



Innovation - Écologie - Territoires

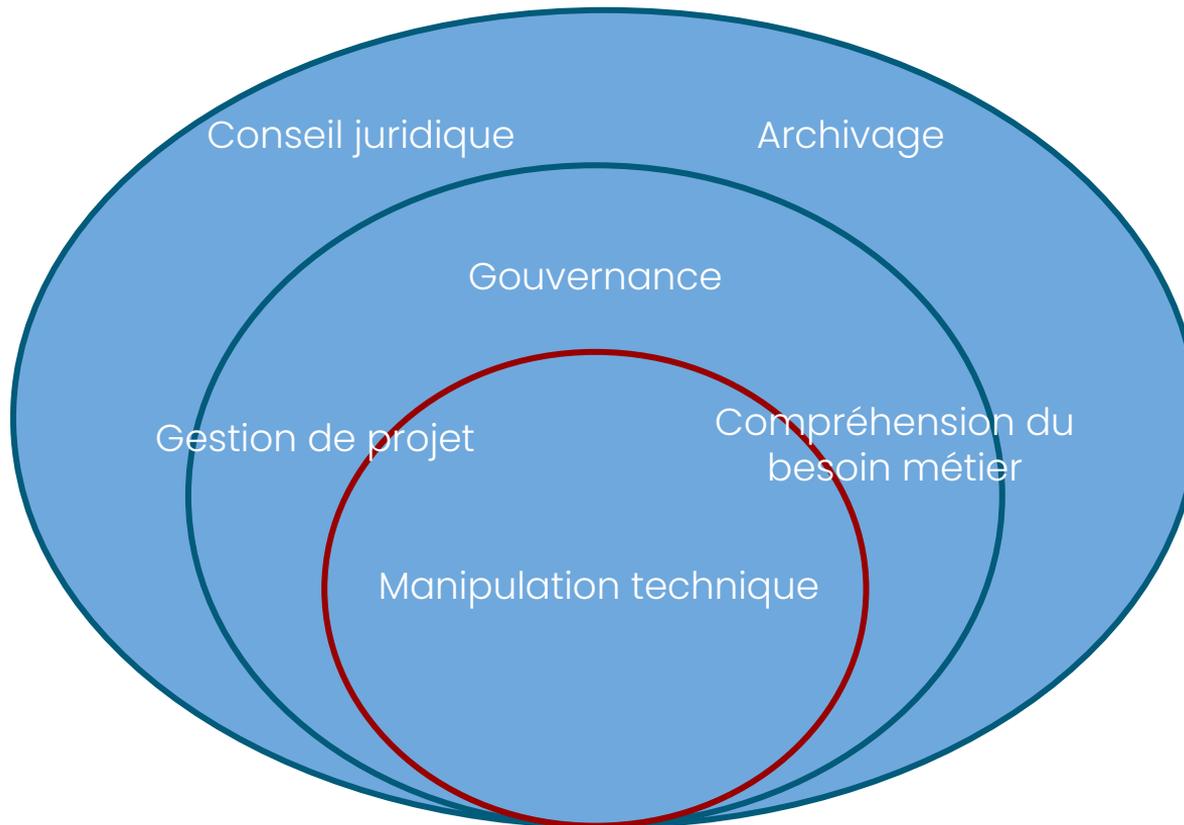


SERVICES DONNÉES

L'offre d'accompagnement de  
l'Ecolab

Webinaire Ecolab - CVRH sur la Culture de la donnée

## Les différents cercles de la donnée



## Le cycle de la donnée et les métiers & compétences associées

1. Le cycle technique de la donnée
  - a. Collecte
  - b. Stockage et transformation
  - c. Visualisation et activation
  - d. Orchestration et hébergement
2. Les métiers & compétences techniques de la donnée  
Les compétences et rôles d'un projet data
3. Quelques exemples concrets  
Illustration du cycle et des métiers de la données sur quelques cas concrets

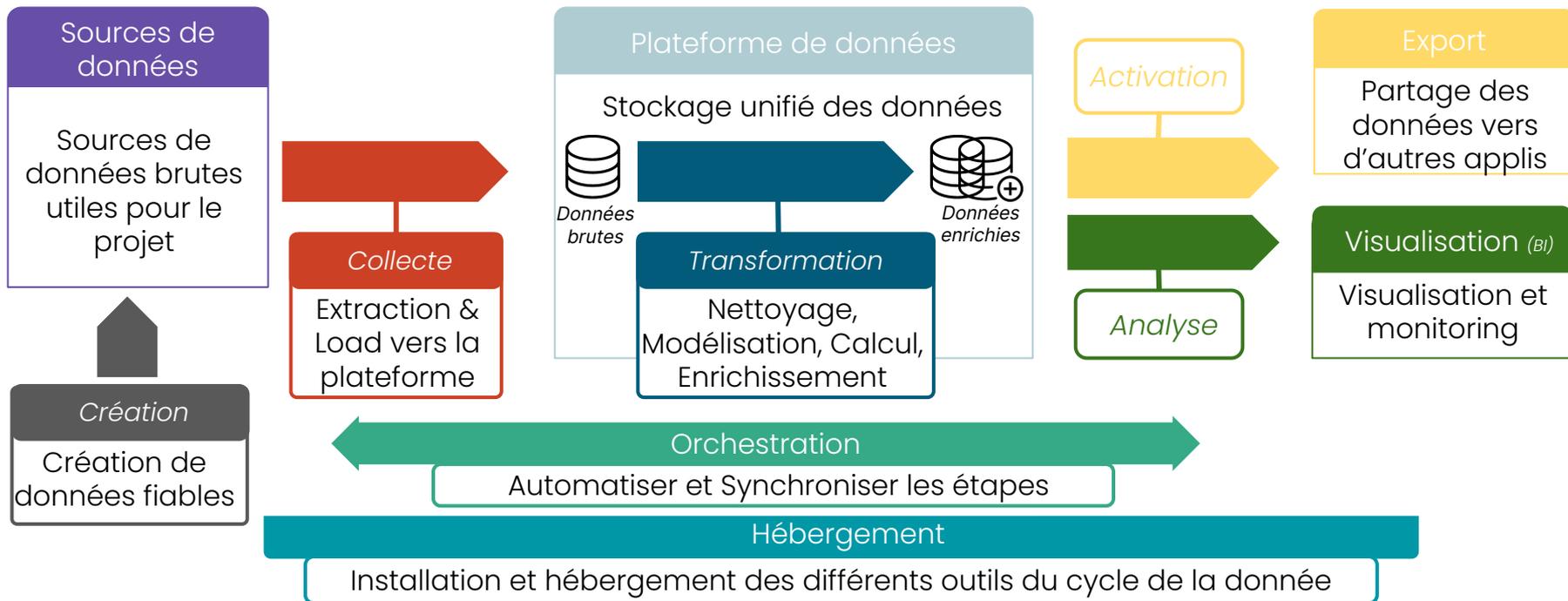
# Le cycle technique de la donnée



# Le cycle de la donnée

## Catalogue

Listing et documentation des outils et des données disponibles pour les analyses



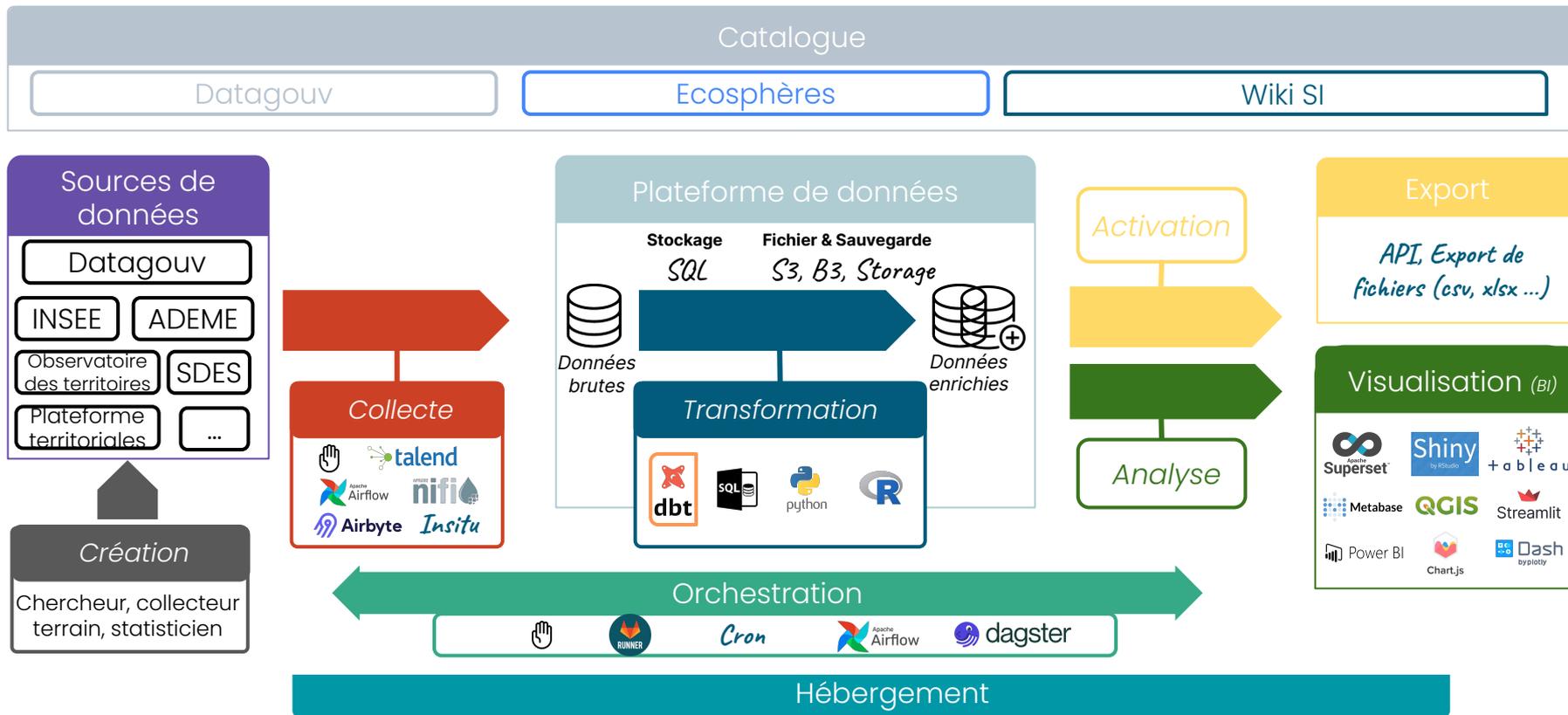


SERVICES DONNÉES

# Le cycle de la donnée

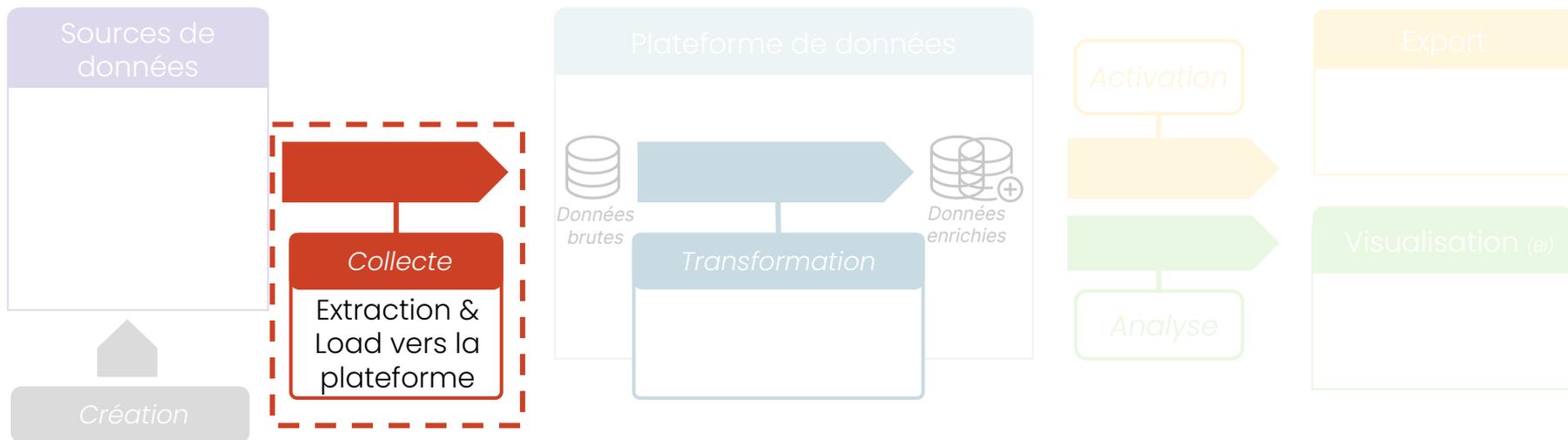


Innovation - Écologie - Territoires



## La collecte

## La collecte (EL)

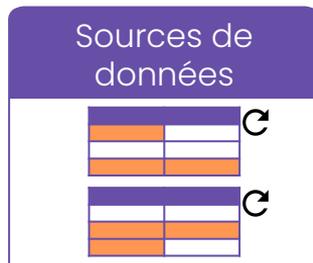


## La collecte (EL)

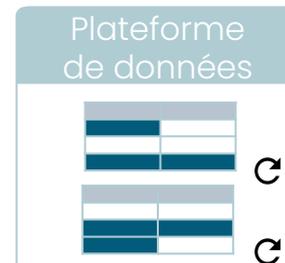
### Enjeu **Extraire les données requises**

- Se connecter à la source de données
- Exporter les données sur la plateforme
- Automatiser la mise à jour régulière

