

Gestion pérenne des données scientifiques



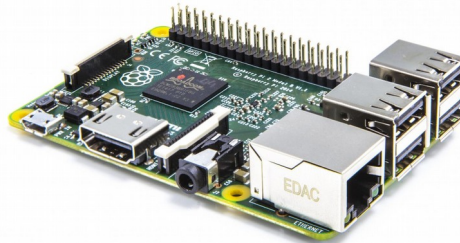
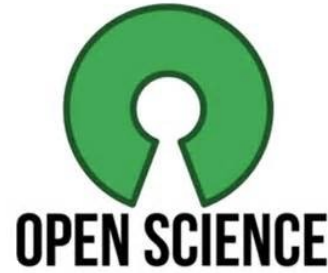
DYNAFOR
UMR 1201 | INRA - INP TOULOUSE



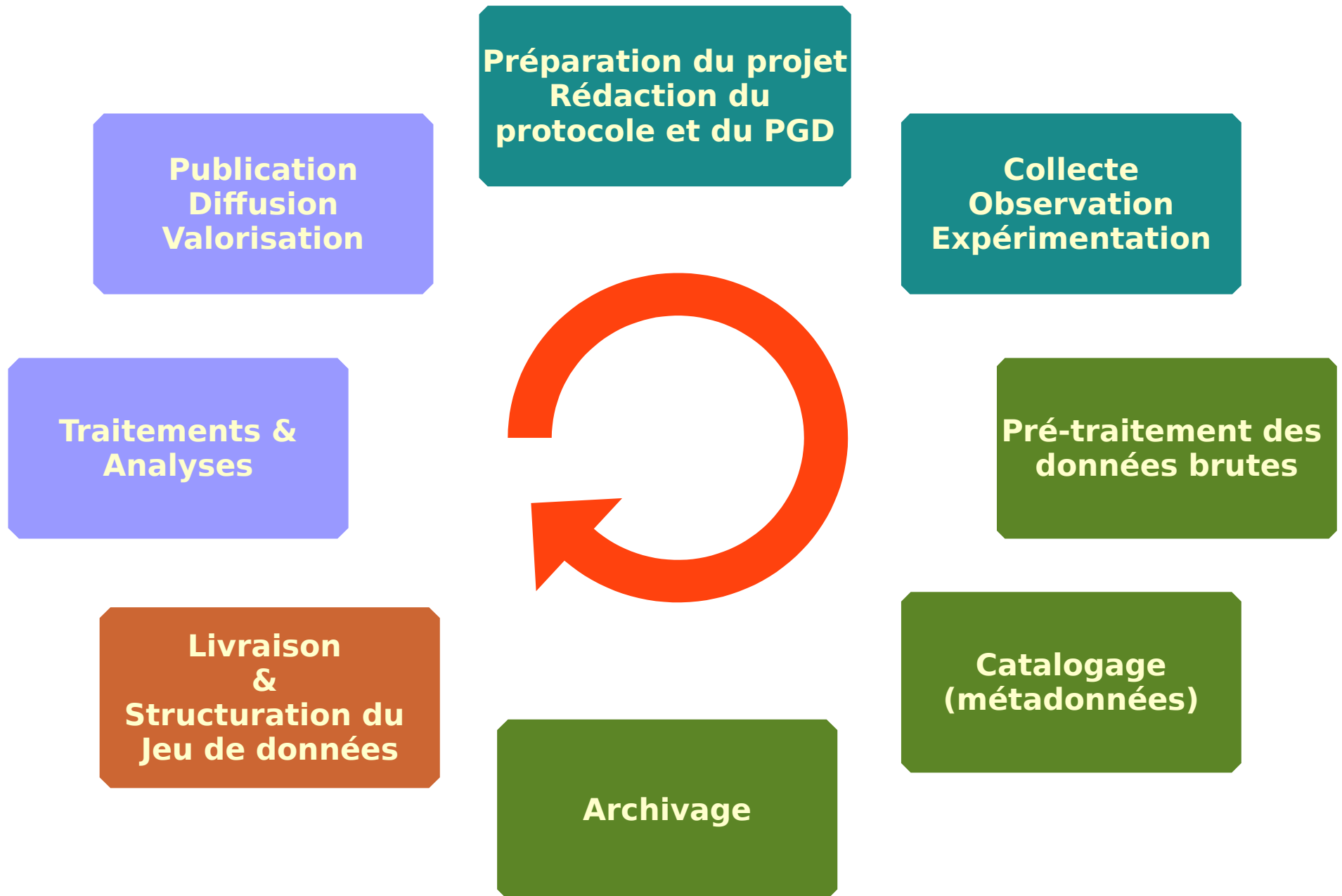
Du Plan de Gestion des données au Datapaper



