



**HAL**  
open science

## Geoflow : un workflow pour une gestion simple, FAIR et durable des données

Wilfried Heintz, Emmanuel Blondel, Julien Barde

### ► To cite this version:

Wilfried Heintz, Emmanuel Blondel, Julien Barde. Geoflow : un workflow pour une gestion simple, FAIR et durable des données. Diffuser ses (méta)données spatiales avec R geoflow, Apr 2021, Webinaire, France. pp.1-70. hal-03420835

**HAL Id: hal-03420835**

**<https://hal.inrae.fr/hal-03420835>**

Submitted on 9 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Geoflow : un workflow pour une gestion simple, FAIR et durable des données



Emmanuel Blondel, Julien Barde, Wilfried Heintz

[Article du numéro spécial du cahier des techniques 2019](#)

# Remerciements

- **Financeurs de développements**



- **Financeurs via projets**



Food and Agriculture Organization of the United Nations



**Blue-Cloud**  
Piloting innovative services for Marine Research & the Blue Economy

Blue-Cloud has received funding from the European Union's Horizon programme call BG-07-2019-2020,862409.

- **Financeurs de packages**

- geometa



**R consortium**

- geosapi

BlueBRIDGE



BlueBRIDGE received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 675680



Grant Agreement No.731011



Food and Agriculture Organization of the United Nations

# Partie 1 : introduction

# Contexte



2007



2018

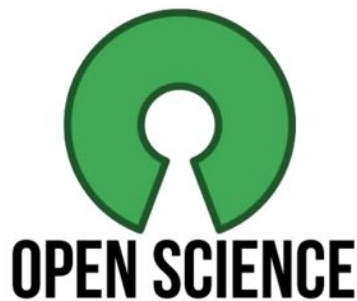
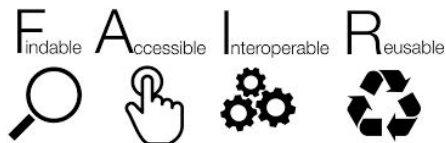
# Contexte



2007



2018



# Gestion de données : rappel

## Données de la Recherche

=



# Gestion de données : rappel

Données **sans** métadonnées (f**A**ir)

=





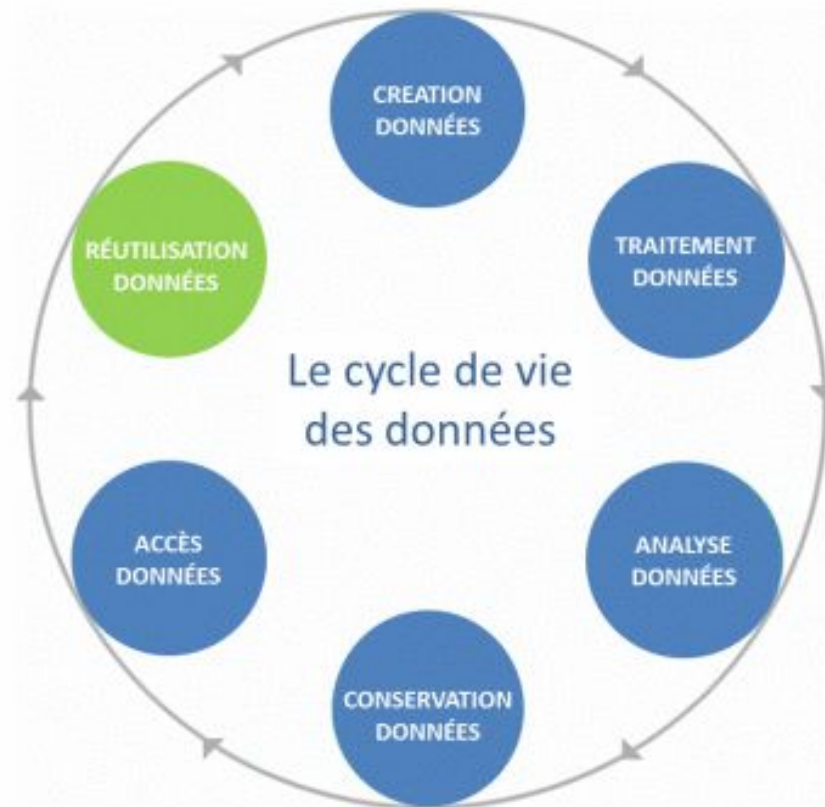
# Gestion de données : rappel

Métadonnées **sans** données (FaIR)