### Outils

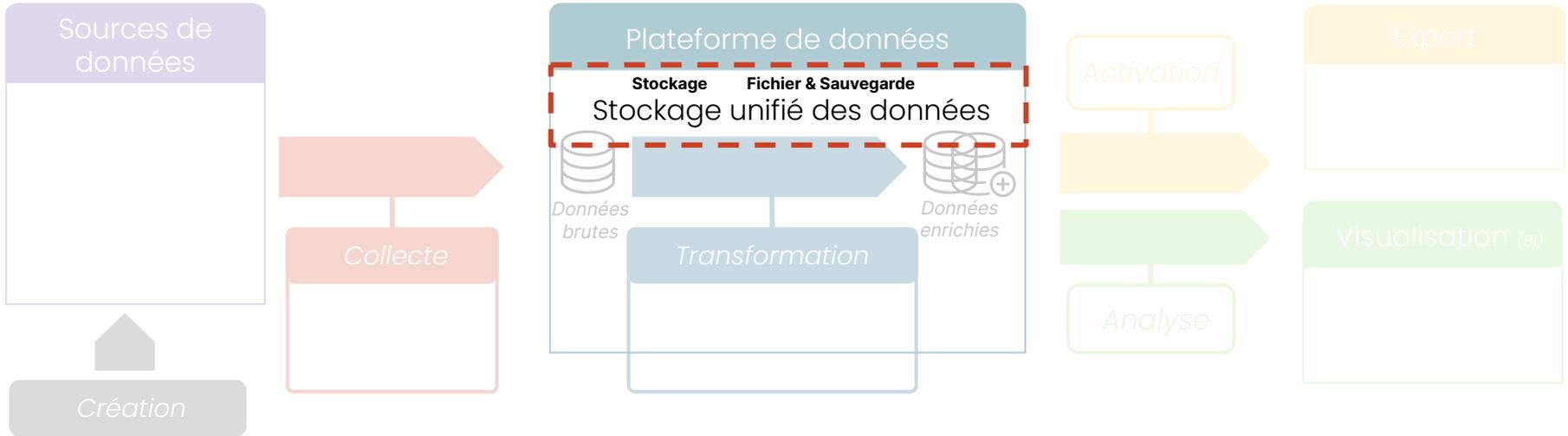


Collecte



# Le stockage

# Le stockage



## Enjeu **Stockage et processing des données**

- Le stockage de données
  - fichiers
  - bases postgresQL et postGIS
- Le processing : transformation + requêtage BI
- La sauvegarde des données

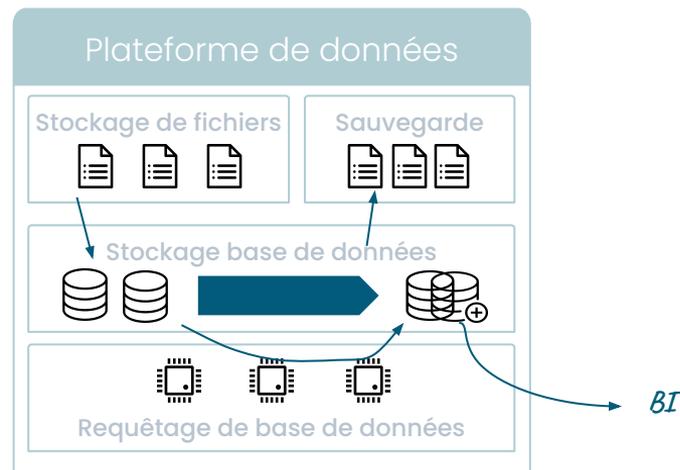
## Outils

### Stockage & requêtage



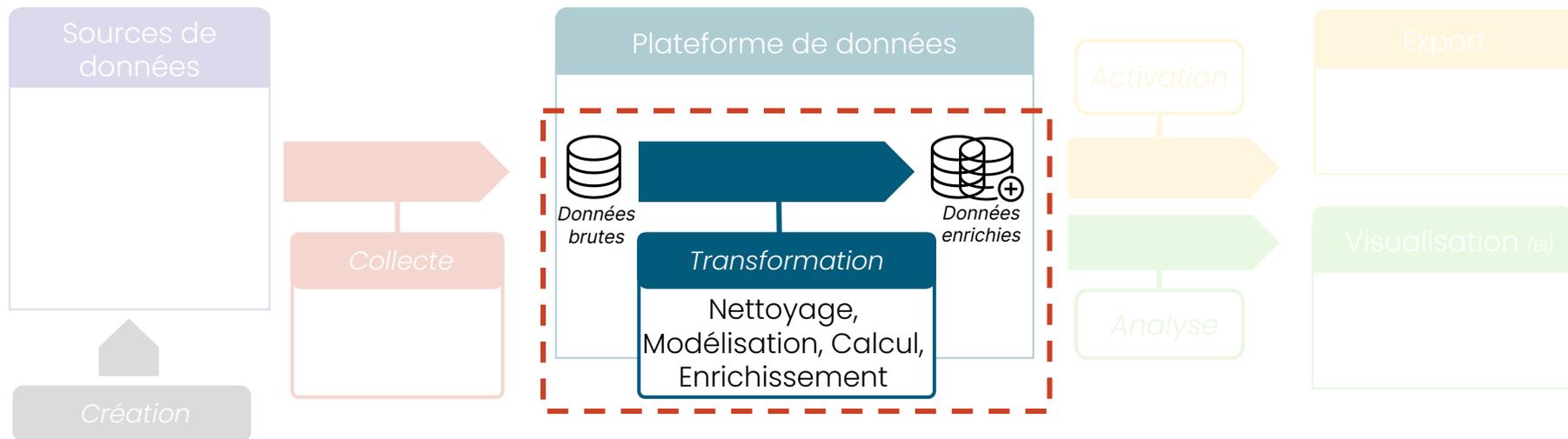
### Sauvegarde

S3 / B3



# La transformation

# La transformation



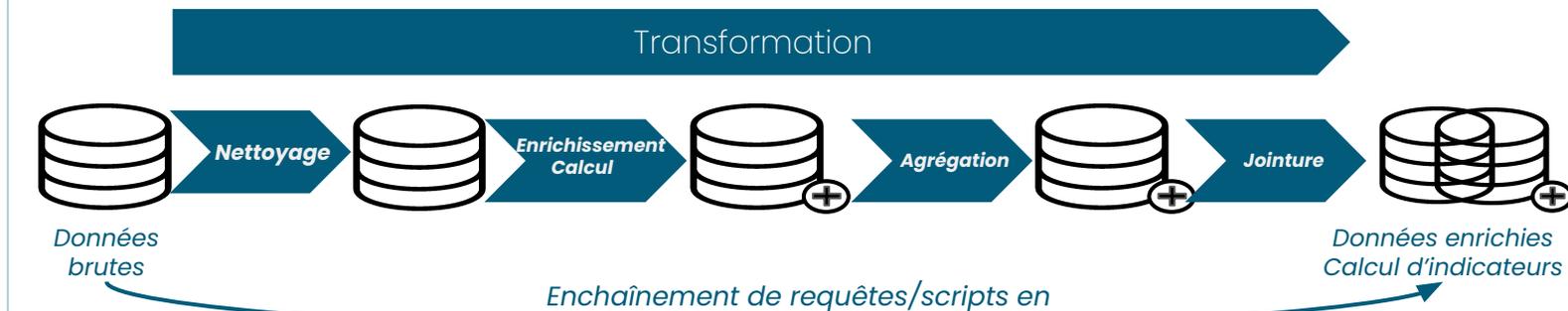
## Enjeu Transformation

- Enrichissement et calcul d'indicateurs
- Croisement des données
- Qualité, Clarté, fiabilité

## Langages



## Plateforme de données



 <i>Outils</i>	Simplicité d'hébergement	Simplicité d'utilisation	Perfs	Transformations avancées	Commentaires
	●●●	●●○	●●●	●○○	Simple, standard et efficace (souvent complémentaire aux 2 autres)
	●●●	●○○	●●○	●●●	Plus avancé et bien adapté pour le ML / data science
	●●●	●○○	●●○	●●●	Plus avancé et bien adapté pour les statistiques

Choisir un langage en fonction :

- des compétences en interne
- du type de transformations à effectuer
- en privilégiant la simplicité

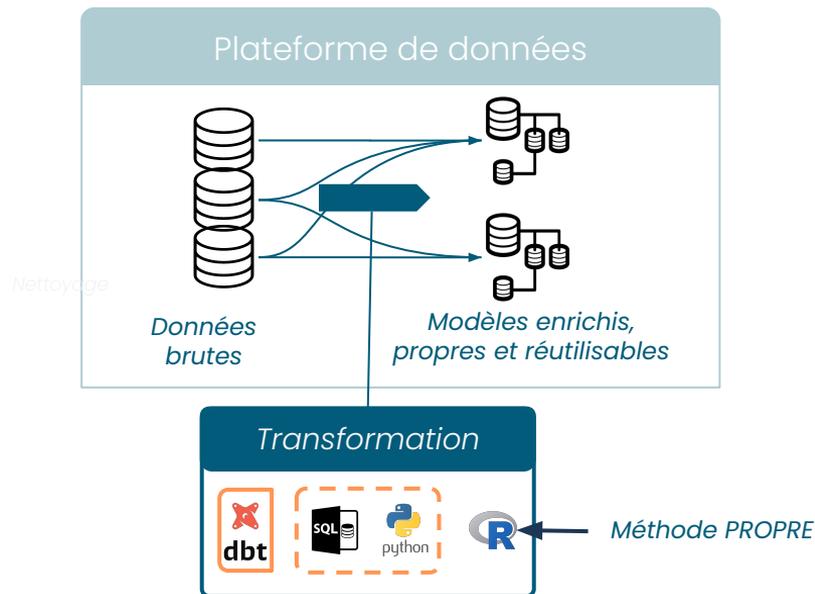
## Enjeu **Modélisation & Gouvernance**

- Fiabilité des données - tests, déduplication
- Maintenance - capitalisation, déduplication
- Clarté - documentation, déduplication

## Outils

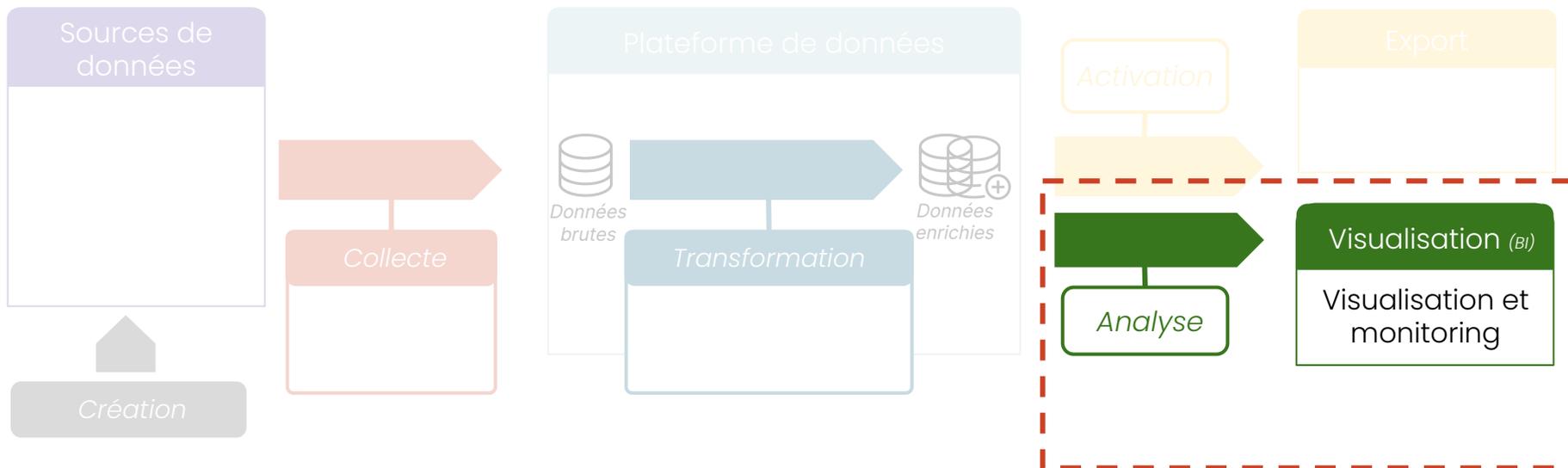


*Process internes*

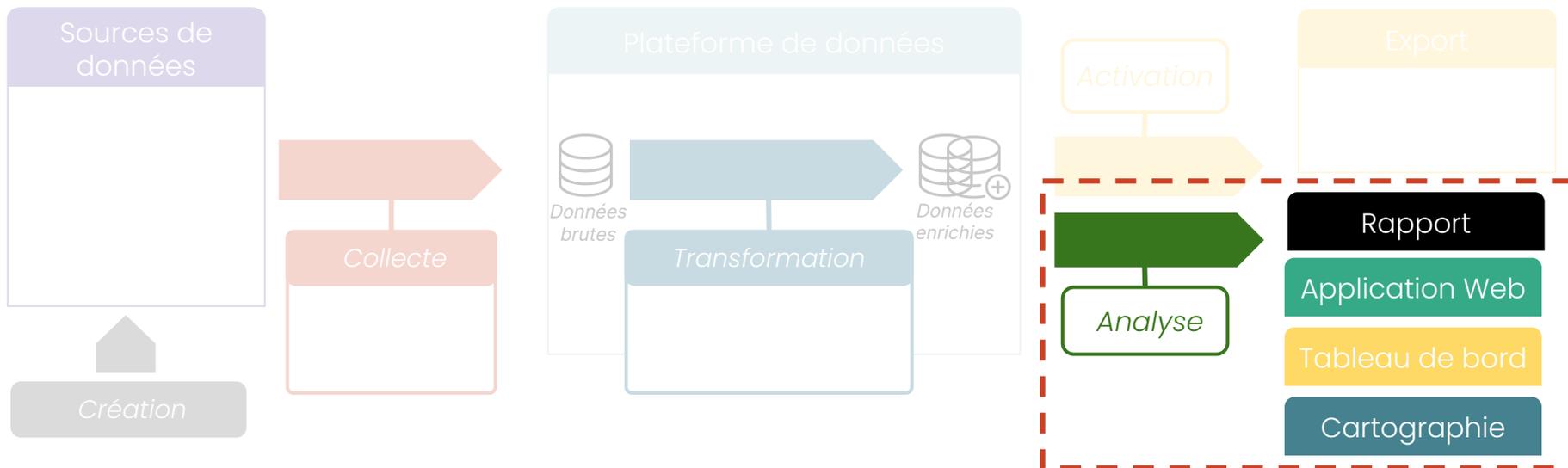


# La visualisation

# La visualisation



# La visualisation





## La visualisation

### Rapport / Notebook

#### Enchaînement de code informatique et de visualisations graphiques

##### Avantages

Utile et pratique pour les tests, les explorations et analyses ponctuelles côté technique

##### Inconvénients

Pas adapté pour les vis' en production  
Pas de MAJ auto, pas interactif, ...

