

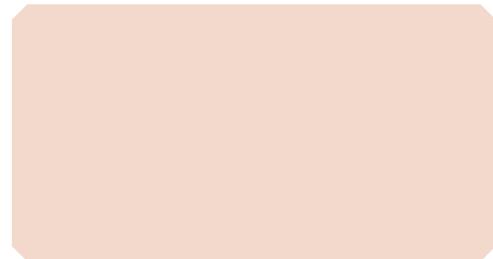
Gestion pérenne des données scientifiques



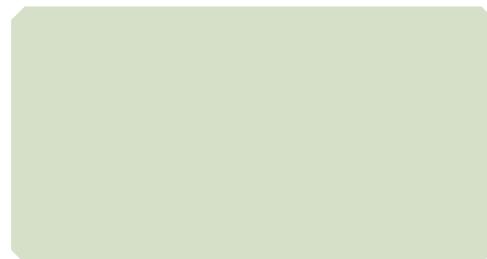
DYNAFOR
UMR 1201 | INRA - INP TOULOUSE

CATI
SIOEA

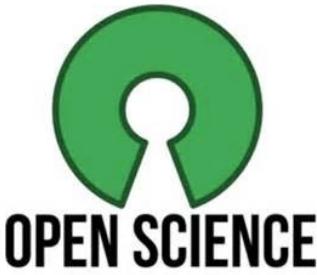
Du Plan de Gestion des données au Datapaper