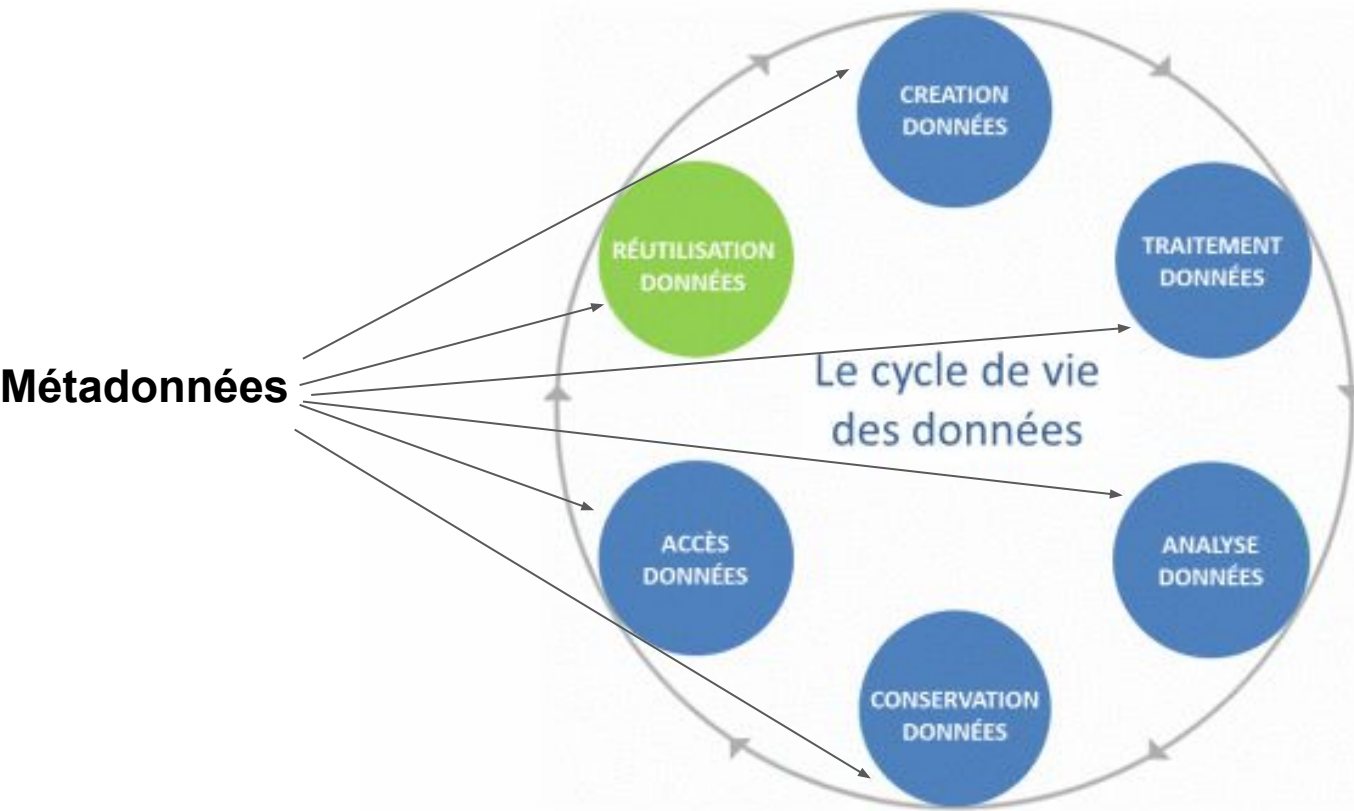
=



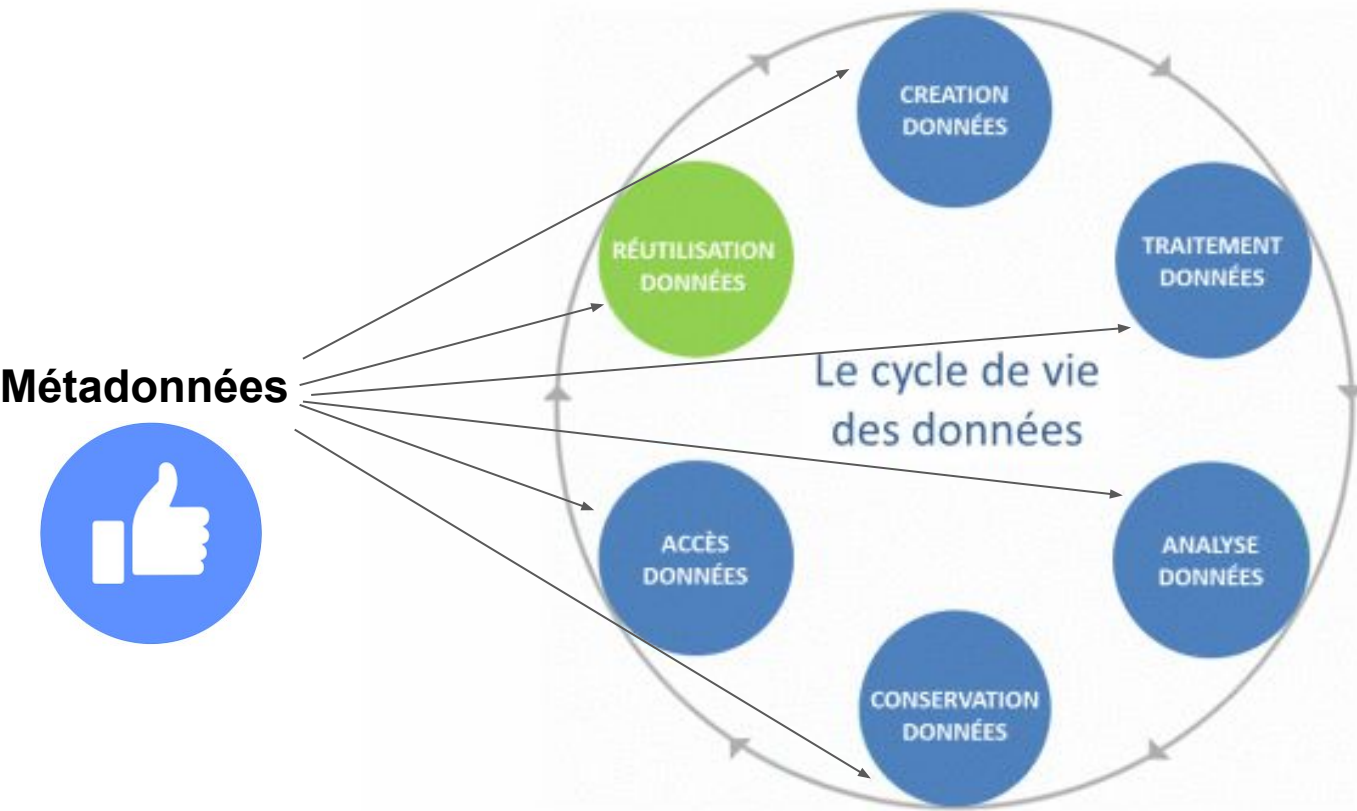
# Cycle de vie des données



# Cycle de vie des données



# Cycle de vie des données



# Une multitude d'outils / interfaces / standards ...



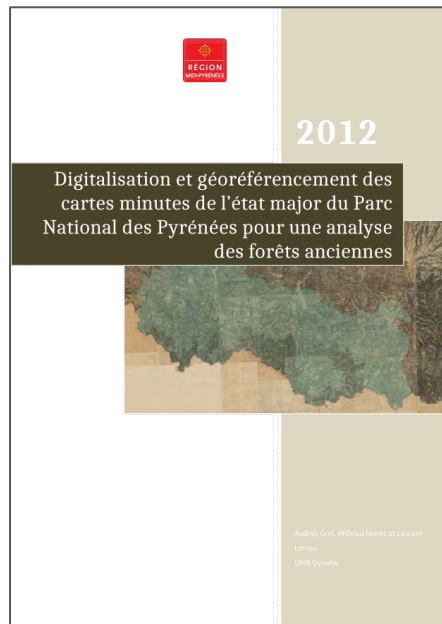
# Un exemple



# Un exemple



Publication