### Application Web

#### Création d'applications web interactives avancées

##### Avantages

Personnalisation forte  
Fonctionnalités avancées  
Peut être plus robuste

##### Inconvénients

Plus long et complexe  
Faible autonomie du métier  
Evolution et maintenance difficile

### Tableau de bord

#### Monitoring et suivi d'indicateurs. Analyses poussées

##### Avantages

Variété de graphiques  
Simple et très flexible  
Autonomie du métier

##### Inconvénients

Moins de personnalisation  
Moins poussé sur la cartographie

### Cartographie

#### Cartes & données géographiques

##### Avantages

Efficace sur les cartes & géo  
Adapté aux données territoriales

##### Inconvénients

Assez spécifique à la cartographie  
Nécessite des compétences en géomatique



SERVICES DONNÉES

# La visualisation

## Rapport / Notebook



...

### Enchaînement de code informatique et de visualisations graphiques

## Application Web

Applis Web

### Création d'applications interactives avancées



Frameworks python / R



...

## Tableau de bord

### Monitoring et suivi d'indicateurs. Analyses poussées



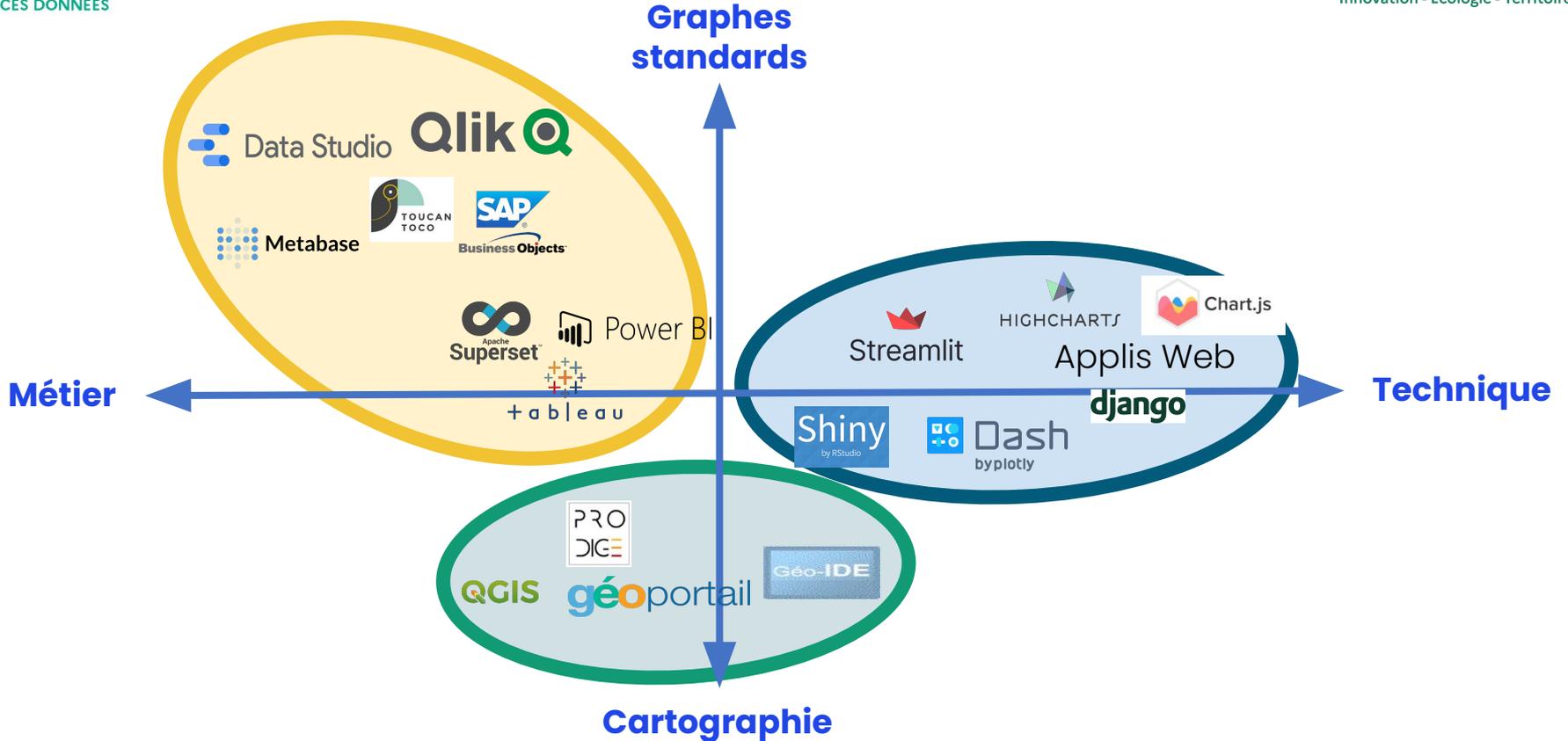
## Cartographie

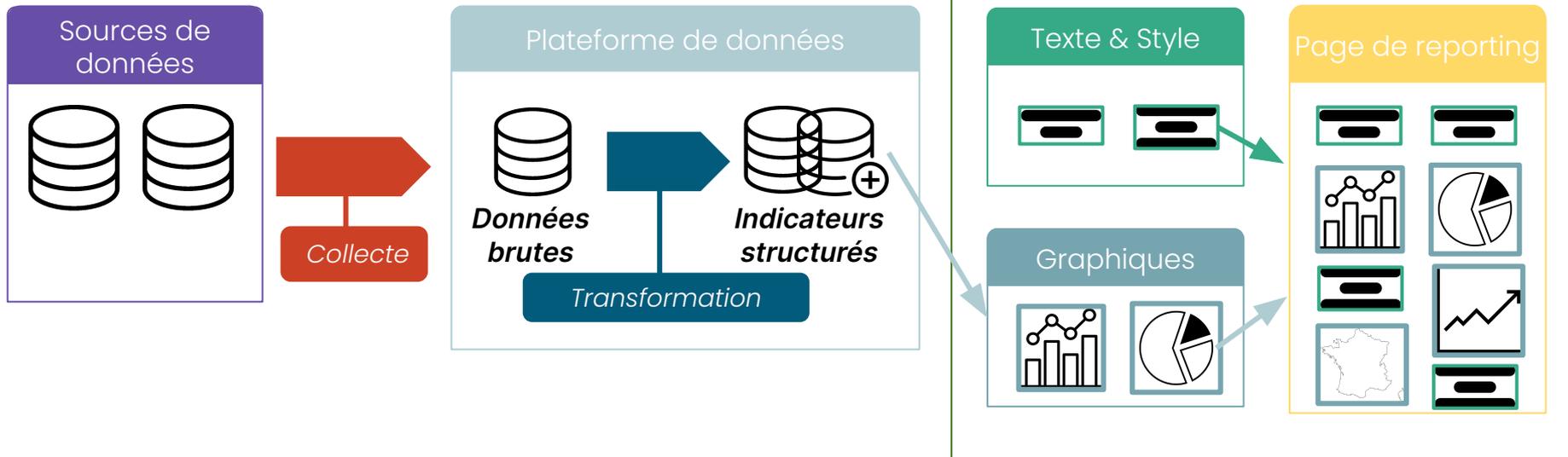
### Cartes & données géographiques



...

# La visualisation





**Les outils des visualisation partagent une structure similaire**

## Le partage & l'activation



SERVICES DONNÉES

# Le partage & l'activation



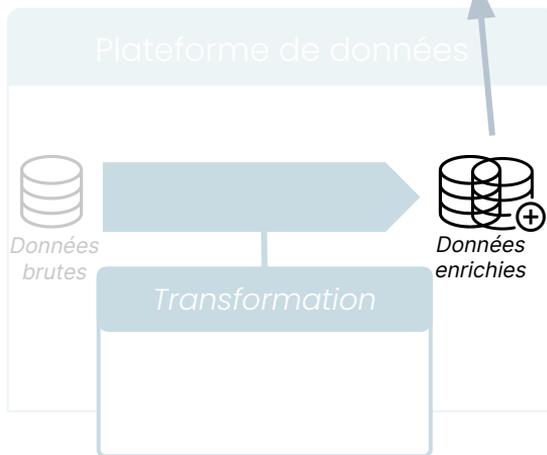
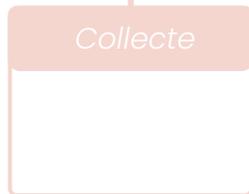
Innovation - Écologie - Territoires

Catalogue

Listing et documentation des outils et des données disponibles pour les analyses



Collecte



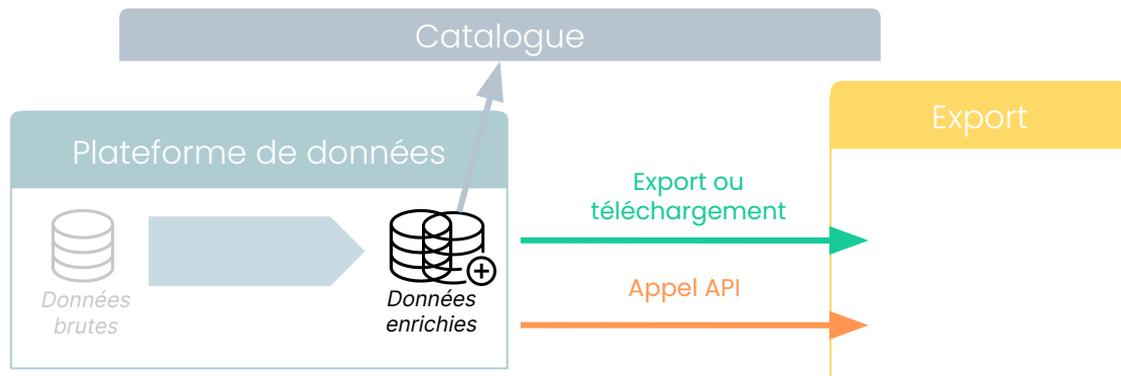
Création

## Enjeu **Partager, activer et valoriser**

- Valoriser les données enrichies et indicateurs
- Activer dans d'autres applications métiers
- Mettre à disposition pour d'autres acteurs

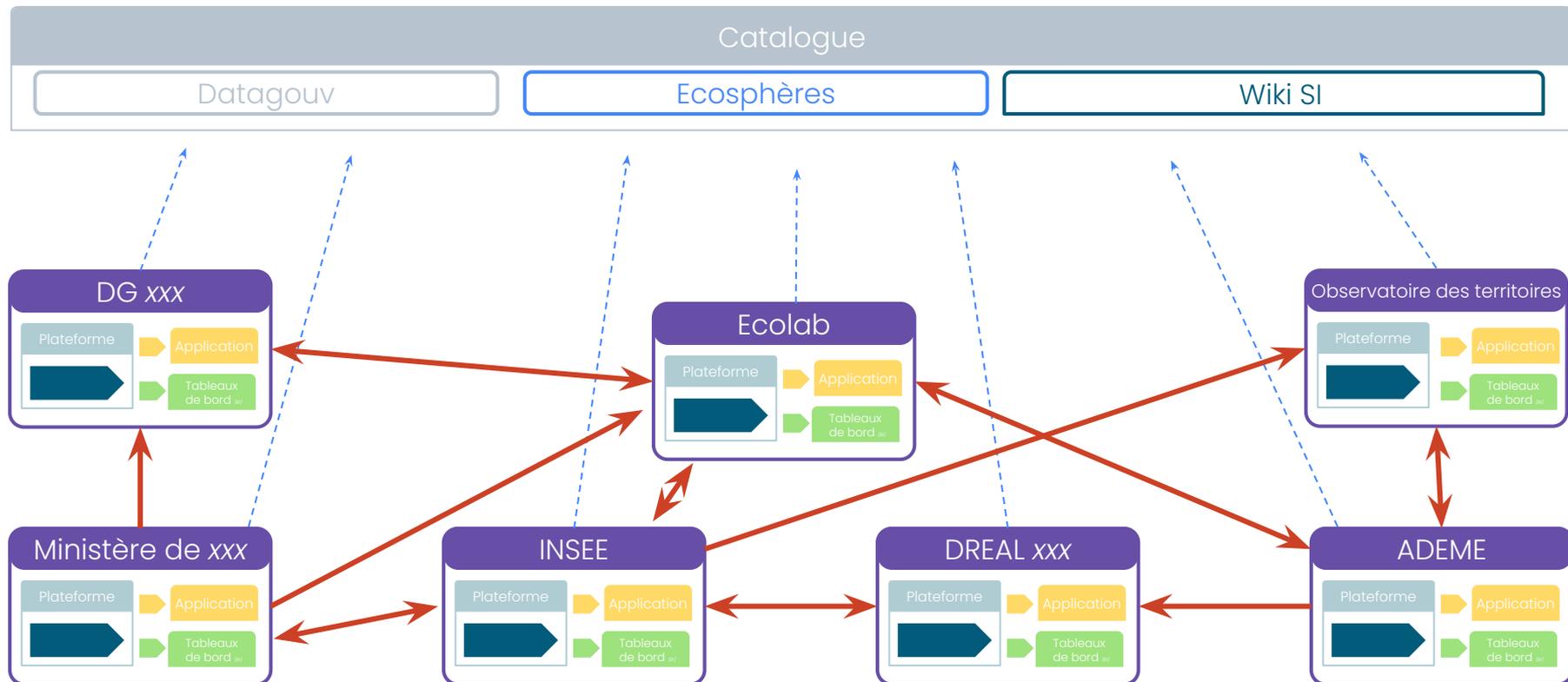
## Outils

*Appel API*      *Export ou téléchargement de fichiers (csv, xlsx, ...)*      *Webhook*



- Mise à disposition des données et indicateurs enrichis sur les bases SQL
- Exposition des données via API
  - Éventuelles options supplémentaires : Webhook / Fichiers plats
  - Partage des informations sur les catalogues de données
  - Respect de certaines conventions (Normes ISO et [DCAT](#), table schema, ...)