Contexte général



Le cycle de vie des données



Une gestion à chaque étape

Les objectifs poursuivis (≠ atteints !)

- Traçabilité des projets et de leurs données
- Accès aux données facilité pour tous les agents de l'UMR
- « Chaîne de la donnée » (= sans rupture)
- Standardisation → Ouverture → Valorisation

Préparation du projet
Rédaction du protocole

Collecte
Observation
Expérimentation

Pré-traitement des
données brutes

Catalogage
(métadonnées)

Archivage

Livraison
&
Structuration du
Jeu de données

Traitements &
Analyses

Publication
Diffusion
Valorisation

Préparation du projet
Rédaction du
Protocole **et** du PGD

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 **et** du PGD

Dialogue entre tous les acteurs + documents de références

- Meilleure anticipation
- Premières **métadonnées**



PLANIFIER

Préparation du projet
Rédaction du
Protocole **et** du PGD

Dialogue entre tous les acteurs + documents de références

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



PLANIFIER



Collec-Science

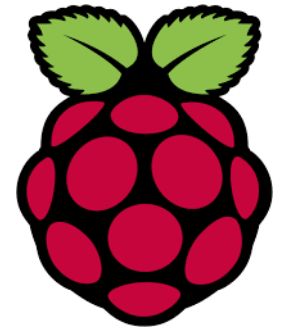


Préparation du projet
Rédaction du
Protocole **et** du PGD

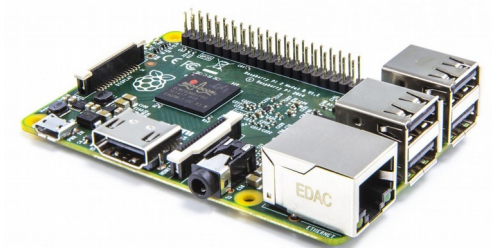
Collecte
Observation
Expérimentation

Etape cruciale mais souvent négligée

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



ACQUÉRIR





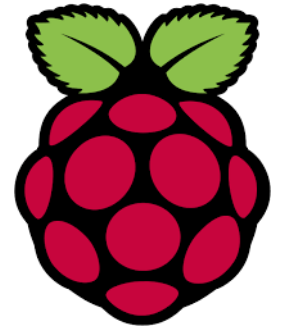
Préparation du projet
Rédaction du
Protocole **et** du PGD

Collecte
Observation
Expérimentation

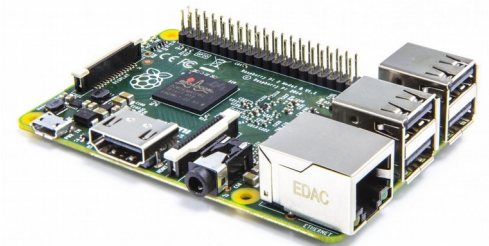
Etape cruciale mais souvent négligée

- Bonne collecte ssi 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 **et** du PGD

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 **et** du PGD

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 **et** du PGD

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 **et** du PGD

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, GeoNAPI R

Pré-traitement des
données brutes

Catalogage
(métadonnées)

DÉCRIRE



Préparation du projet
Rédaction du
Protocole **et** du PGD



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

Catalogage
(métadonnées)

Archivage





Préparation du projet
Rédaction du
Protocole **et** du PGD



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**

Pré-traitement des
données brutes

STOCKER

Catalogage
(métadonnées)