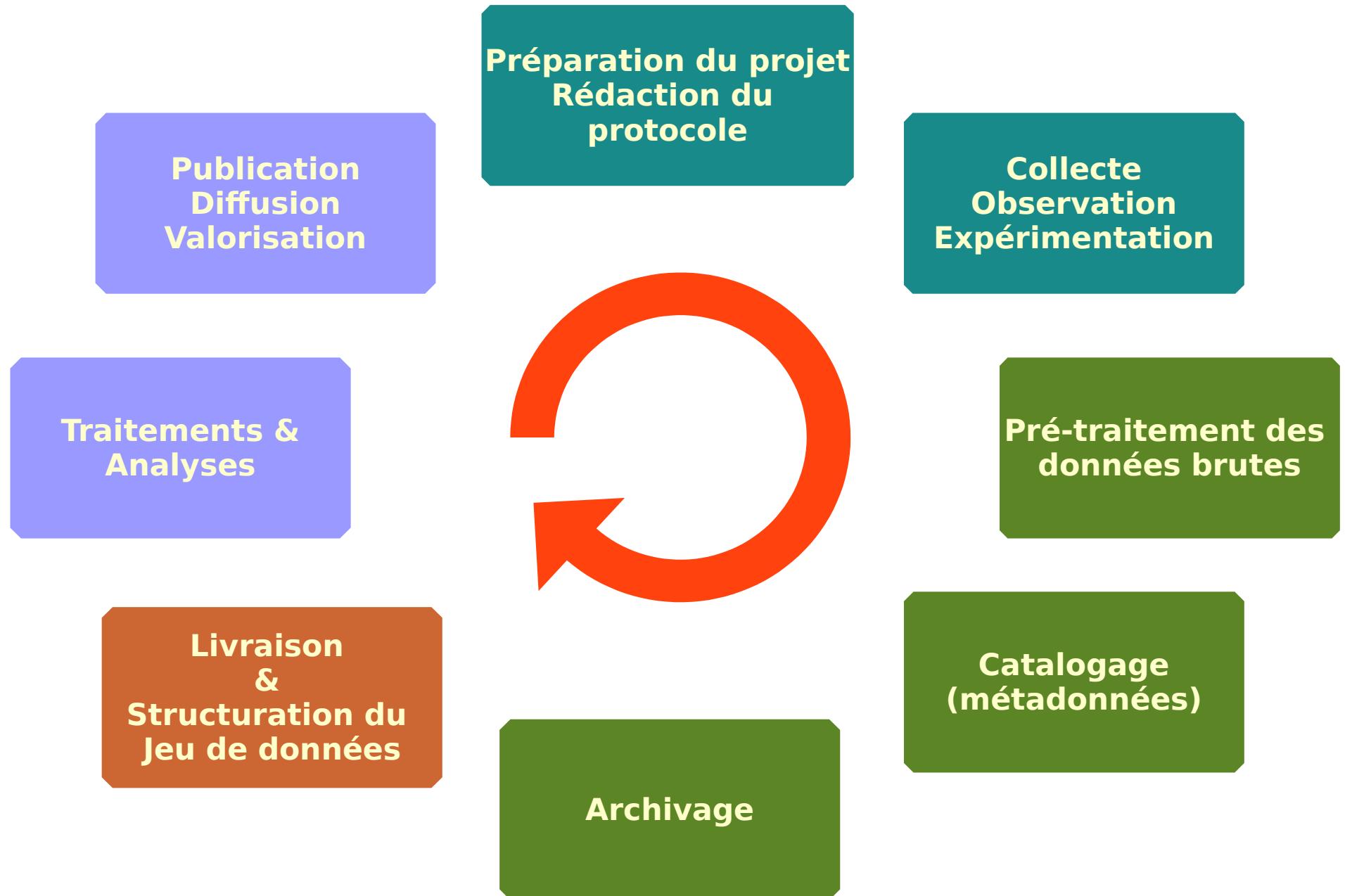
Mars 2018



Contexte général



Le cycle de vie des données



Une gestion à chaque étape

Les objectifs poursuivis (\neq atteints !)

- Traçabilité de toutes les données
- Accès aux données facilité pour tous les agents de l'UMR
- « Chaîne de la donnée » (= sans rupture)
- Standardisation → Ouverture → Valorisation
- Stockage = fil conducteur

Préparation du projet
Rédaction du
protocole

Pour garantir la réussite de la campagne de collecte

- Quels objectifs poursuivis ?
- Quel cadre scientifique ?
- Quel type de données à collecter ?
- Qui fait quoi ?
- Mode opératoire ...

PLANIFIER

Préparation du projet
Rédaction du
protocole

Dialogue entre tous les acteurs + documents de références

- Meilleure anticipation
- Premières métadonnées



PLANIFIER

Préparation du projet
Rédaction du
protocole

Dialogue entre tous les acteurs + documents de références

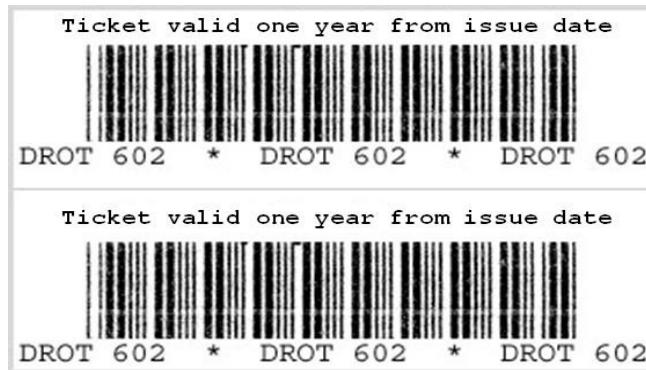
- Meilleure anticipation
- Premières métadonnées
 - ⇒ Protocole écrit et diffusé à l'avance
(Sharepoint)
 - ⇒ Plan de gestion des données
(Opidor)



PLANIFIER



Collec-Science

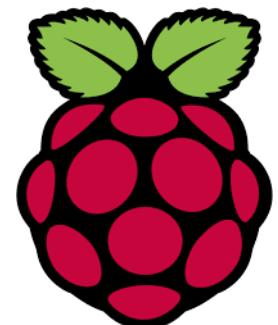


Préparation du projet
Rédaction du
protocole

Collecte
Observation
Expérimentation

Etape cruciale mais souvent négligée

- Bonne collecte ssi protocole complet et disponible
- Contrôle de la qualité *in situ*
- Gestion des échantillons



ACQUÉRIR





Collec-Science

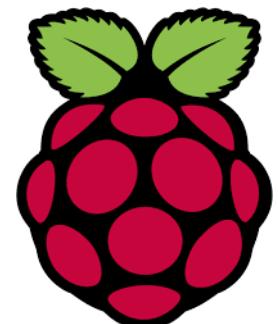


Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Etape cruciale mais souvent négligée

- Bonne collectessi protocole complet et disponible
- Contrôle de la qualité *in situ*
- Gestion des échantillons
- ⇒ Apport des outils embarqués pour initier le flux des données :
Assistance à la saisie – BDD alimentée sur le terrain