# Le partage & l'activation

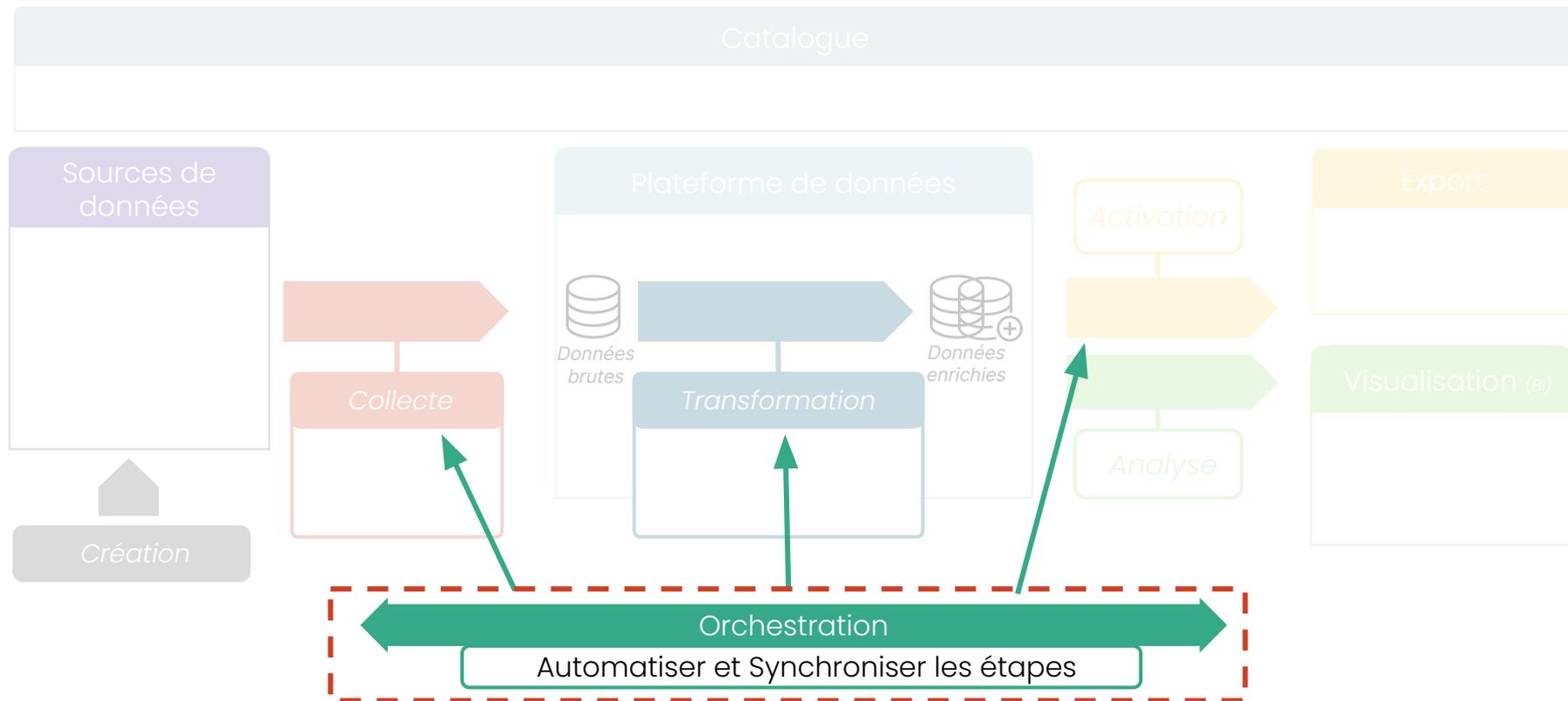


# L'orchestration



SERVICES DONNÉES

# L'orchestration



# L'orchestration

## Enjeu **Synchroniser et automatiser**

- Mise à jour automatisée et planifiée des données
- Lancement successif des tâches du pipeline
- Monitoring et alerting du pipeline de données

## Outils



Cron



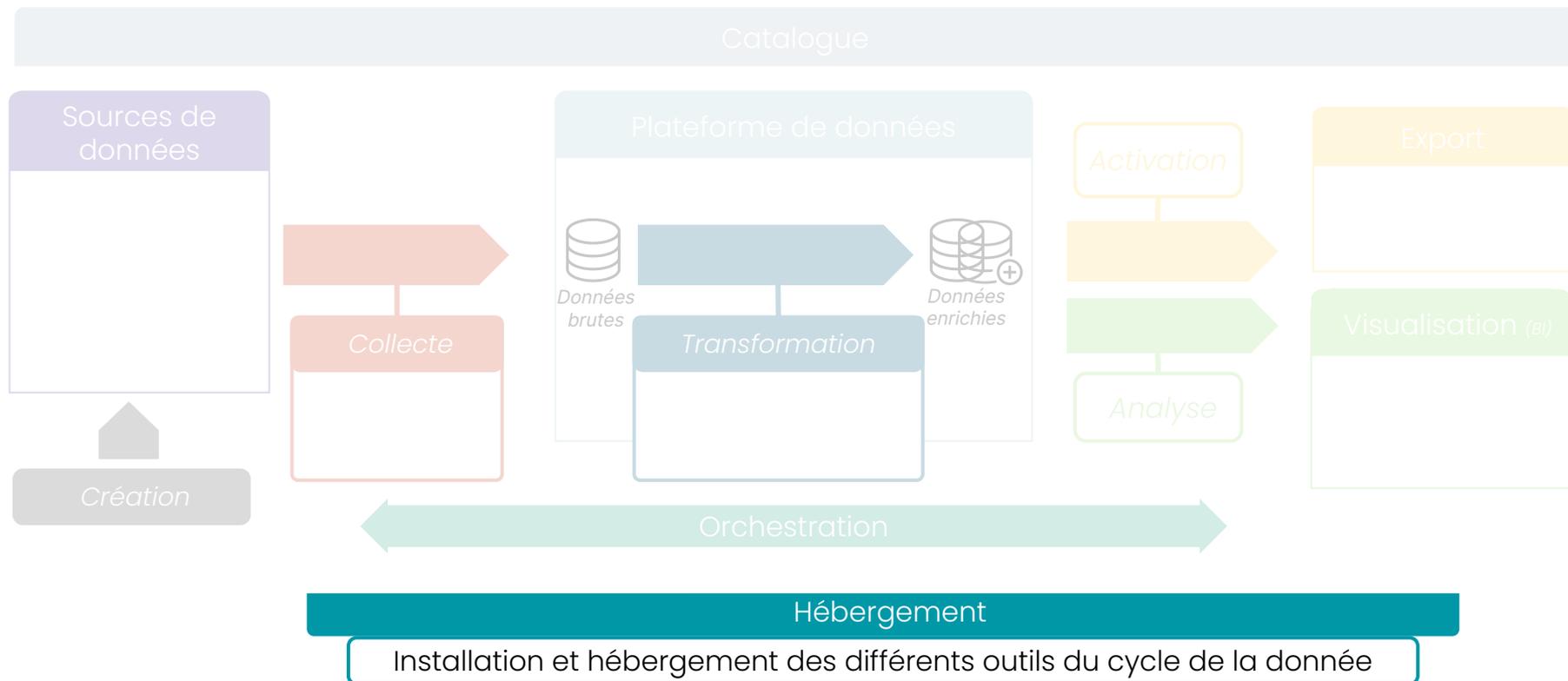
Connexion à l'outil  
de collecte (EL)

