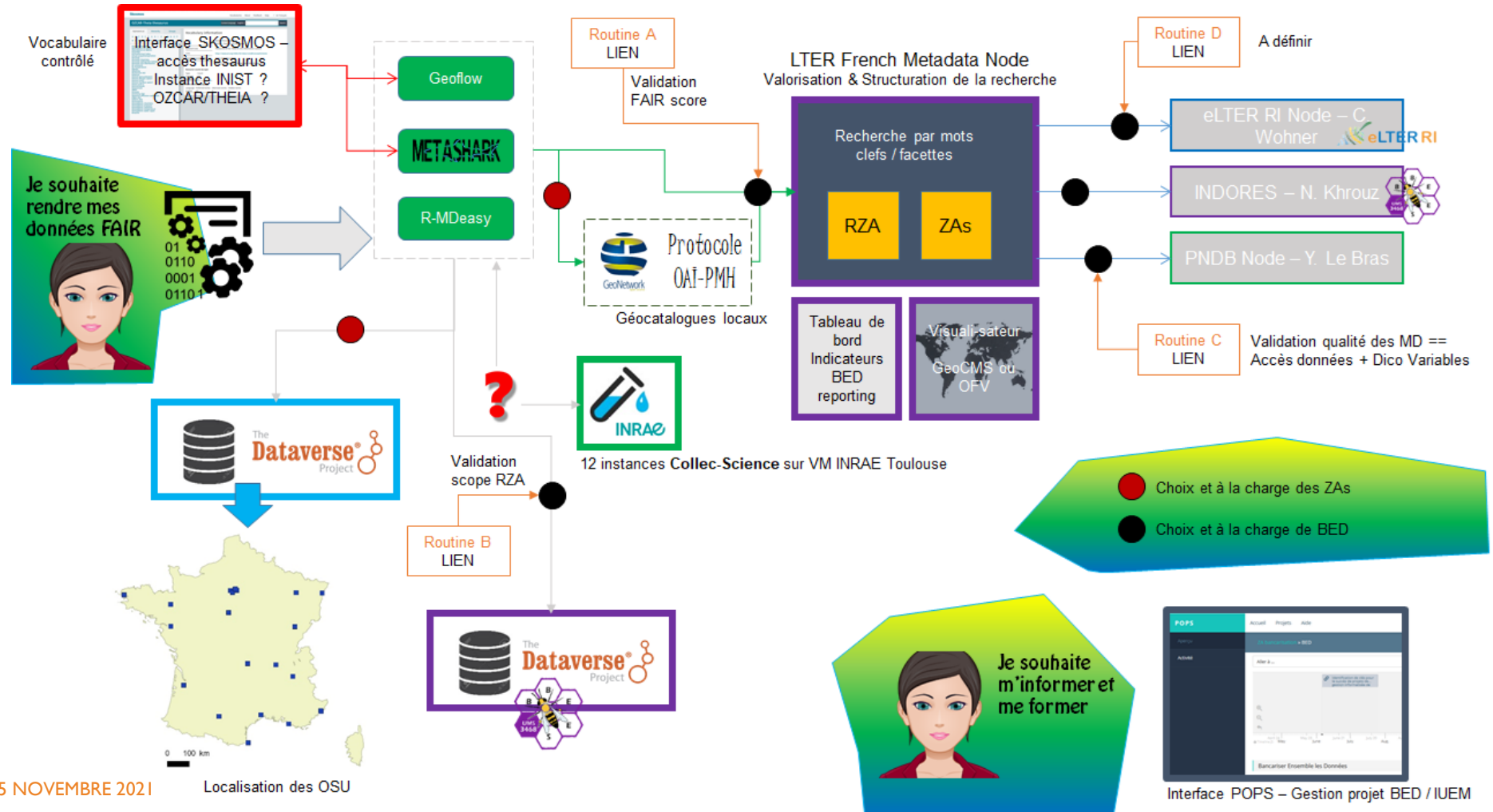
14+1 ZAs et > 80 sites ateliers

# FOCUS sur l'IR RZA : BED un collectif en appui à la gestion des données

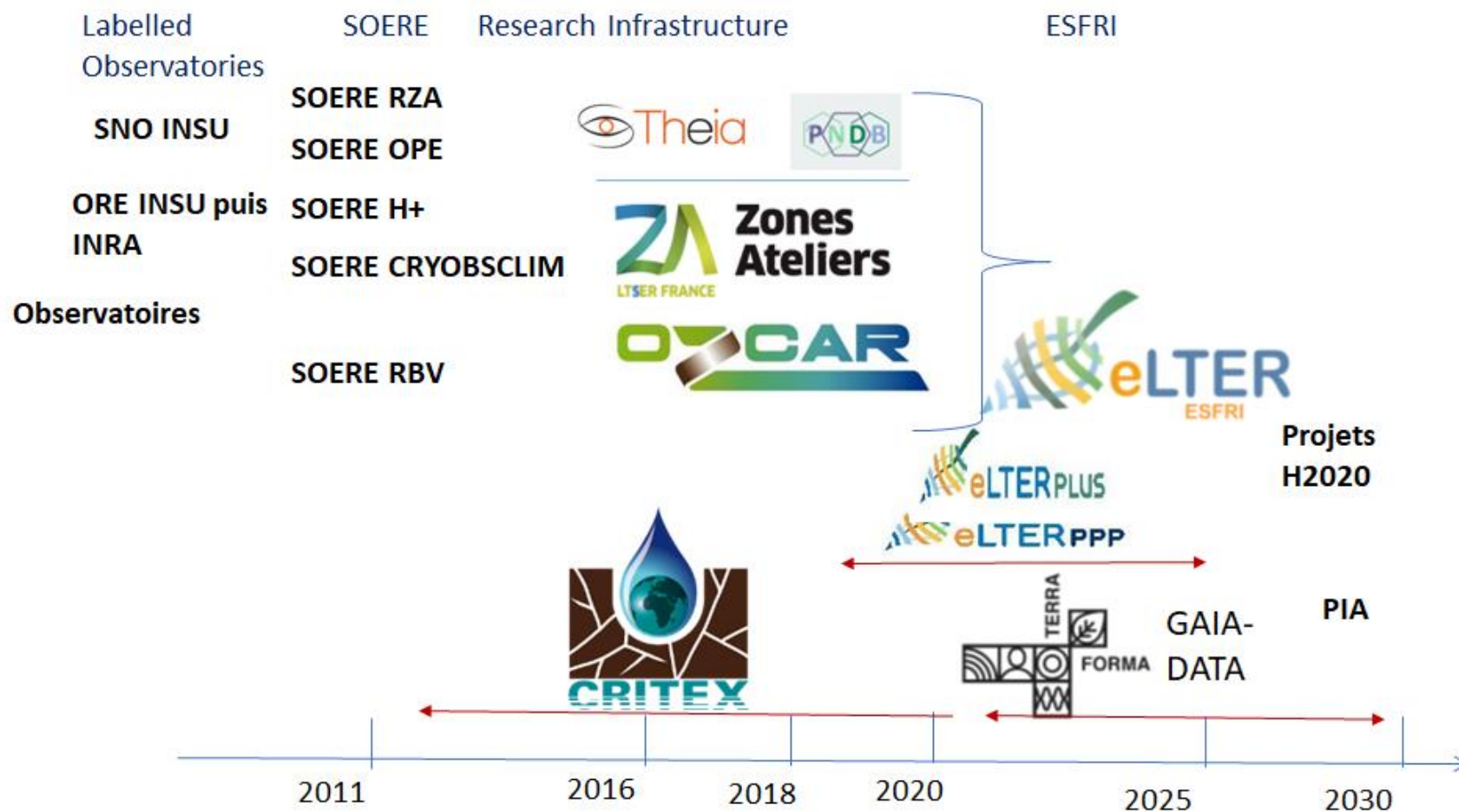
- un centre de ressource, qui rassemble des compétences dans le domaine
- en appui à la direction du RZA (reporting, stratégie sur les données)
- complémentaire à ce qui existe en local
- aux interfaces avec les IRs et e-IRs
- en relation avec les réseaux de données