ACQUÉRIR





Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Vérifications post-terrain

- Contrôle de l'intégrité des données
- Éliminations des valeurs aberrantes

Pré-traitement des données brutes

Contrôler



Préparation du projet
Rédaction du protocole

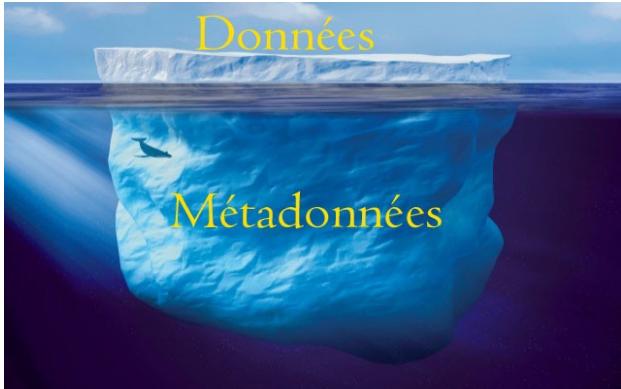
Collecte
Observation
Expérimentation

Vérifications post-terrain

- Contrôle de l'intégrité des données
- Éliminations des valeurs aberrantes
- ⇒ Scripts R, OpenRefine, ETL ...

Pré-traitement des données brutes

Contrôler



Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

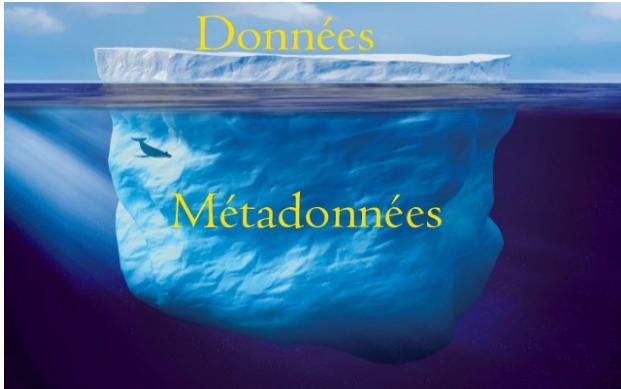
Garantir la pérennité du jeu de données

- Décrire les données et leur contexte
- Faciliter la réutilisation
- Normaliser pour une compréhension commune

Pré-traitement des données brutes

Catalogage (métadonnées)

DÉCRIRE



Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Garantir la pérennité du jeu de données

- Décrire les données et leur contexte
- Faciliter la réutilisation
- Normaliser pour une compréhension commune
- ⇒ Geonetwork, IPT, Librairies R

Pré-traitement des données brutes

Catalogage (métadonnées)

DÉCRIRE



Préparation du projet
Rédaction du protocole



DataCenter

Collecte
Observation
Expérimentation

Stocker le jeu de données sur le long terme

- Sauvegarder les données brutes
- Permettre l'accès futur aux données
- Garantir l'intégrité de la ressource

Pré-traitement des données brutes

STOCKER



Archivage

Catalogage (métadonnées)



Préparation du projet
Rédaction du protocole



DataCenter

Collecte
Observation
Expérimentation

Stocker le jeu de données sur le long terme

- Sauvegarder les données brutes
- Permettre l'accès futur aux données
- Garantir l'intégrité de la ressource
 ⇒ Datacenter

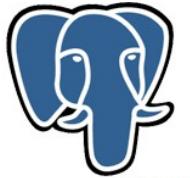


STOCKER

Archivage

Pré-traitement des données brutes

Catalogage (métadonnées)



Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Intégrer les données dans le Sys. d'information

→ Bases de données partagées

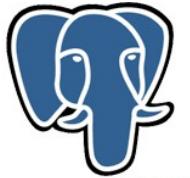
Livraison
&
Structuration du
Jeu de données

ORGANISER

Pré-traitement des
données brutes

Catalogage
(métadonnées)

Archivage



Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Intégrer les données dans le Sys. d'information

- Bases de données partagées
- ⇒ Infrastructure de Données Géographiques

Pré-traitement des données brutes

ORGANISER

Livraison & Structuration du Jeu de données

Catalogage (métadonnées)

Archivage



Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Traitements « scientifiques »

→ Statistiques, SIG, algorithmes, interprétation ...

Traitements &
Analyses

ANALYSER

Pré-traitement des
données brutes

Livraison
&
Structuration du
Jeu de données

Catalogage
(métadonnées)

Archivage



Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Traitements « scientifiques »

→ Statistiques, SIG, algorithmes, interprétation ...