Connexion aux scripts  
de transformation

Connexion à l'outil  
d'activation/analyse

Orchestration – Synchro automatisée & planifiée

# Hébergement



# Synthèse briques



SERVICES DONNÉES

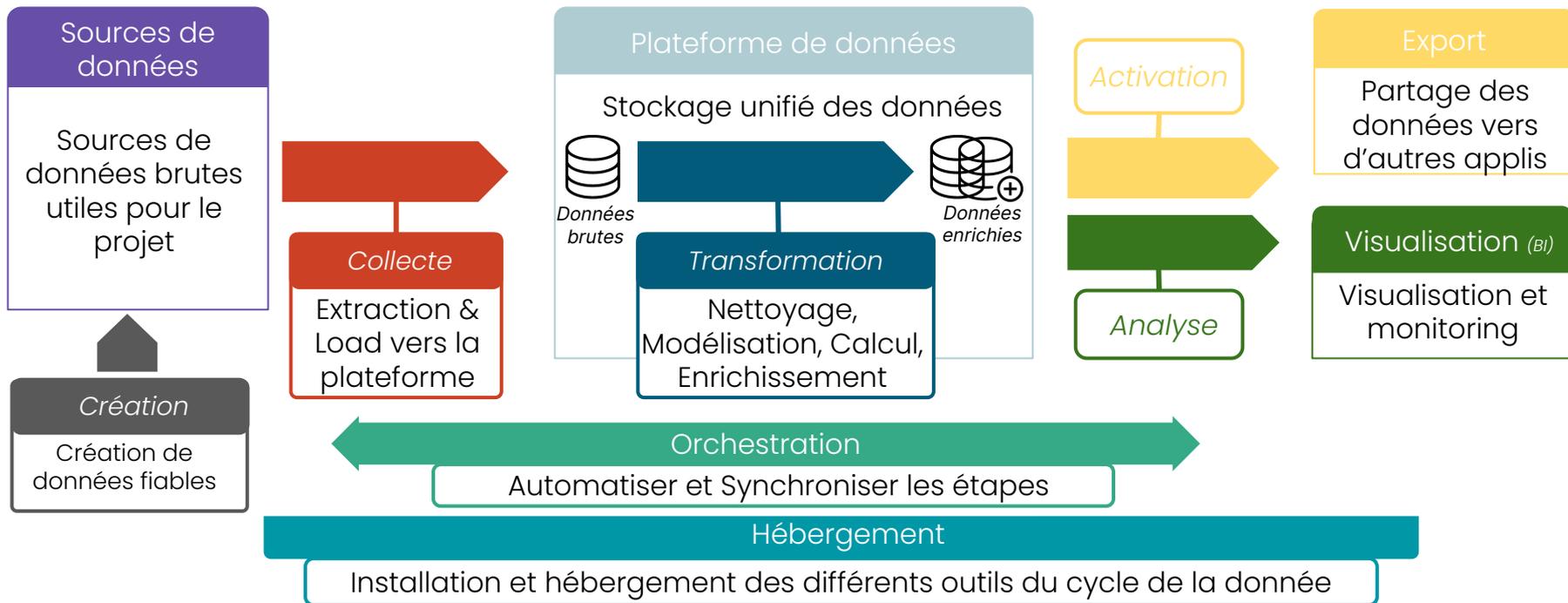
# Le cycle de la donnée



Innovation - Écologie - Territoires

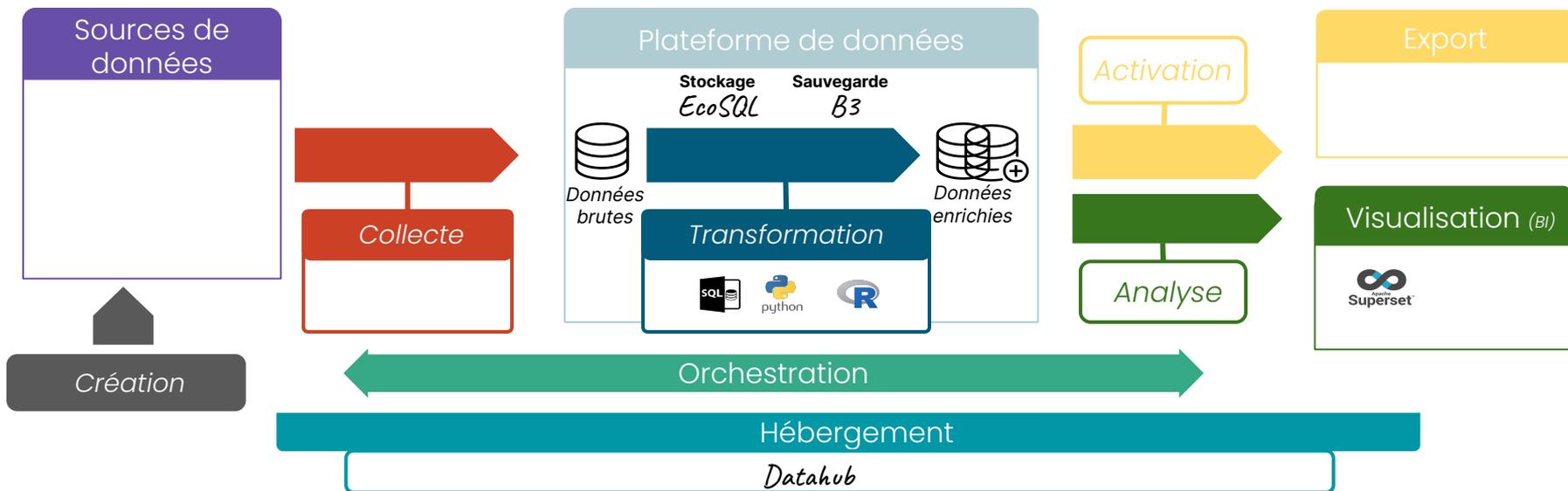
Catalogue

Listing et documentation des outils et des données disponibles pour les analyses



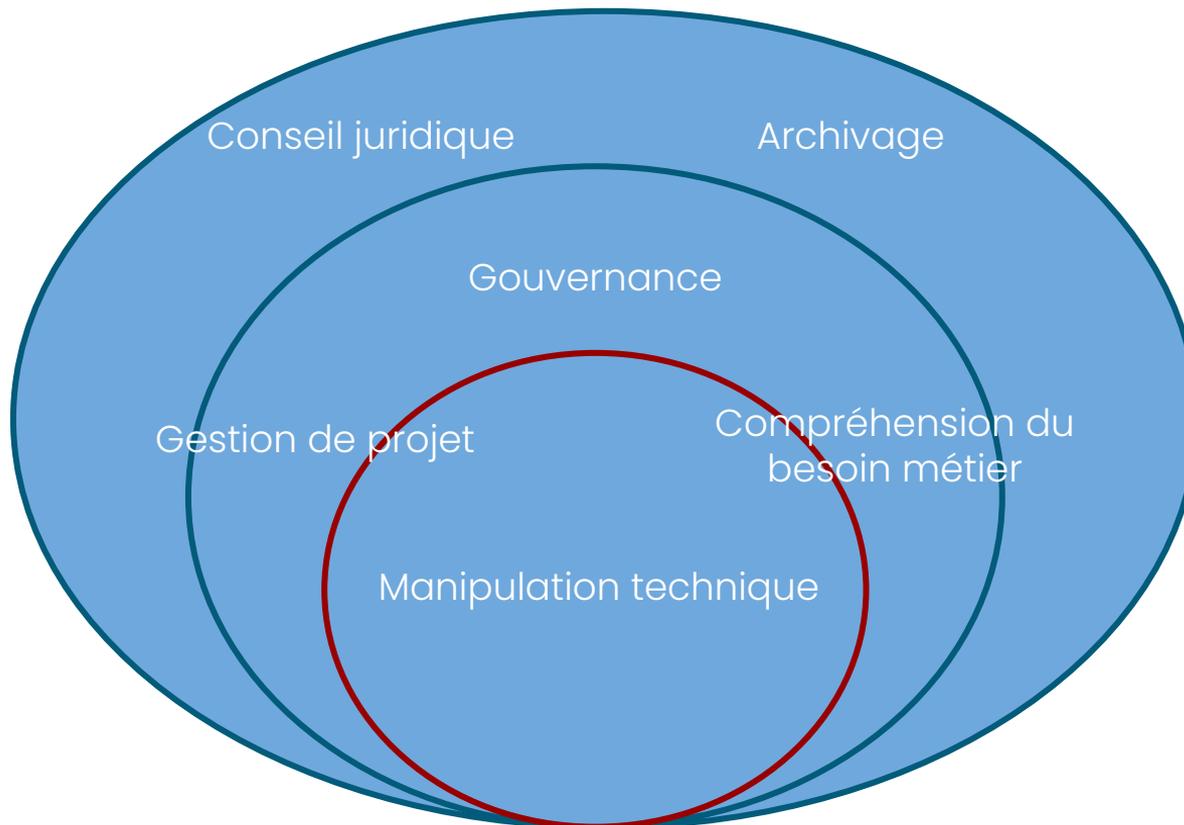
## A retenir

- Cycle : On retrouve ces différentes étapes dans presque tout projet data.
- Stockage : Stocker les données sur des serveurs SQL, notamment l'offre EcoSQL
- Transformation : Séparer collecte & transformation  
Stocker les données sur la base SQL (brutes + enrichies)
- Transformation : Choisir le langage adapté à votre besoin et à vos compétences
- Offre DNUM : développement d'outils par la DNUM (EcoSQL, B3, Superset, Datahub, ...)

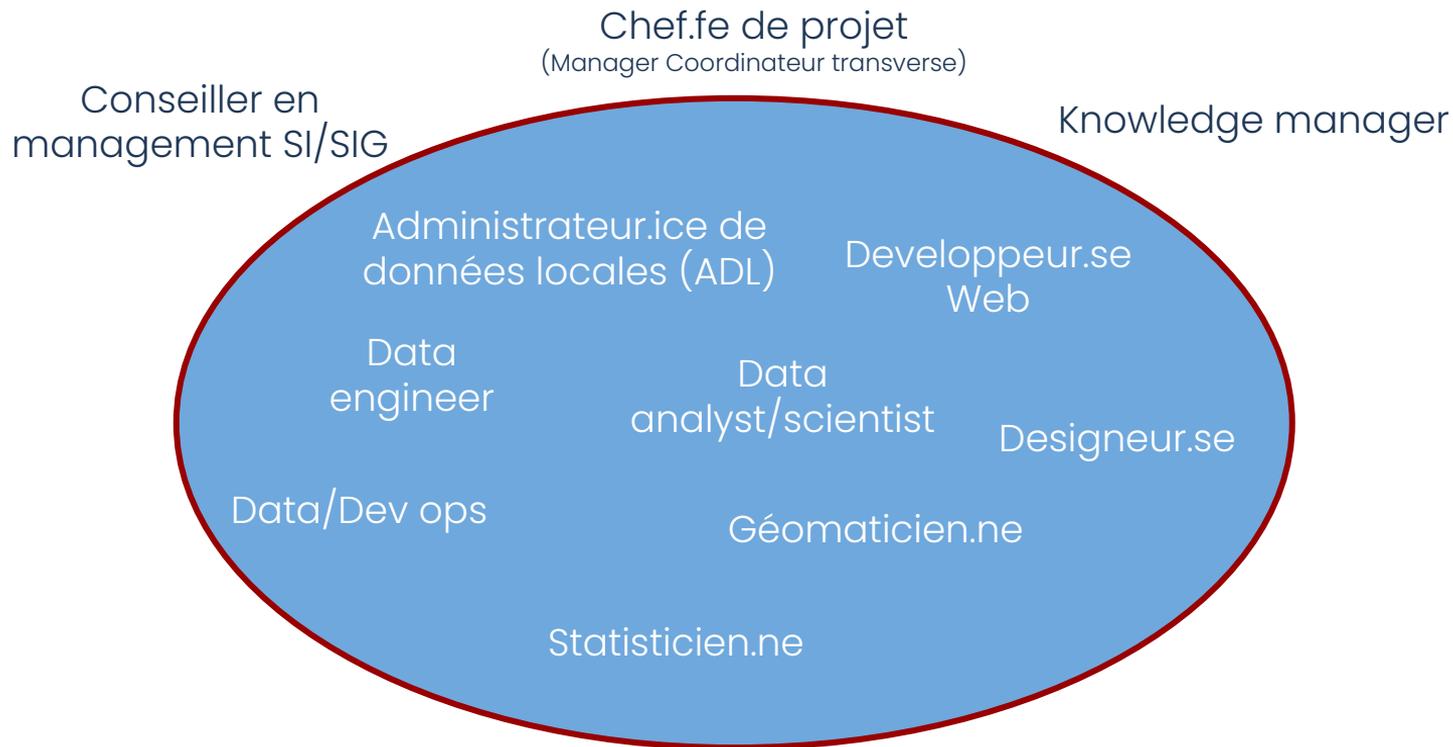


# Les métiers techniques de la donnée

## Les différents cercles de la donnée



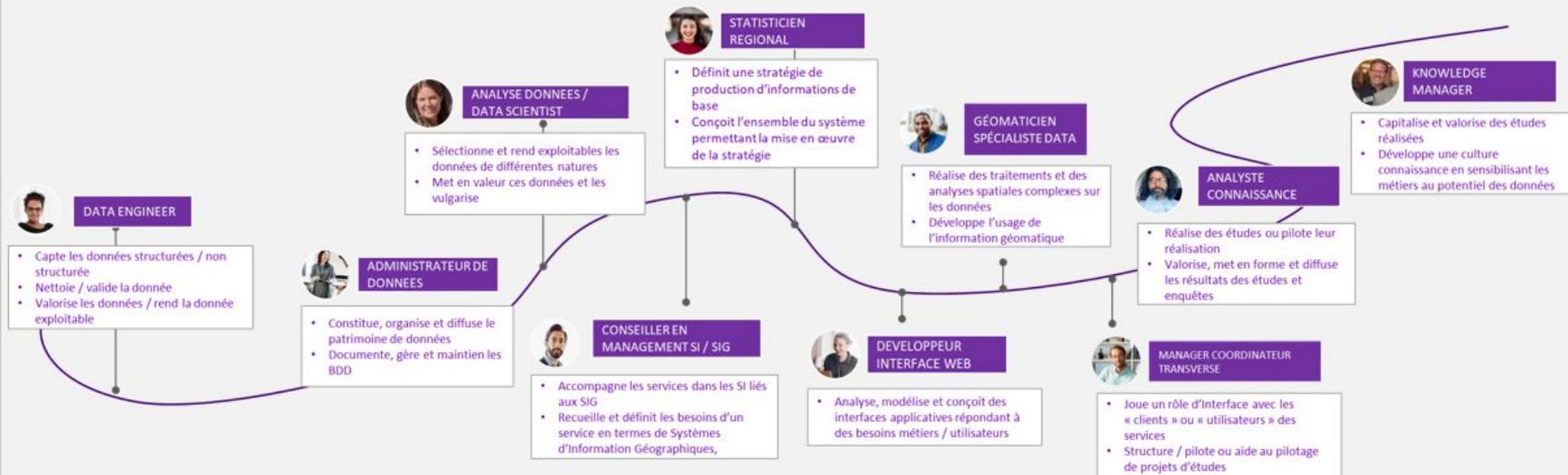
## Les métiers techniques de la donnée



**Manipulation technique**

# Panorama des emplois-type au sein de la chaîne de valeur « connaissance et data »

Collecte des données ➔ Structuration des données ➔ Administration des données ➔ Analyse des données ➔ Représentation des données ➔ Etudes / coordination ➔ Capitalisation



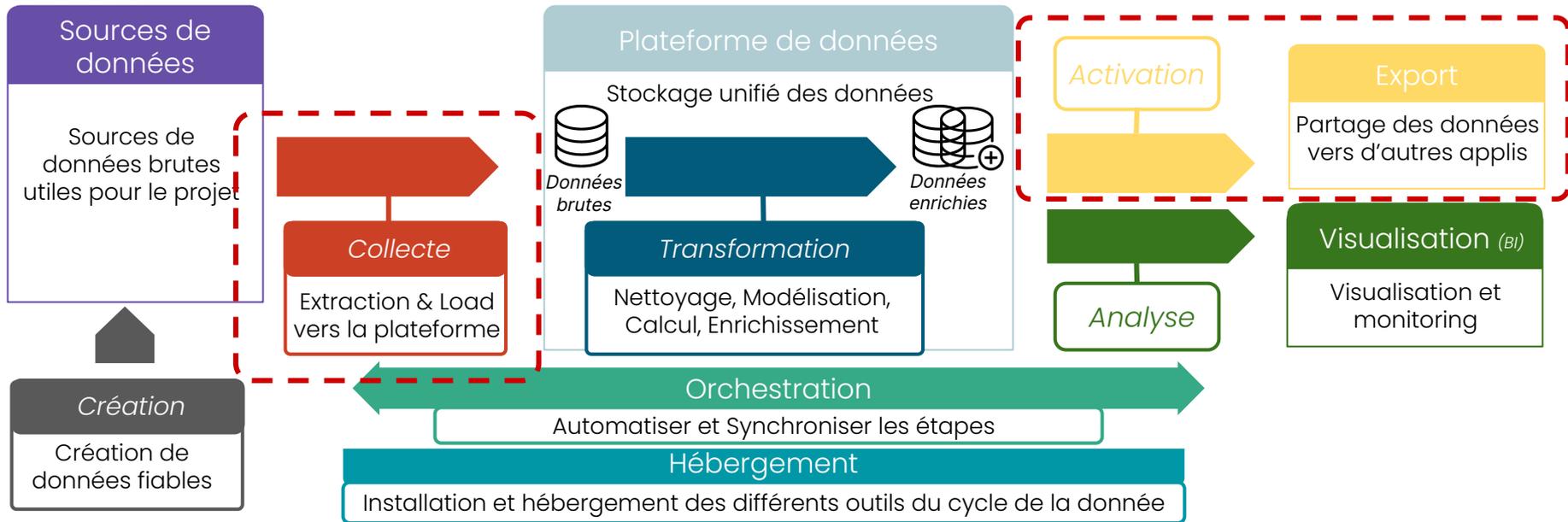


SERVICES DONNÉES

# Data engineer



Innovation - Écologie - Territoires



## Data engineer

Le plombier - gestion des flux + qualité des données.  
Imports & Exports, Accès & Partage