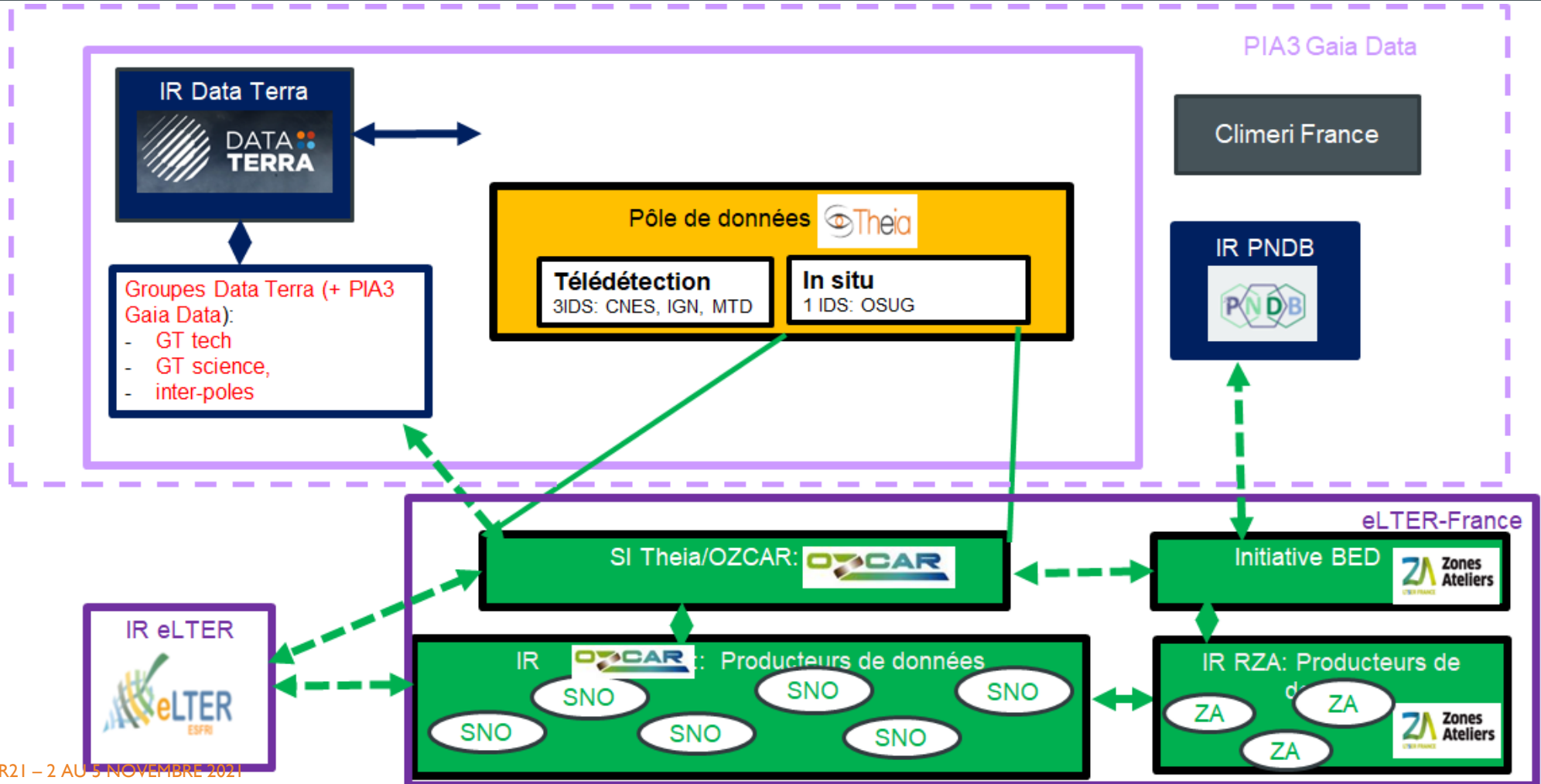
# FOCUS sur l'IR RZA : SI du RZA (en construction)