⇒ WPS, publication de scripts

Traitements &
Analyses

ANALYSER

Pré-traitement des données brutes

Livraison &
Structuration du Jeu de données

Archivage

Catalogage (métadonnées)



GLOBAL
BIODIVERSITY
INFORMATION
FACILITY



Publication
Diffusion
Valorisation

Préparation du projet
Rédaction du
protocole

Collecte
Observation
Expérimentation

Traitements &
Analyses

Pré-traitement des
données brutes

Livraison
&
Structuration du
Jeu de données

PARTAGER

→ Mettre en oeuvre la réutilisation

Archivage

Catalogage
(métadonnées)



GLOBAL
BIODIVERSITY
INFORMATION
FACILITY



Publication
Diffusion
Valorisation

Préparation du projet
Rédaction du
protocole

Collecte
Observation
Expérimentation

Traitements &
Analyses

Pré-traitement des
données brutes

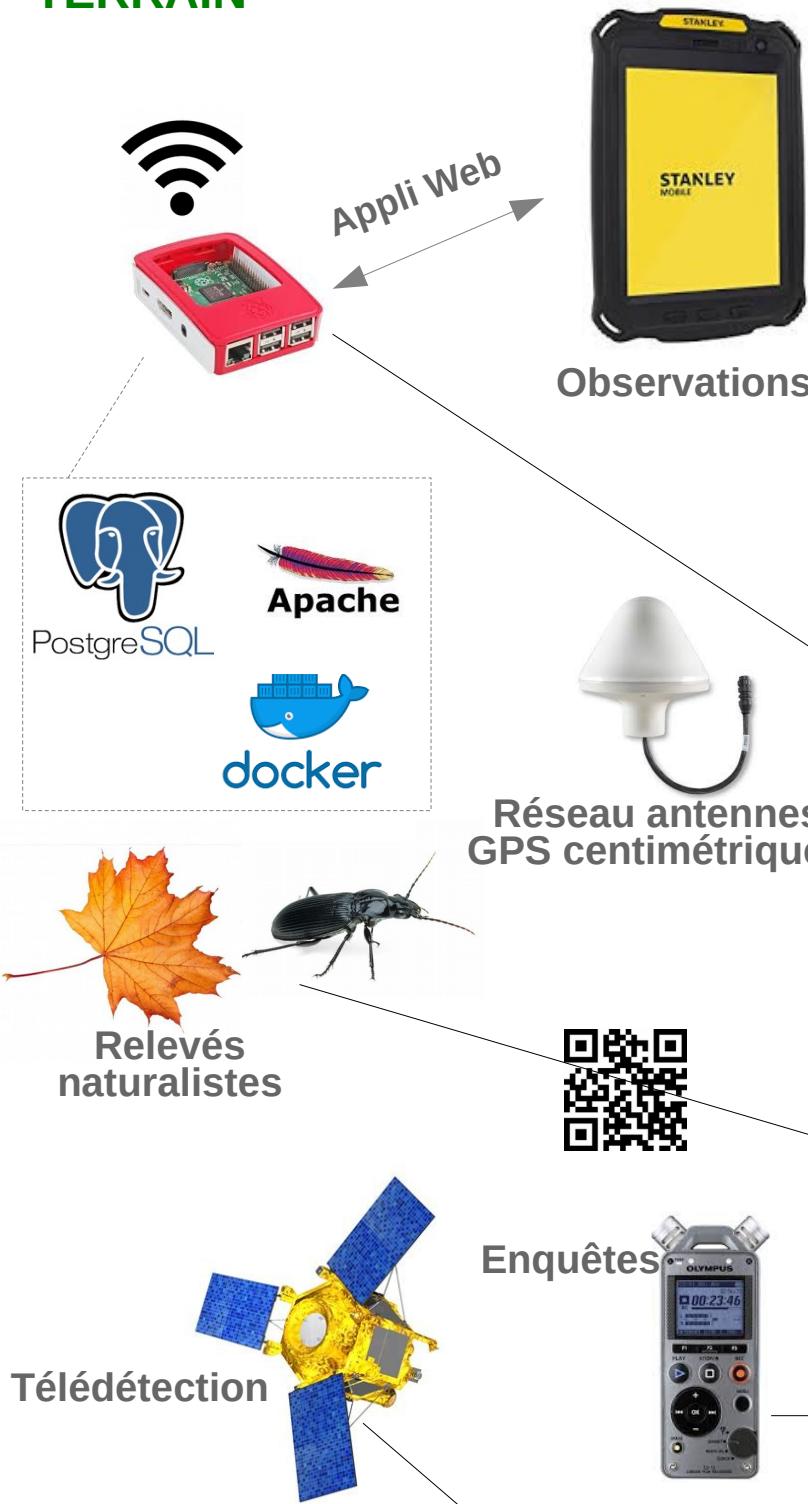
Livraison
&
Structuration du
Jeu de données