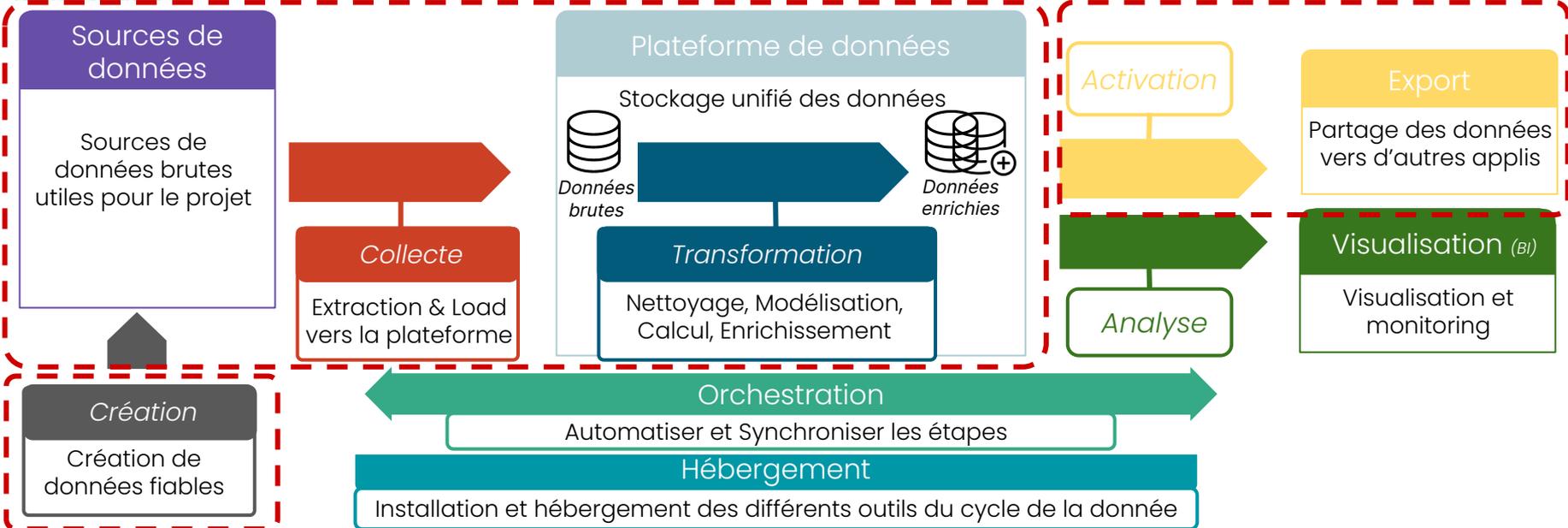


SERVICES DONNÉES

# Statisticien

ECO LAB

Innovation - Écologie - Territoires



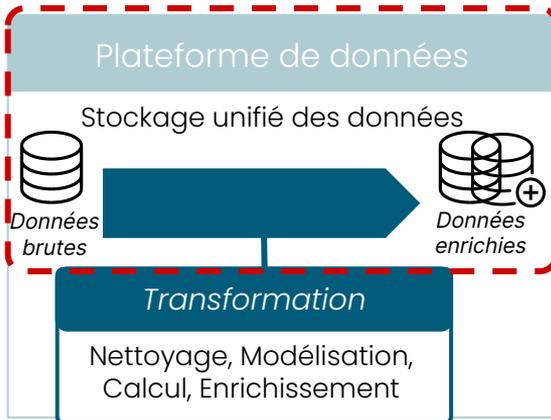
## Statisticien.ne

Mise en place et exploitation du système permettant la production de données statistiques



SERVICES DONNÉES

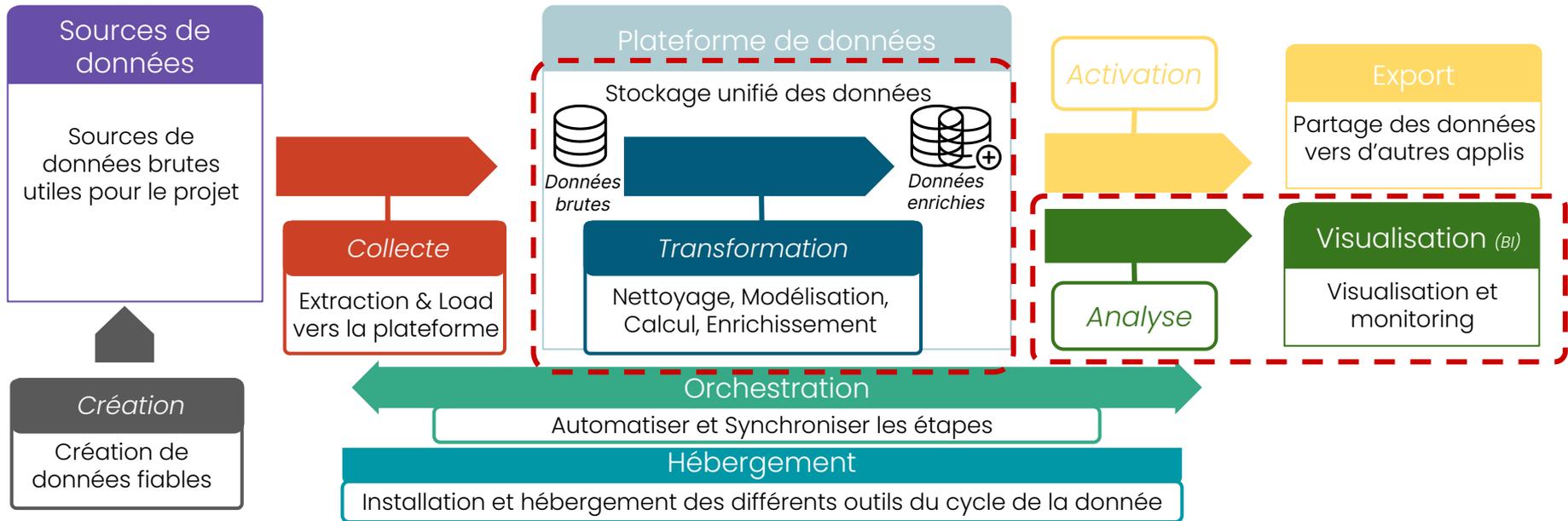
# Data engineer





SERVICES DONNÉES

# Data analyst / scientist



## Data scientist / Data analyst

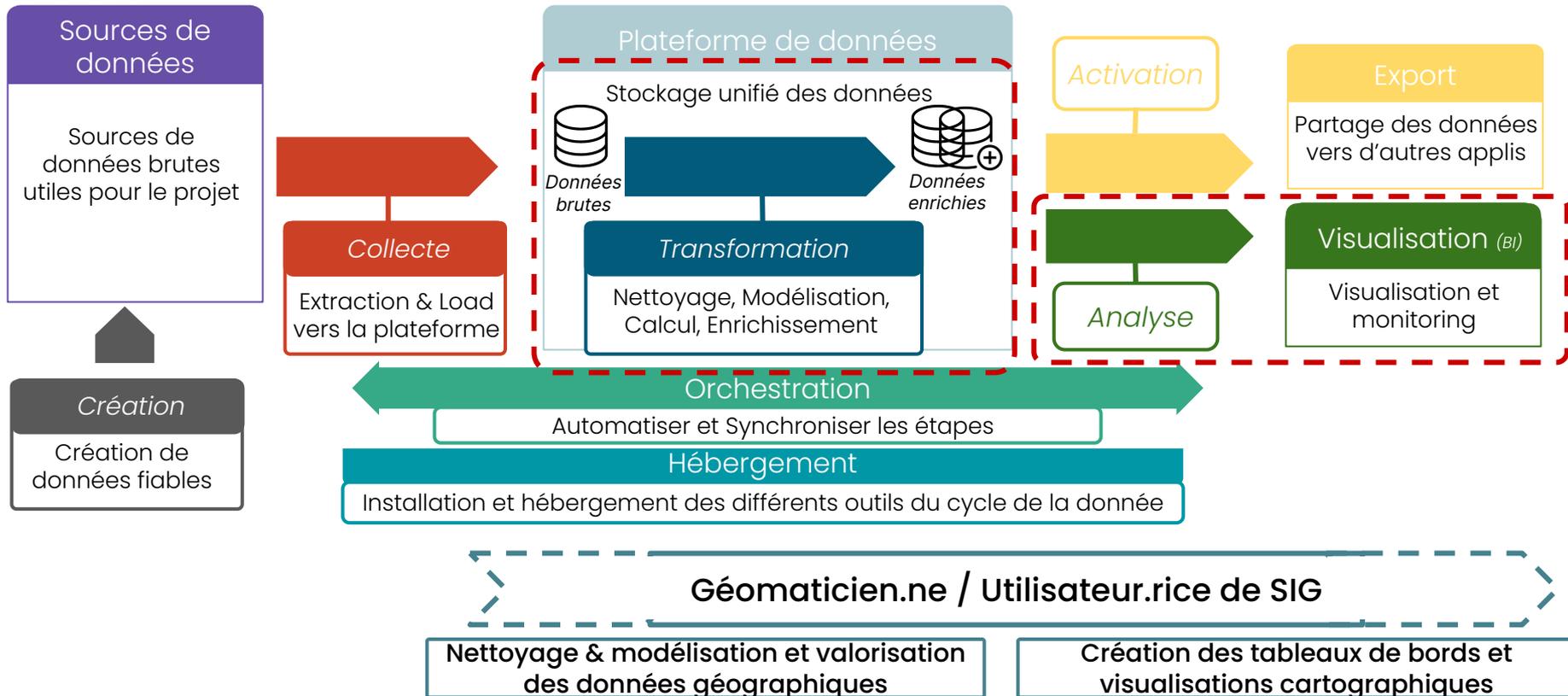
Nettoyage & modélisation data, calcul d'indicateurs et valorisation des données

Création de tableaux de bords efficaces pour l'analyse et le monitoring



SERVICES DONNÉES

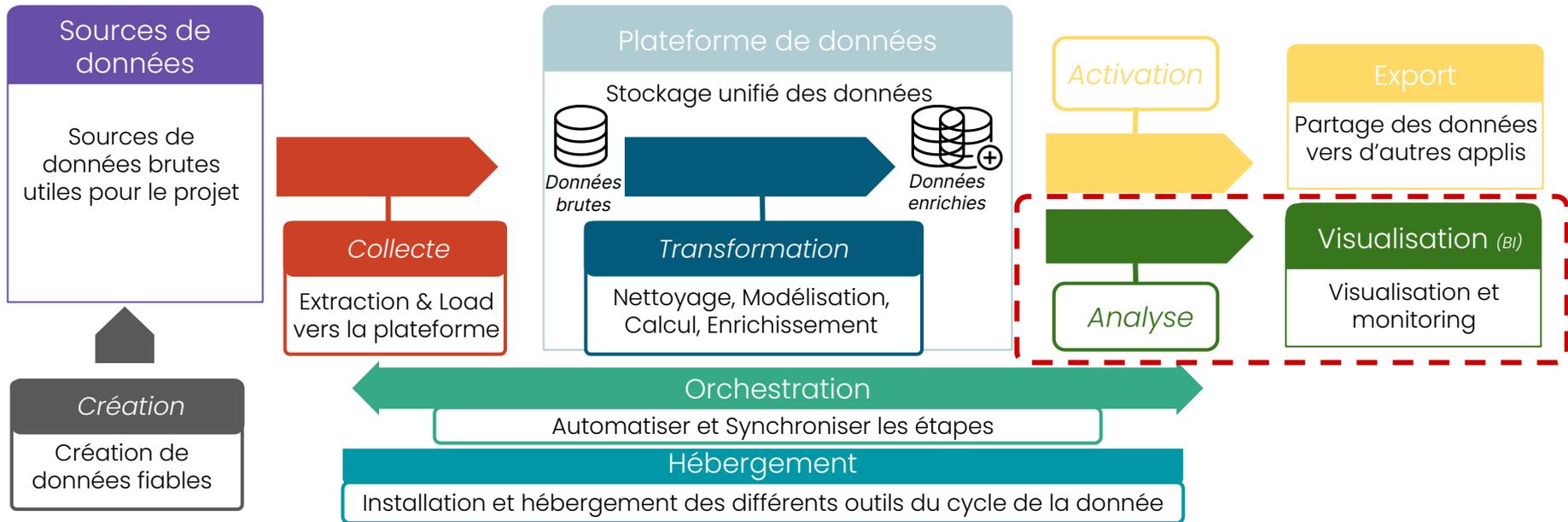
# Géomaticien.ne / Utilisateur.rice de SIG