Cartographie

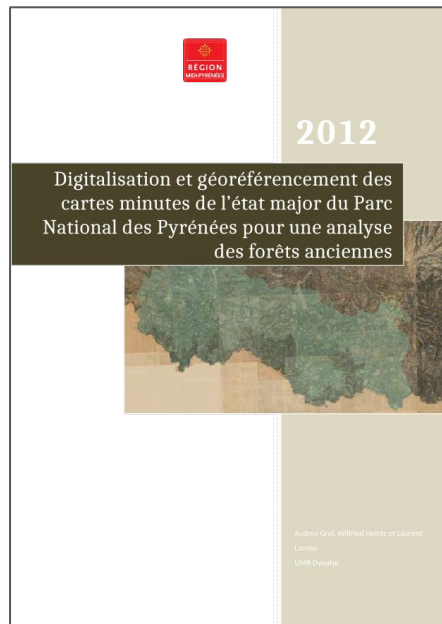


Métadonnées

# Un exemple

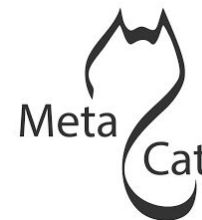


Publication



OpenLayers

Cartographie



Métadonnées

Minimum : 3 interfaces logicielles ... jusqu'à 6 ou plus



Back to search

&lt; Previous

Next &gt;