Archivage

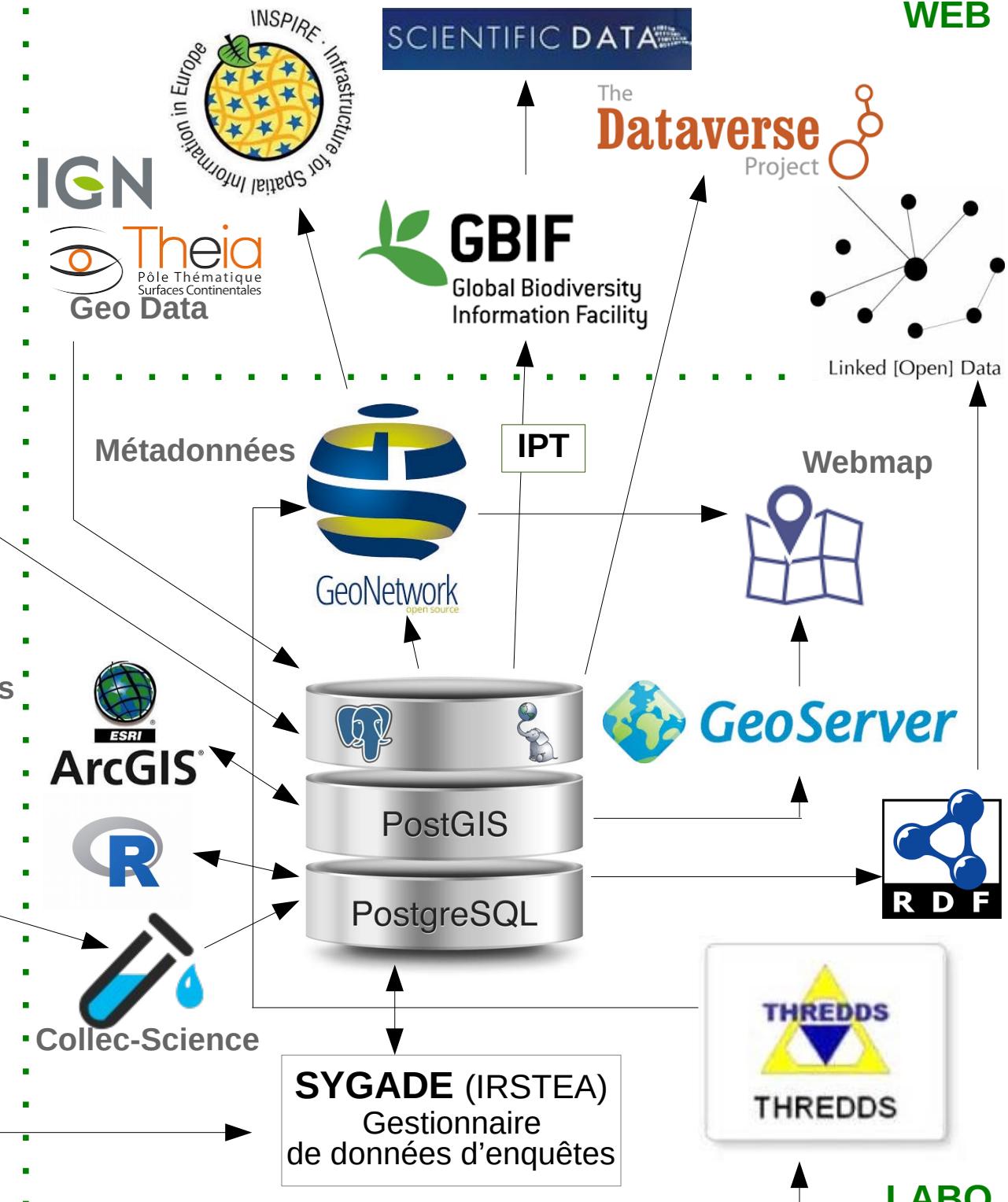
Catalogage
(métadonnées)