Archivage





PostgreSQL



Préparation du projet
Rédaction du
Protocole **et** du PGD

Collecte
Observation
Expérimentation

Intégrer les données dans le Sys. d'information

→ Bases de données partagées

Pré-traitement des
données brutes

Livraison
&
Structuration du
Jeu de données

ORGANISER

Catalogage
(métadonnées)

Archivage



PostgreSQL



Préparation du projet
Rédaction du
Protocole **et** du PGD

Collecte
Observation
Expérimentation

Intégrer les données dans le Sys. d'information

→ Bases de données partagées

⇒ **Infrastructure de Données Géographiques (IDG)**

Pré-traitement des
données brutes

Livraison
&
Structuration du
Jeu de données

ORGANISER

Catalogage
(métadonnées)

Archivage



Préparation du projet
Rédaction du
Protocole **et** du PGD

Collecte
Observation
Expérimentation

Traitements « scientifiques »

→ Statistiques, SIG, algorithmes, interprétation ...

Traitements &
Analyses

ANALYSER

Pré-traitement des
données brutes

Livraison
&
Structuration du
Jeu de données

Archivage

Catalogage
(métadonnées)



Préparation du projet
Rédaction du
Protocole **et** du PGD

Collecte
Observation
Expérimentation

Traitements « scientifiques »

→ Statistiques, SIG, algorithmes, interprétation ...

Traitements &
Analyses

⇒ WPS, publication de scripts

ANALYSER

Pré-traitement des
données brutes

Livraison
&
Structuration du
Jeu de données

Archivage

Catalogage
(métadonnées)



Publication
Diffusion
Valorisation

Préparation du projet
Rédaction du
Protocole **et** du PGD

Collecte
Observation
Expérimentation

OPEN SCIENCE !

→ Mettre en oeuvre la réutilisation

Traitements &
Analyses

Pré-traitement des
données brutes

PARTAGER

Livraison
&
Structuration du
Jeu de données

Archivage

Catalogage
(métadonnées)



Publication
Diffusion
Valorisation

Préparation du projet
Rédaction du
Protocole **et** du PGD

Collecte
Observation
Expérimentation

OPEN SCIENCE !