Download

Display mode

Browse

## Forêts anciennes des Pyrénées

Cartographie de l'occupation ancienne du sol du massif des Pyrénées réalisée à partir de la digitalisation des minutes de cartes d'Etat Major. (2012-04-01)

### Download and links



Digital Object Identifier

 DOI <http://dx.doi.org/10.15454/BNEUSM>

Open link



Cartographie des forêts anciennes de France - Objectifs bilan et perspectives

 publication <https://prodinra.inra.fr/record/386328>

Open link



Digitalisation et géoréférencement des cartes minutes de l'état major du Parc National des Pyrénées pour une analyse des forêts anciennes

 publication <https://prodinra.inra.fr/record/185677>

Open link



WMS

pyrenees

Forêts anciennes des Pyrénées

Add service layer to map

This dataset is published in the view service (WMS) available at [https://dynids.toulouse.inra.fr:8444/geoserver/forets\\_anciennes/ows?service=WMS](https://dynids.toulouse.inra.fr:8444/geoserver/forets_anciennes/ows?service=WMS).

Overview



pyrenees









Logo

DOI 10.15454/BNEUSM

### Serveur

-  État du service
-  Logs GeoServer
-  Information sur le point de contact
-  À propos de GeoServer

### Données

-  Prévisualisation de la couche
-  Espaces de travail
-  Entrepôts
-  Couches
-  Agrégations de couches
-  Styles

### Services

-  WMTS
-  WFS
-  WCS
-  WMS

### Configuration

-  Globale
-  JAI
-  Accès à une couverture

## Éditer la couche

Éditer les données de la couche et de publication

## forets\_anciennes:pyrenees

Configurer la ressource et les informations associées à la couche

**Données**

Publication

Dimensions

Cache de tuiles

## Éditer la couche

### Information de base sur les données

Nom

pyrenees

Activé

Publié

Titre

Forêts anciennes des Pyrénées (doi:10.15454/BNEUSM)

Résumé

Cartographie de l'occupation ancienne du sol du massif des Pyrénées réalisée à partir de la digitalisation des minutes de cartes d'Etat Major. (2012-04-01)

Dynafor dataverse (www.inra.fr)

Portail Data INRAE > Experimental - Observation - Simulation Dataverse > Dynafor dataverse > **Cartographie de l'occupation ancienne des sols du massif des Pyrénées**



Statistiques

50 téléchargements


 Contact  Partager

 Lier

 Modifier ▾

## Cartographie de l'occupation ancienne des sols du massif des Pyrénées Version 3.0

Heintz, Wilfried; Larrieu, Laurent; Grel, Audrey, 2019, "Cartographie de l'occupation ancienne des sols du massif des Pyrénées", <https://doi.org/10.15454/BNEUSM>, Portail Data INRAE, V3

 Citer le dataset ▾

 Pour en apprendre davantage sur le sujet, consulter le document [Data Citation Standards \[en\]](#).

### Description

Cartographie de l'occupation ancienne du sol du massif des Pyrénées réalisée à partir de la digitalisation des minutes de cartes d'Etat Major. (2012-04-01)

### Subject

Forests and Forest Products; Biodiversity and Ecology

### Mot-clé

Occupation du sol, Forêts anciennes, Digitalisation, Pyrénées, Minutes d'Etat Major

### Related Publication

Grel, A., Heintz, W., LARRIEU, L. (2012). Digitalisation et géoréférencement des cartes minutes de l'état major du Parc National des Pyrénées pour une analyse des forêts anciennes. 18 p.

### Link to data

[https://dynafor.toulouse.inrae.fr/dataviewer/?&baseview=World%20Imagery&views=%5B%22pid%3Dpyrenees%2Cstrategy%3Dogc\\_filters%2C%22%5D&extent=-3.0647675384868647,41.343502497685755,3.9664824615131353,44.760250544560755&center=0.45085746151313544,43.051876521123255&zoom=9](https://dynafor.toulouse.inrae.fr/dataviewer/?&baseview=World%20Imagery&views=%5B%22pid%3Dpyrenees%2Cstrategy%3Dogc_filters%2C%22%5D&extent=-3.0647675384868647,41.343502497685755,3.9664824615131353,44.760250544560755&center=0.45085746151313544,43.051876521123255&zoom=9)

### Query

#### Attributs

Occupation du sol

En plaine, culture - En montagne, alpages ou rochers [11]