OPEN SCIENCE !
→ Mettre en oeuvre la réutilisation
 ⇒ IPT, DOI & Datapaper
PARTAGER

TERRAIN



SCIENTIFIC DATA

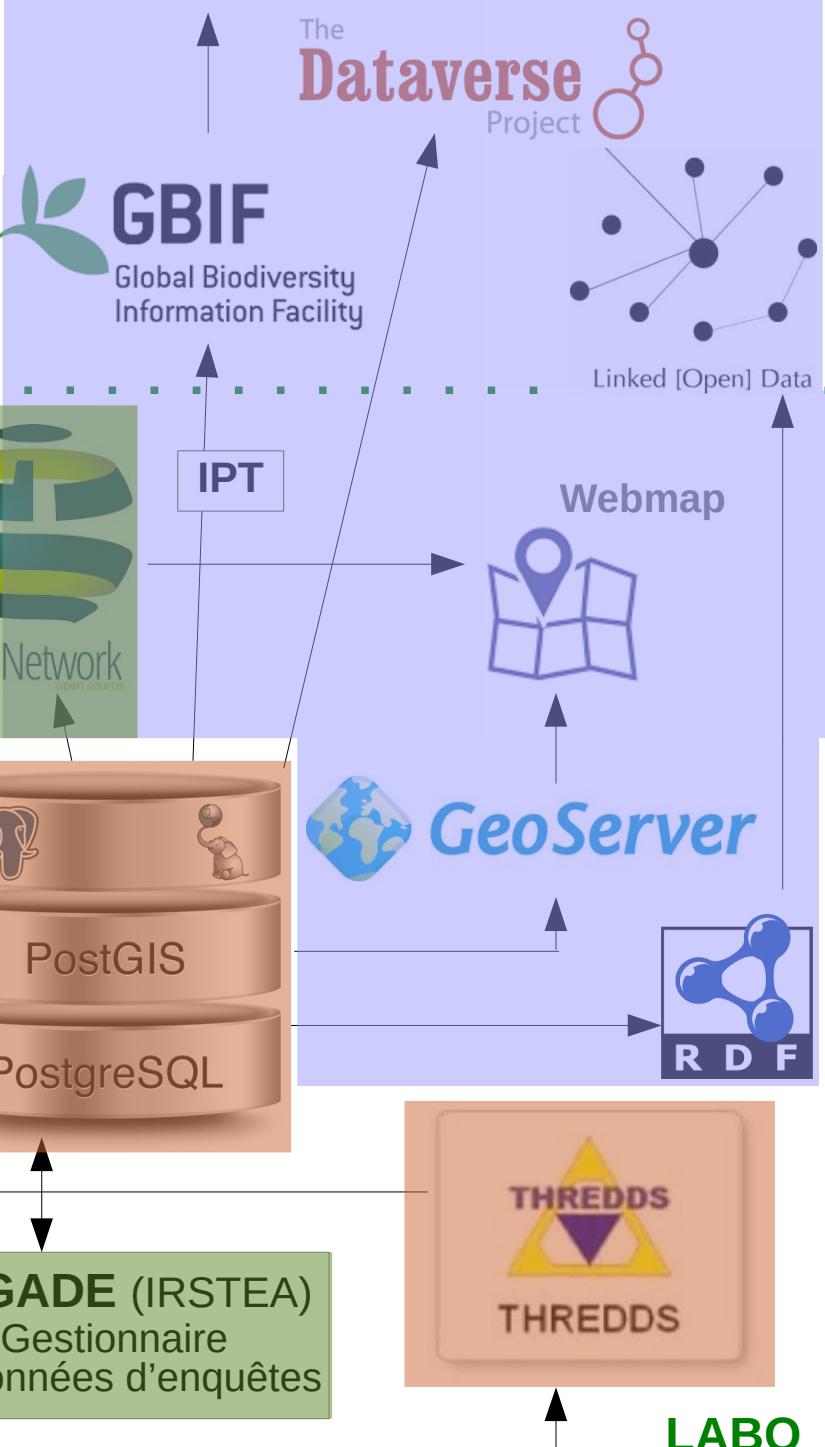


WEB

TERRAIN



SCIENTIFIC DATA



WEB