SERVICES DONNÉES

# Developpeur.se Web



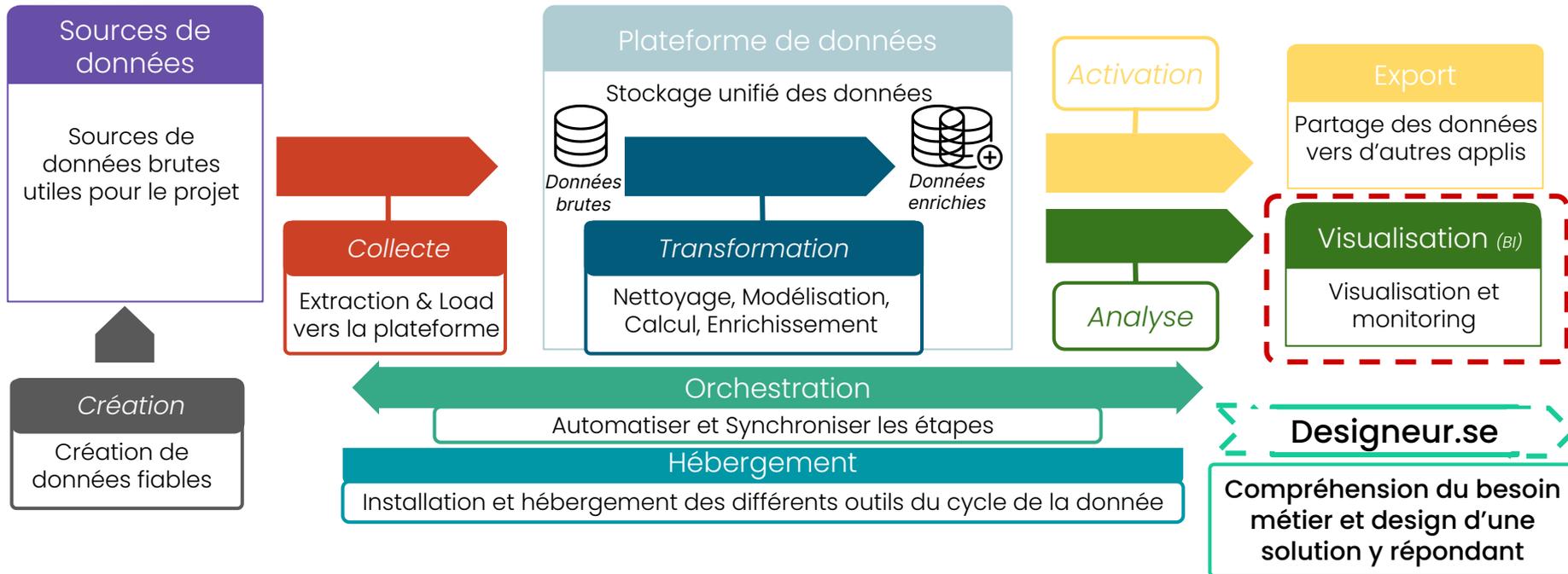
Développeur.se Web

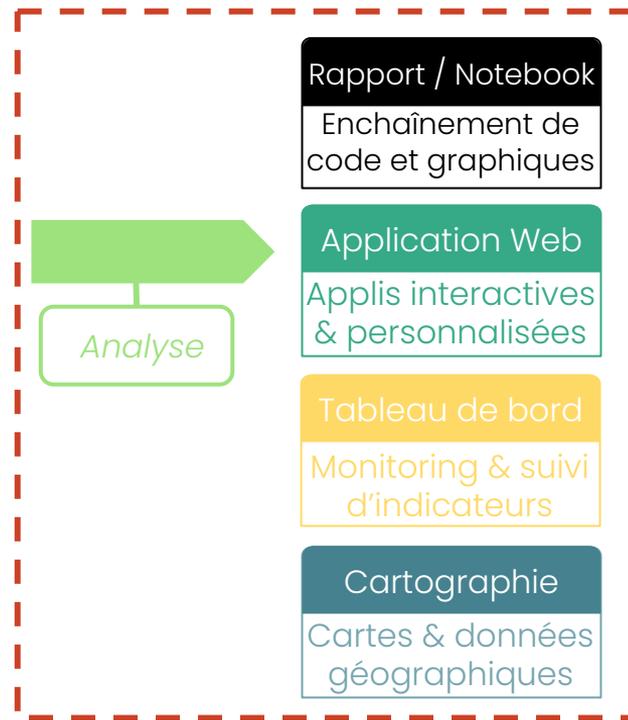
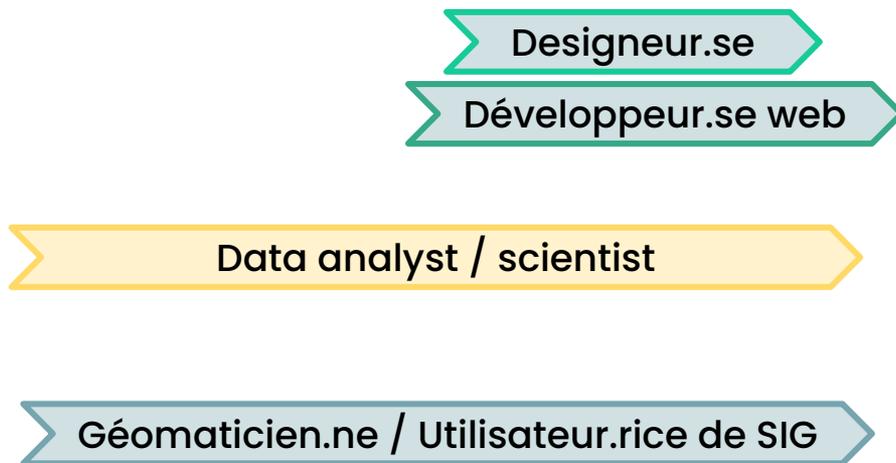
Conception d'applications web avancées



SERVICES DONNÉES

# Designeur.se





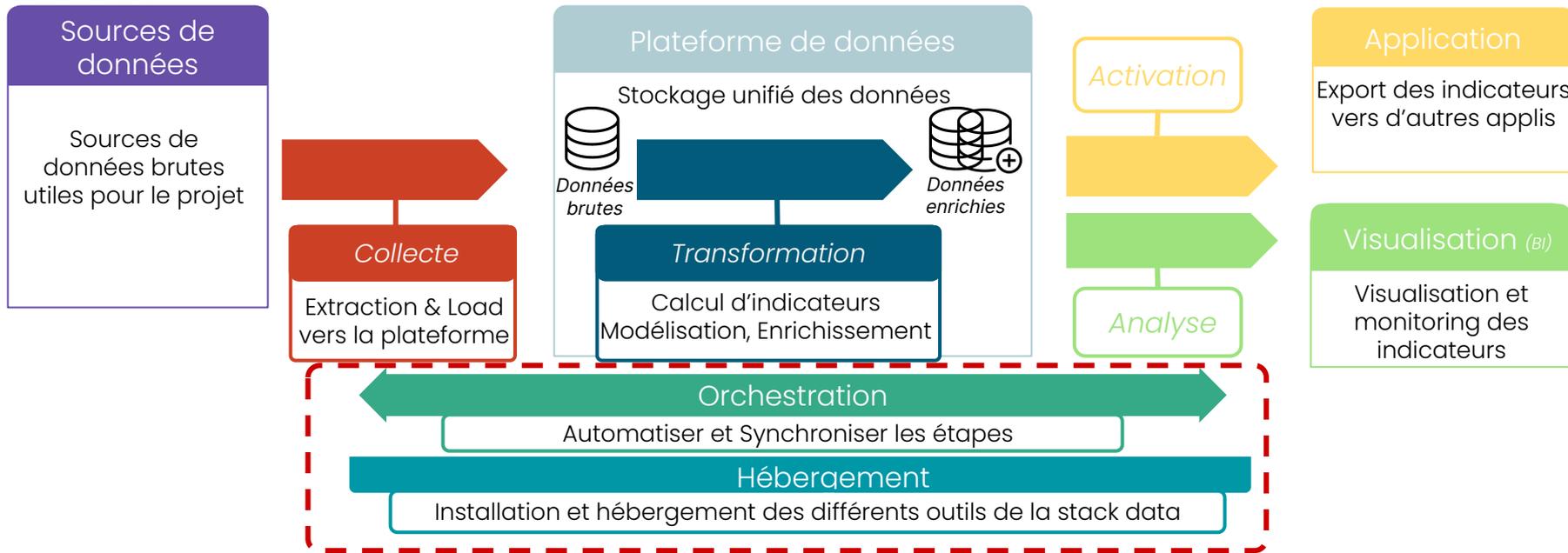


SERVICES DONNÉES

## Data ops



Innovation - Écologie - Territoires



## Data / Dev ops

Hébergement - Installe et maintient les outils data nécessaires aux autres équipes