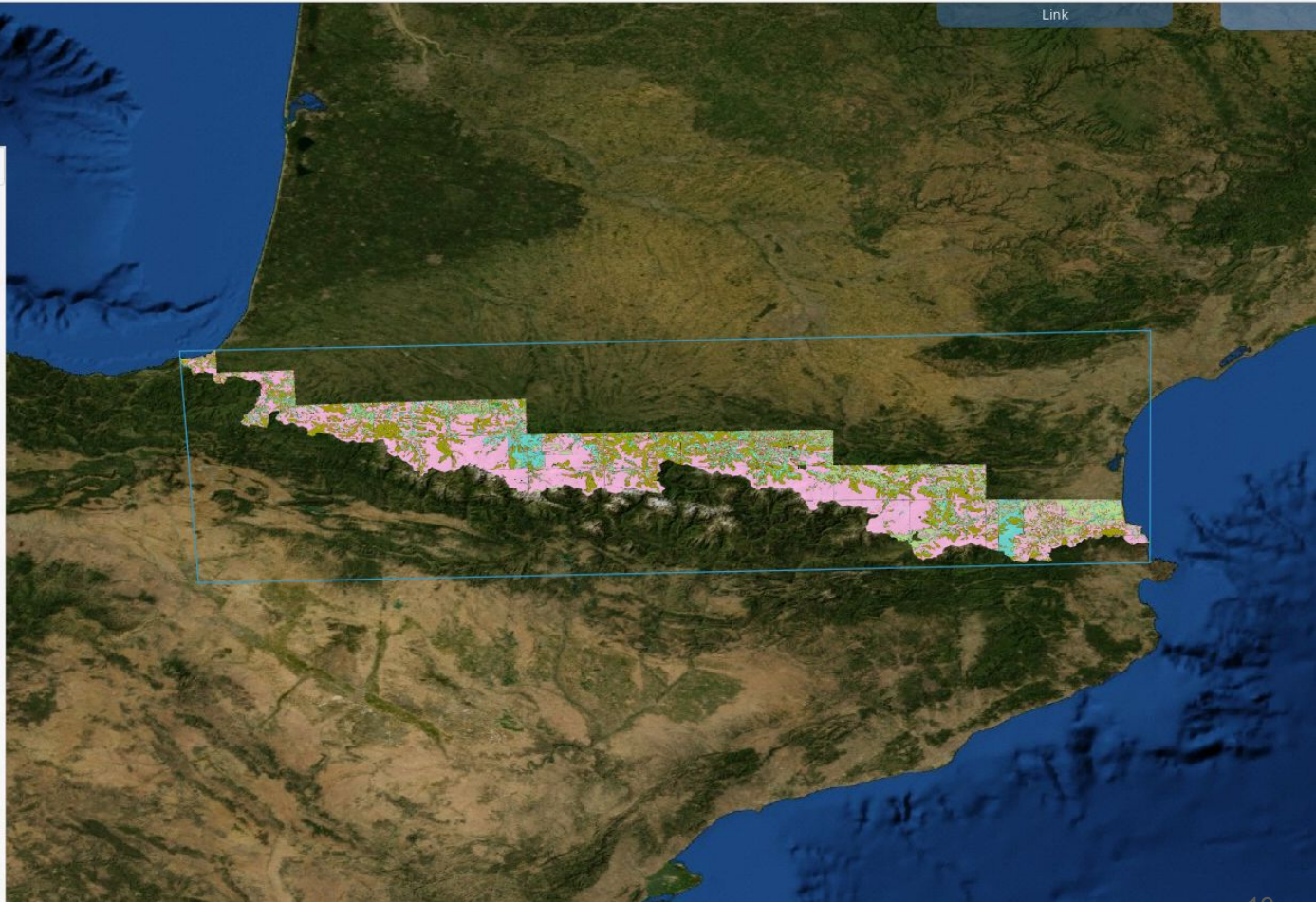
Village [7]

Forêt [1]

Pâquis, pâtures [13]

Rivières [32]

Prairie [3]





▼ Date de mise à jour

▼ Format de téléchargement

▼ Droits d'usage

Gratuits

## 13 ensembles de données trouvés

**Cartographie de l'occupation ancienne des sols du massif des Pyrénées**[search.datacite.org](https://search.datacite.org)

Dernière mise à jour : 2019

D

**[DREAL OCCITANIE] Trait de Côte de 1850 à 2003 sur le littoral en Occitanie**[www.data.gouv.fr](http://www.data.gouv.fr)