Traitements &
Analyses

→ Mettre en oeuvre la réutilisation
⇒ **WFS, R, IPT, DOI & Datapaper**

Pré-traitement des
données brutes

PARTAGER

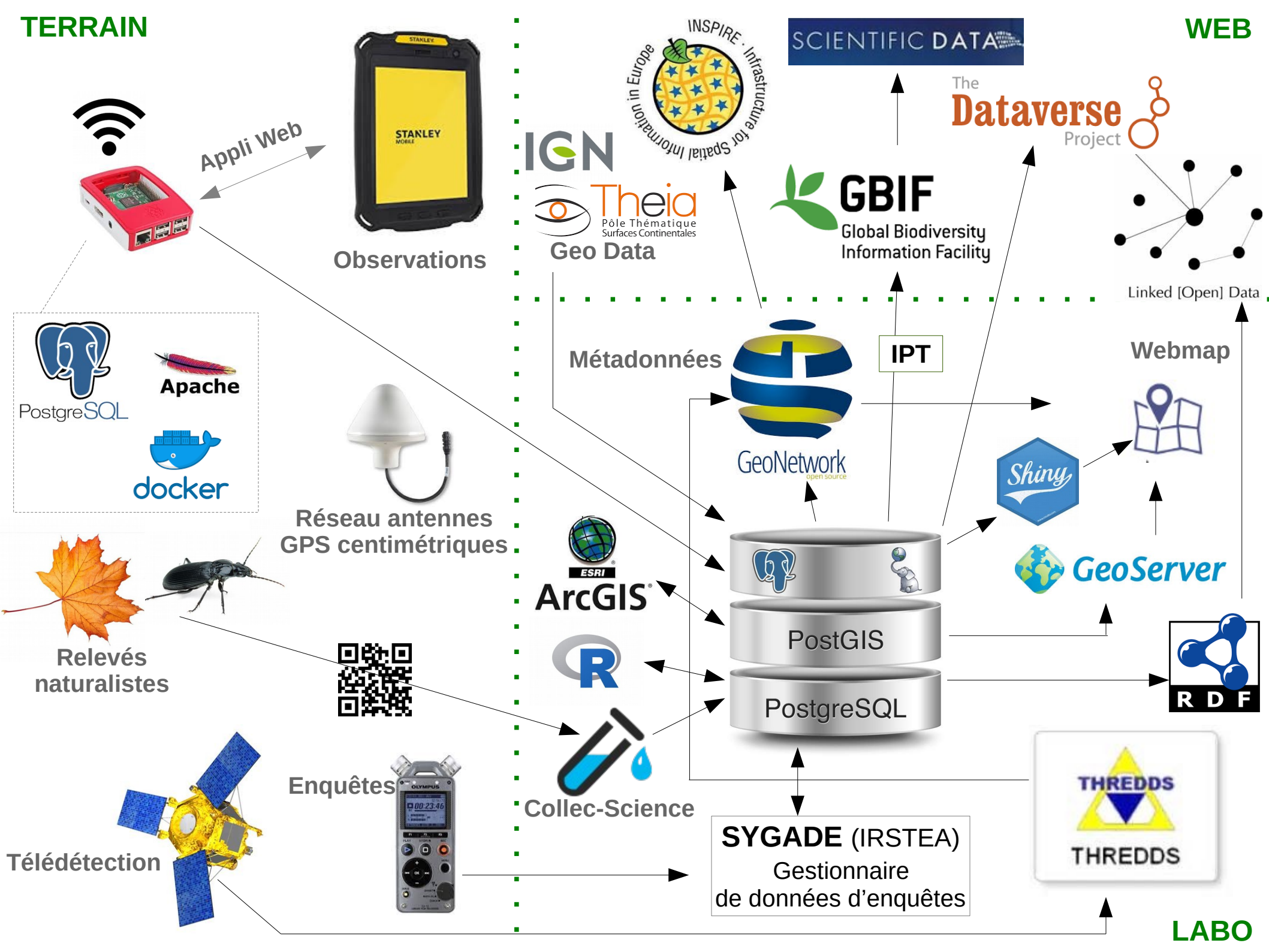
Livraison
&
Structuration du
Jeu de données

Archivage

Catalogage
(métadonnées)

TERRAIN

WEB



Appli Web



Observations

PostgreSQL

Apache

docker



Réseau antennes GPS centimétriques



Relevés naturalistes



Enquêtes



Télédtection



IGN

Theia

Pôle Thématique Surfaces Continentales

Geo Data

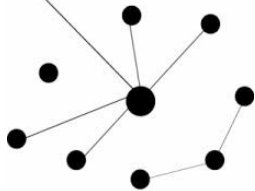


SCIENTIFIC DATA

GBIF

Global Biodiversity Information Facility

The Dataverse Project



Linked [Open] Data

Métadonnées



IPT

Webmap



GeoServer

ESRI

ArcGIS



Collec-Science

PostGIS

PostgreSQL

RDF

SYGADE (IRSTEA)