# FOCUS sur l'IR OZCAR & RZA : Structuration nationale et internationale - eLTER RI et Equipex



# Le paysage de la gestion des données



# REMERCIEMENTS

cnrs

- CNRS soutien financier
  - R. OMNES (conseillère formation, DR10) suivi de la formation
  - P. RUPP (ICUBE), C. DIDIERJEAN (EDYTEM) gestion administrative et financière
  - Le comité organisateur : BED/RZA, OZCAR/THEIA, UMS BBEES, PNDB, avec la participation de la MISHA
  - Centre Saint-Thomas, hébergement
- 
- Retrouvez les informations sur : <https://e-envir-21.sciencesconf.org/>



Maison Interuniversitaire des  
Sciences de l'Homme  
- Alsace -





# RÉFÉRENCES



Diapo 1. Cartoon: "Data Management« . P. Hochstenbach. 2015. <https://www.fosteropenscience.eu/content/cartoon-data-management> (consulté le 24 janvier 2022)

Diapo 2. (1) Data Sharing and Management Snafu in 3 Short Acts. <https://www.youtube.com/watch?v=N2zK3sAtr-4> (consulté le 24 janvier 2022). (2) Gibney, E., Van Noorden, R. Scientists losing data at a rapid rate. *Nature*(2013). <https://doi.org/10.1038/nature.2013.14416https://www.nature.com/articles/nature.2013.14416>. (3) Michener et al. 1997 Nongeospatial metadata for the ecological sciences. *Ecological Applications*, 7(1), pp. 330–342

Diapo 4. (1) P. Genoud & D. Zibelin. Introduction au websémantique. E-envir19. Gif-sur-Yvette.

(2) V. Bardot <https://valbardot.wordpress.com/2013/12/02/evolution-du-web/> (consulté le 24 janvier 2022)

Diapo 5. (1) Guillaume Monnain / Akenium. BD Antoine Petit & la feuille de route CNRS. <https://www.inist.fr/nos-actualites/science-ouverte-le-cnrs-se-dote-dune-feuille-de-route/> (consulté le 24 janvier 2022) (2) site internet OUVRIER LA SCIENCE <https://www.ouvrirlascience.fr/plan-national-pour-la-science-ouverte/>

Diapo 6. Des outils pratiques pour gérer et valoriser vos données de recherche Formation URFIST Bordeaux 17 mai 2018 Atelier 3 : « Décrire et documenter ses données avec des métadonnées standardisées ». Martine Barale CIRAD. Schéma : <https://www.fosteropenscience.eu/foster-taxonomy/open-science>

Diapo 7. The Fourth Paradigm: Data-Intensive Scientific Discovery, T. Hey, S. Tansley, and K. Tolle, Eds. Redmond, VA: Microsoft Research, 2009, ISBN 978-0-9825442-0-4, <http://fourthparadigm.org>

Diapo 8. Feuille de route du CNRS 2019 et Livre Blanc Allenvi 2020 sur les IRs - 2020-2030.

Diapo 16. FAIR data training <https://www.andis.org.au/working-with-data/fairdata/training>

Diapo 18. F. André (DIST/CNRS). Les données de la recherche il y a beaucoup à dire ...et à FAIR! URFIST Rennes. 18/05/2017 à Rennes

Diapo 22-31. Ressources Irs OZCAR et RZA