+4

Dernière mise à jour : 12 juil. 2019

E

**Service de visualisation cartographique (WMS) du jeu de données : Surface...**[www.europeandataportal.eu](http://www.europeandataportal.eu)

wms

Dernière mise à jour : 8 avr. 2018

**Cartographie de l'occupation ancienne des sols du massif des Pyrénées**[Découvrir sur search.datacite.org](https://search.datacite.org)**Identifiant unique**<https://doi.org/10.15454/bneusm>**Ensemble de données mis à jour** 2019**Ensemble de données fourni par**[DataCite](#)

Portail Data INRAE

**Auteurs**

Laurent Larrieu; Audrey Grel; Wilfried Heintz

**Description**

Cartographie de l'occupation ancienne du sol du massif des Pyrénées réalisée à partir de la digitalisation des minutes de cartes d'Etat Major.







fx

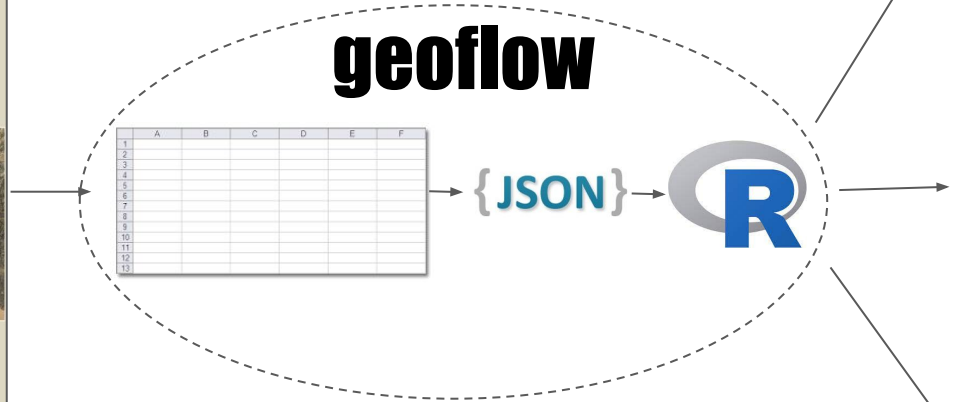
Identifier

	A	B	C	D	E
1	<b>Identifier</b>	<b>Title</b>	<b>Description</b>	<b>Subject</b>	<b>Creator</b>
2	id:pyrenees_ doi:10.15454/BN	Forêts anciennes des Pyrénées	Cartographie de l'occupa	GENERAL:Occupation du sol, Forêt	metadata:wilfried.heintz@inrae.fr_ owner:laurent.larrieu@inrae.fr,wilfried.heintz@inrae.fr_ pointOfContact:wilfried.heintz@inrae.fr

1 tableau source + 1 fichier de configuration + 1 commande R → ***n* actions**



# geoflow : principe de fonctionnement



Cartographie



Métadonnées



# Historique de geoflow : principes généraux

- Optimisation progressive de codes (ateliers Agropolis 2018, Zones Ateliers..)
  - Codes basés sur des paquets plus ciblés (CRAN): geometa, ows4R, geonapi, geosapi, zen4R
  - ⇒ paquet geoflow sur [repository github](#), à terme sur CRAN
- Implémentation orientée sur la donnée géospatiale mais modèle générique conçu pour gérer d'autres standards (DCMI, Datacite, OGC, TDWG..),
- Gestionnaire de workflow:

Orchestrer l'exécution d'1..\* **action(s)** appliquée(s) à 1..\* **entité(s)** en utilisant 1..\* **logiciel(s)** :

- **action** = traitement + paramètres
- **entity** = métadonnée + donnée
- **software** = logiciel (composant d'une infrastructure)

# Historique de geoflow : principes généraux

## Principaux objectifs

- Approche orientée métadonnée (concept central / input du workflow)
- Créer des services sur la (méta)donnée
  - Services FAIR (implémentation de DMPs) : découverte, accès, interopérabilité (OGC, TDWG..), visualisation, utilisabilité...
  - Services spécifiques (ex: ETL..)
- (Presque) s'affranchir des interfaces graphiques
  - Saisie des métadonnées dans un tableau “.csv” ... mais édition collaborative possible (GDrive)
  - Configuration du workflow avec un fichier “.json”... mais création d'une Shiny app (en cours)
  - Exécution du workflow avec des scripts “.R”
- Science ouverte & reproductible : dans une même table, les métadonnées capturent la description des données mais aussi des traitements avec leurs paramètres.