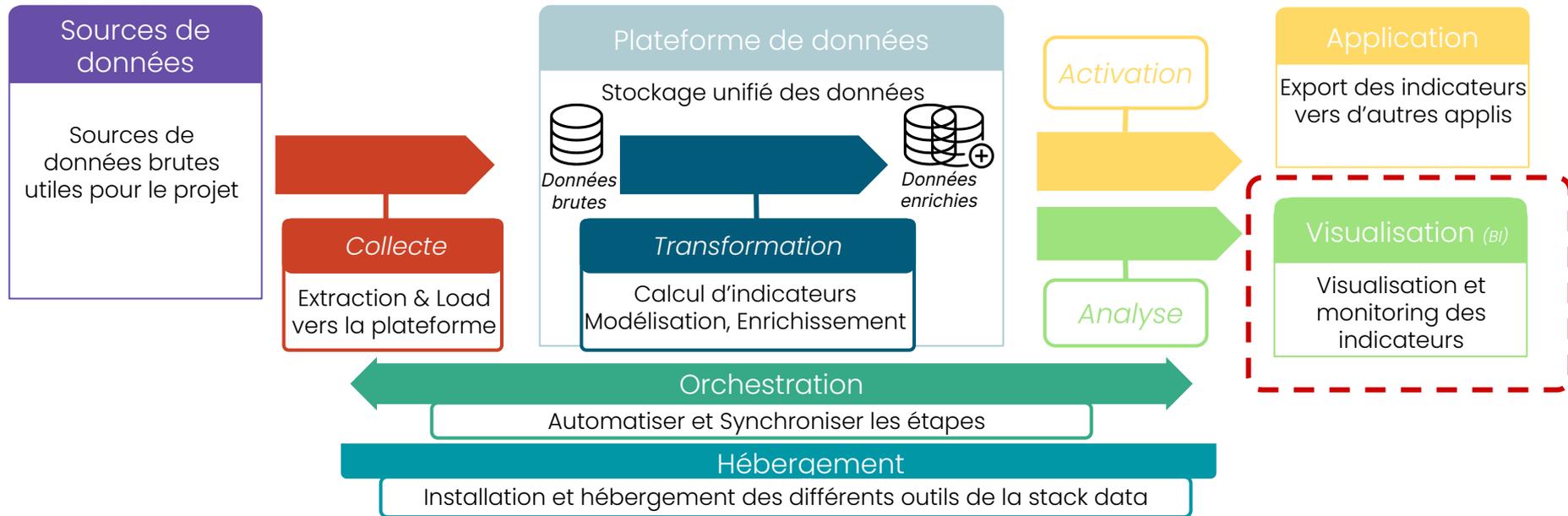


SERVICES DONNÉES

# Chef.fe de projet



Innovation - Écologie - Territoires



Chef.fe  
de projet

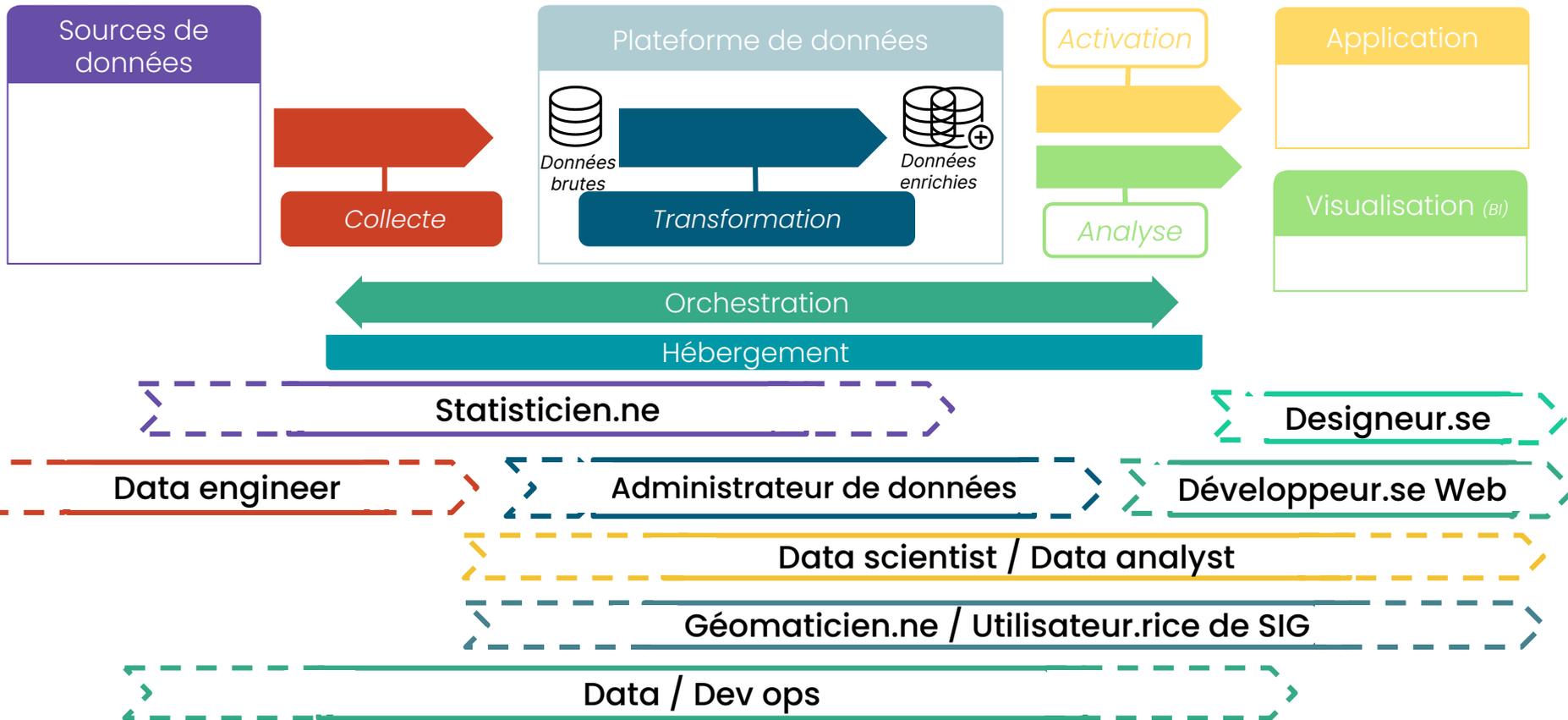


SERVICES DONNÉES

# Les métiers de la donnée

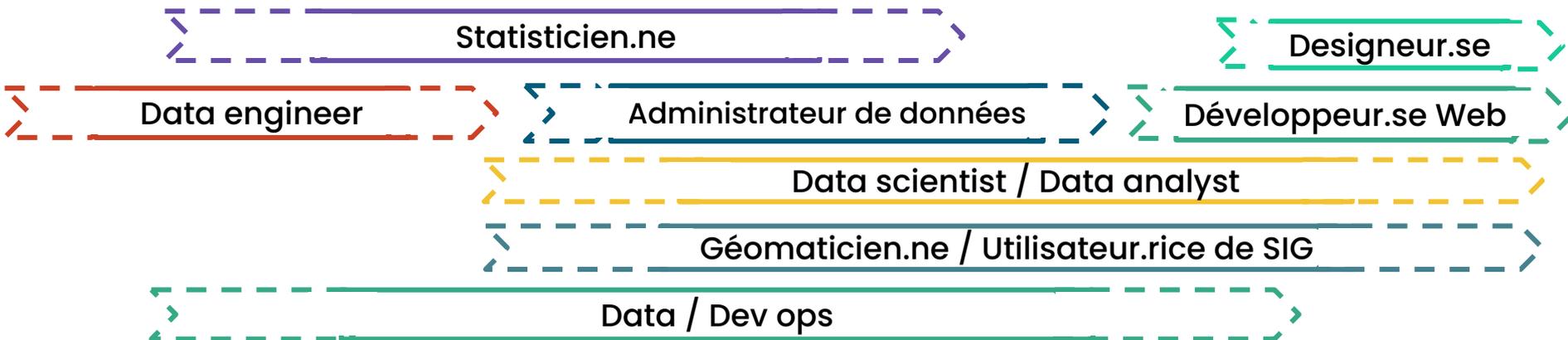
ECO LAB

Innovation - Écologie - Territoires



## A retenir

- Différentes compétences et métiers interviennent aux différentes étapes du cycle de la donnée
- Toutes les métiers n'interviennent pas forcément dans un projet selon les ressources et le besoin
- Aux mêmes étapes peuvent intervenir différents métiers selon le besoin et les choix techniques
- Les informations sont indicatives avec beaucoup de métiers intervenant à de multiples étapes



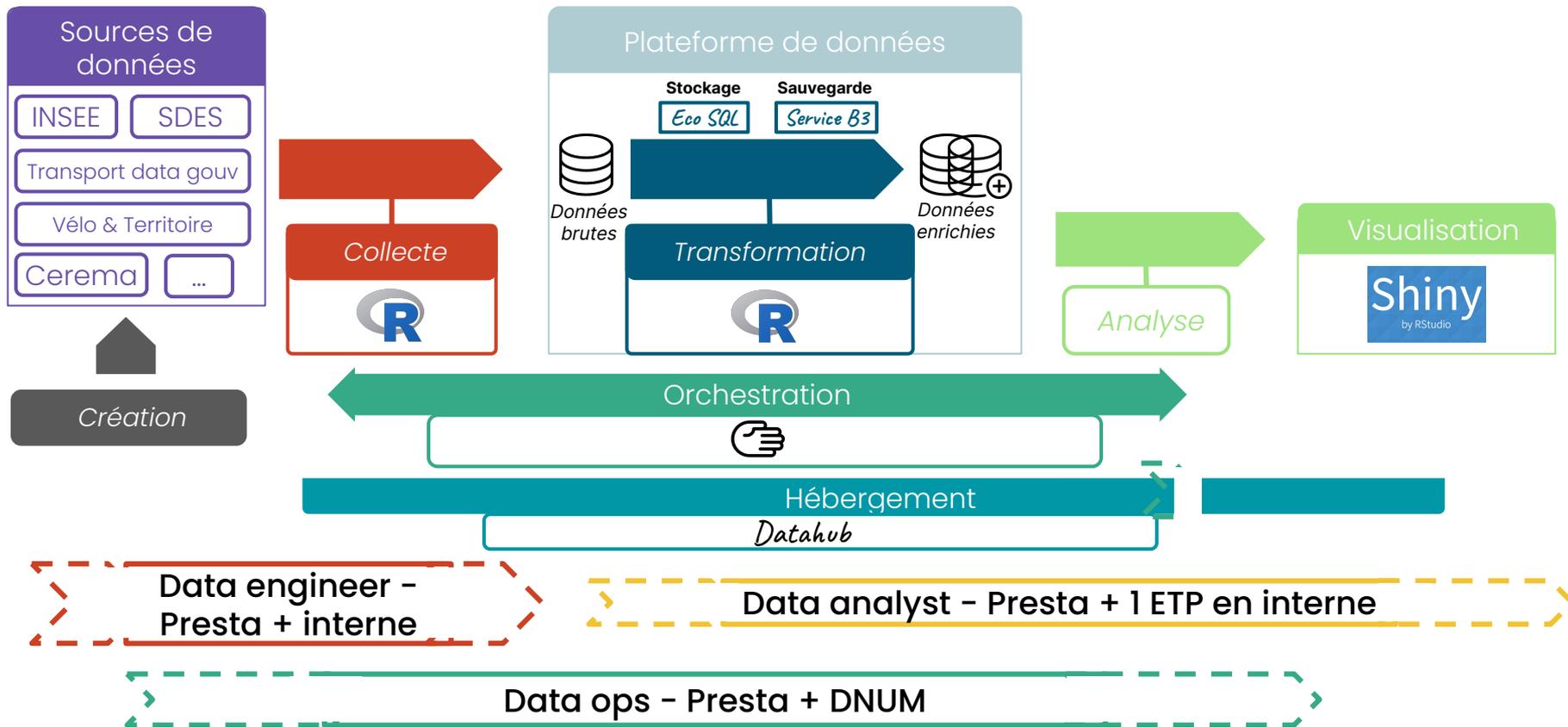
## Quelques exemples concrets



SERVICES DONNÉES

## Pôle Mobilité durable - Appli R-Shiny

ECO LAB  
Innovation - Écologie - Territoires



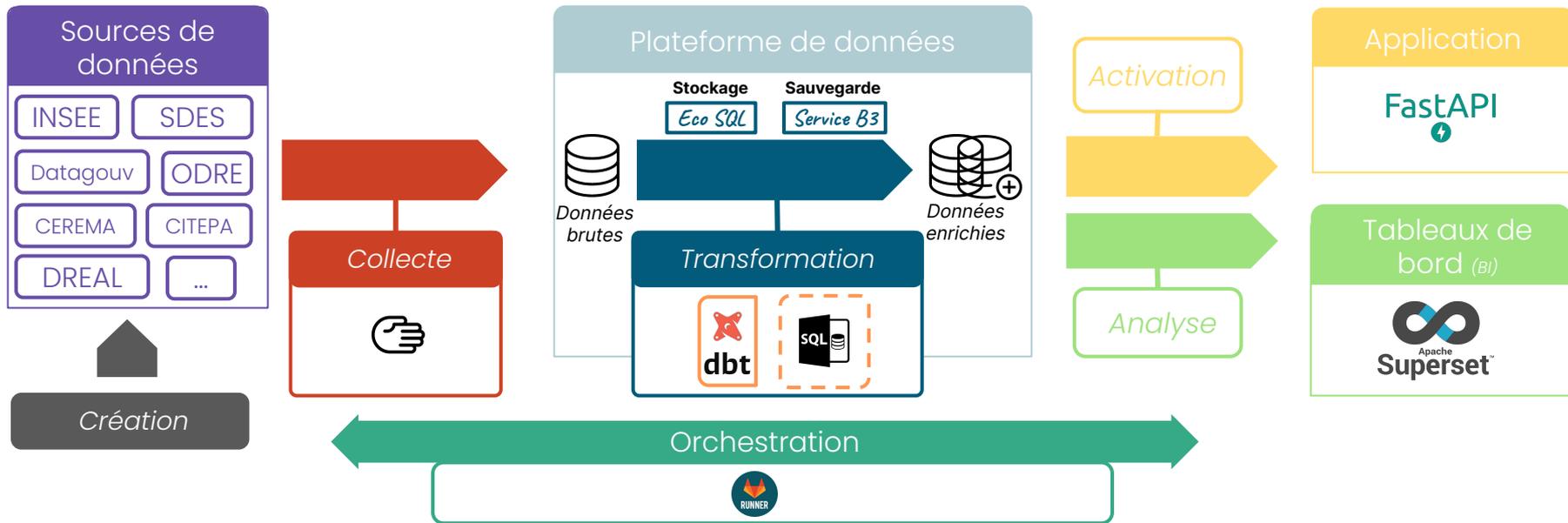


SERVICES DONNÉES

# Indicateurs Territoriaux



Innovation - Écologie - Territoires



1 data engineer

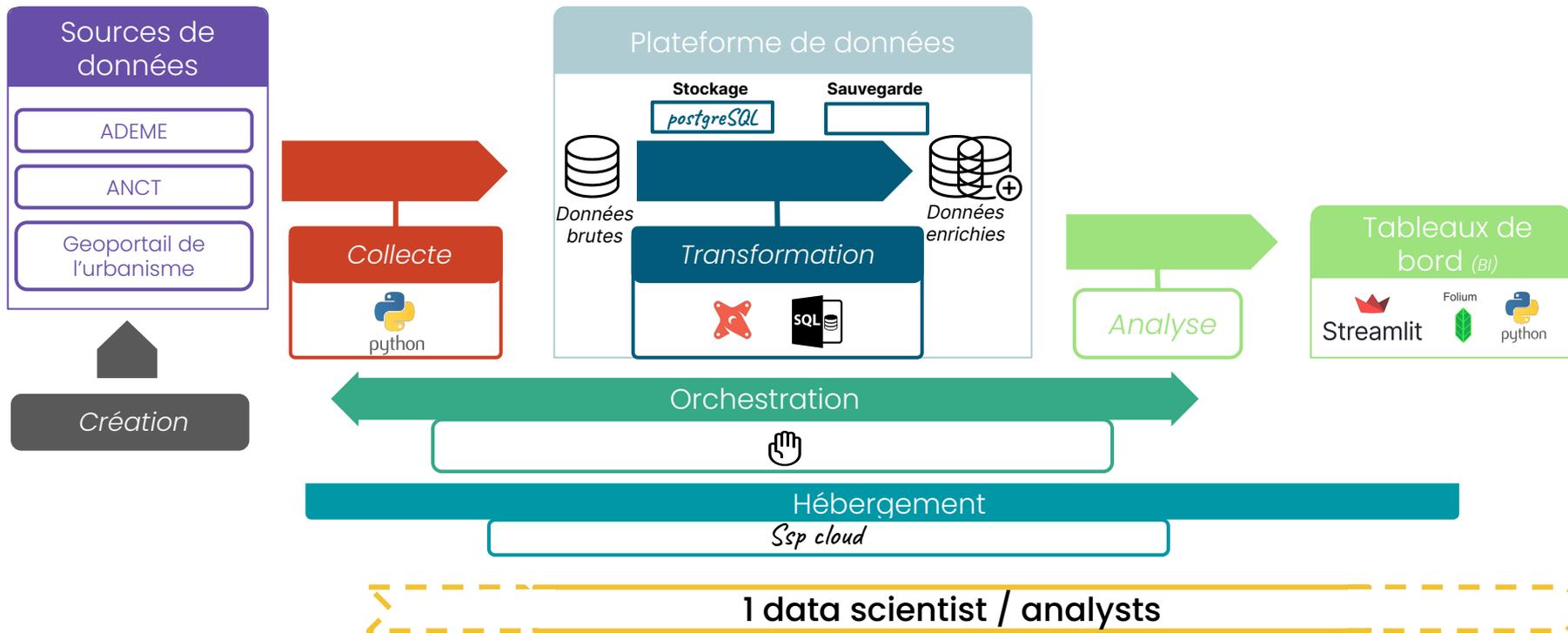
2 data analysts

Data ops - accompagnement DNUM



SERVICES DONNÉES

# Cartographie des documents du PCAET





SERVICES DONNÉES

# L'outil SIGNE



Innovation - Écologie - Territoires

**Sources de données**

Services métiers  
DREAL - ZGE  
Services régionaux de l'Etat

**Collecte**

**Plateforme de données**

Stockage: *postgreSQL* | Sauvegarde

Données brutes → Transformation → Données enrichies

**Activation**

**Analyse**

**Application**

**Tableaux de bord (BI)**

Création

**Orchestration**

Designeur.se

Administrateur de données

Data scientist / Data analyst

Géomaticien.ne / Utilisateur.rice de SIG

Data / Dev ops

**Des questions ?**