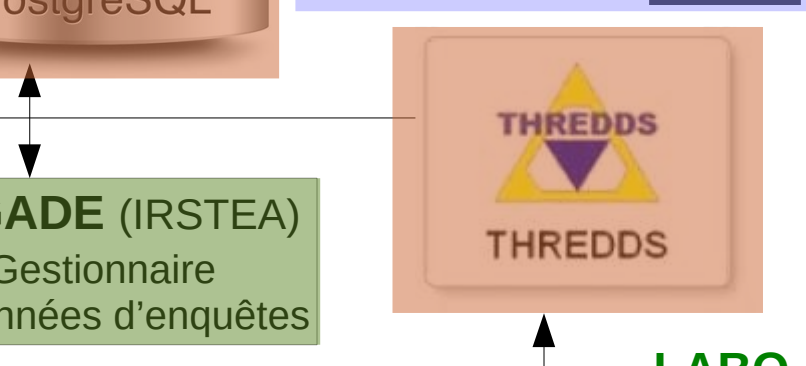
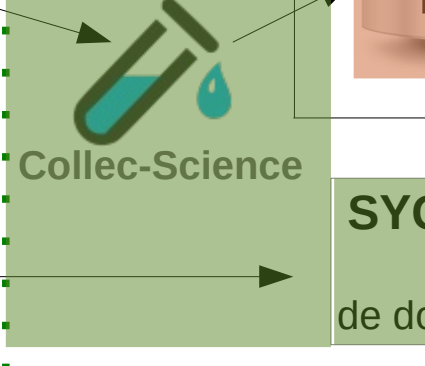
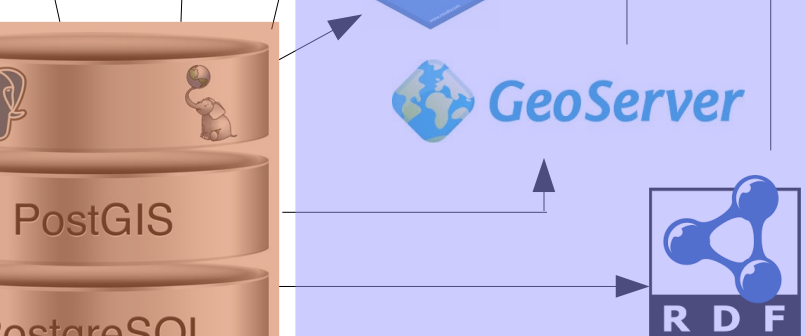
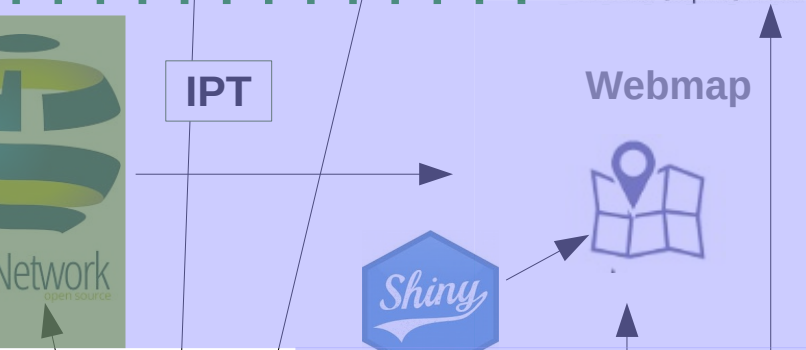
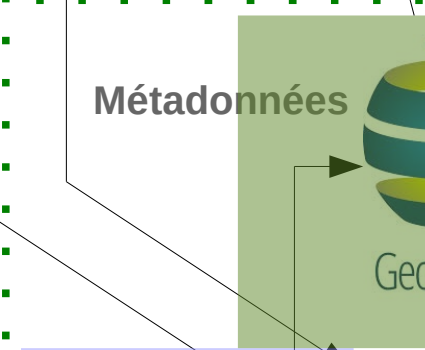
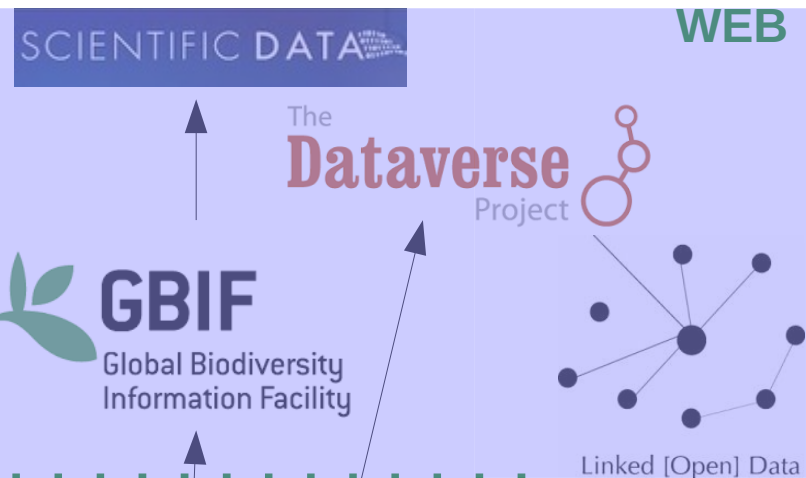
Gestionnaire de données d'enquêtes

THREDDS

THREDDS

LABO

TERRAIN





wilfried.heintz@inra.fr