# Historique de geoflow : pour quels cas d'utilisation ?

## Automatiser & synchroniser l'exécution de tâches récurrentes

- À partir de sources de données classiques (entités):
  - Semi-structurées : csv, shp, geopackage ...
  - Structurées : SQL, NetCDF ...
  - Standardisées OGC (CSW, WMS, WFS ...), TDWG (EML, DwC) ...
- Actions qui créent des services classiques (FAIR):
  - Métadonnées : **Découverte** (ex: Geonetwork / CSW, Metacat..) & **Mapping** entre standards
  - **Accès** aux données (ex: Geoserver)
  - **DOIs** (ex: Zenodo, Dataverse)
  - **Visualisation & requêtage**: Shiny, Openlayers, Open Fair Viewer, Jupyter notebook ...
- Actions qui créent des services spécifiques
  - Charger une base de données à partir de différents fichiers (~ ETL)
  - Exécuter des scripts (analyse de données..)

# Historique de geoflow : problèmes identifiés

## Travail encore récent, meilleure prise en main si :

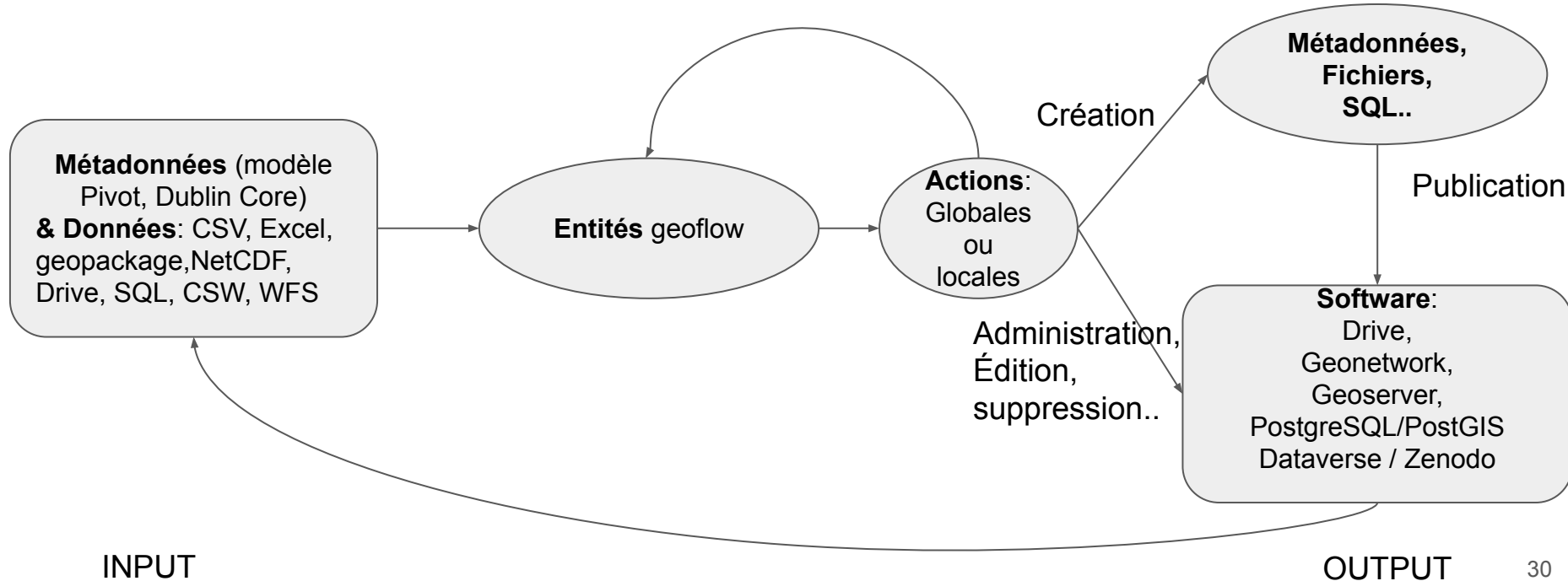
- Documentation du travail en cours (manuel utilisateur):
  - Modèle pivot de métadonnées ⇒ documenter la syntaxe pour la valuation contrôlée,
  - Configuration du workflow (fichier json)
  - Exemples reproductibles (DOIs sur releases de codes & versions de tableaux de données)
- Consolidation:
  - procédure d'installation + listing des erreurs connues
  - actions (dataverse, zenodo, EML)
- Assurance Qualité (QA)
  - Tests d'intégration
  - Publication du paquet (repository github) sur CRAN
- Interface graphique (R Shiny) pour éditer & exécuter le json de configuration,
- Organisation d'ateliers pour aider les utilisateurs / analyser les besoins.

# Les concepts de geoflow : terminologie

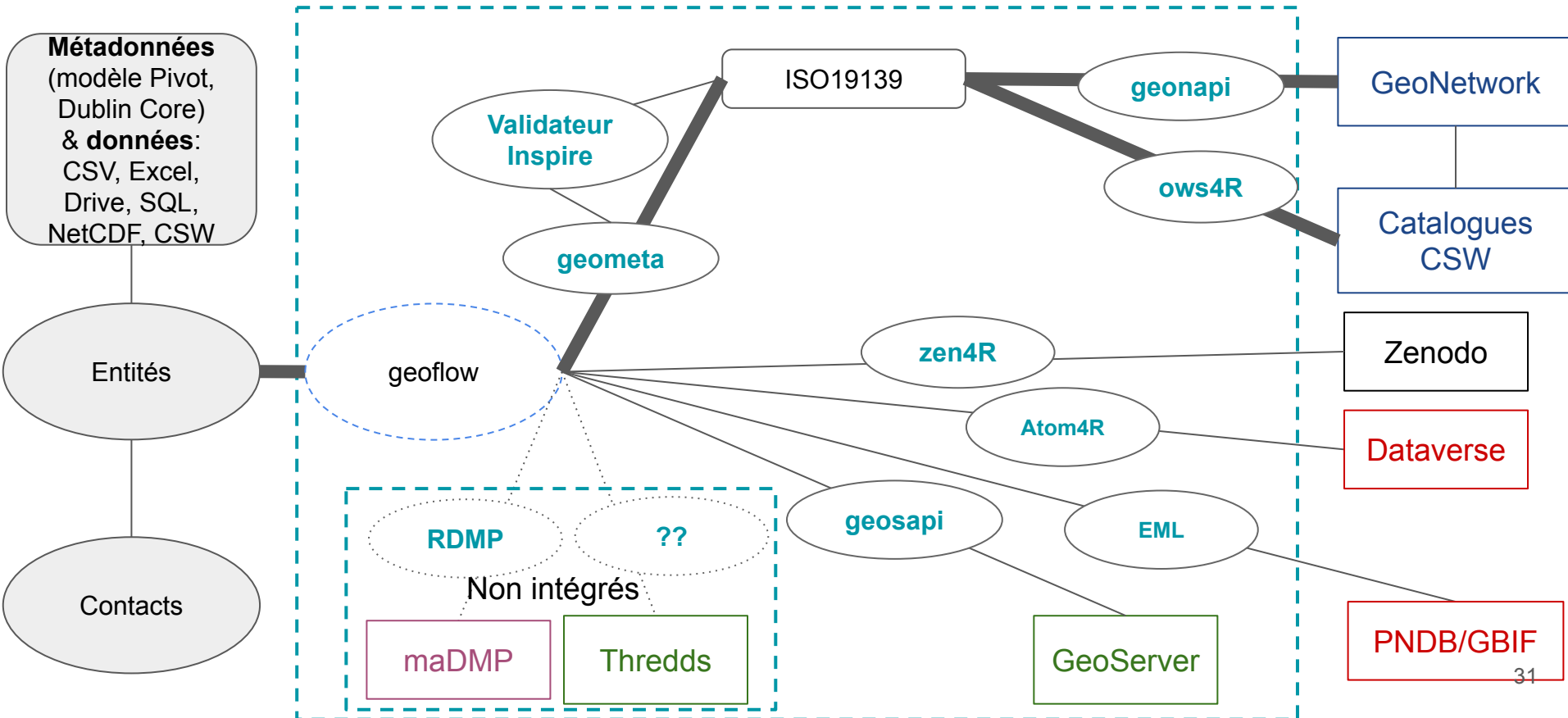
Orchestrer l'exécution d'1..\* **action**(s) appliquée(s) à 1..\* **entité**(s) en utilisant 1..\* **software**

- **Action**
  - **actions globales** (appliquées à toutes les entités) ou **locales** (ciblent une entité spécifique)
  - **actions natives** de geoflow ou **actions spécifiques** de l'utilisateur (inconnues de geoflow).
- **Entité** (input) = métadonnées + données
  - **métadonnées** = 16 éléments du modèle pivot (~ DCMI)
  - **données** semi-structurées (ex: fichiers plats: csv, shp..) ou structurées (ex: résultats de requêtes SQL / OPeNDAP / WFS..)
- **l'Output** d'une action peut être :
  - un nouvel objet: une fiche de métadonnée (ex: 19139 / XML), un rapport..
  - une nouvelle entité (qui peut devenir input d'une autre action..)
- **Software** : composant d'infrastructure de données (IDS, VRE..): input ou output

# Les concepts de geoflow : schéma d'un workflow

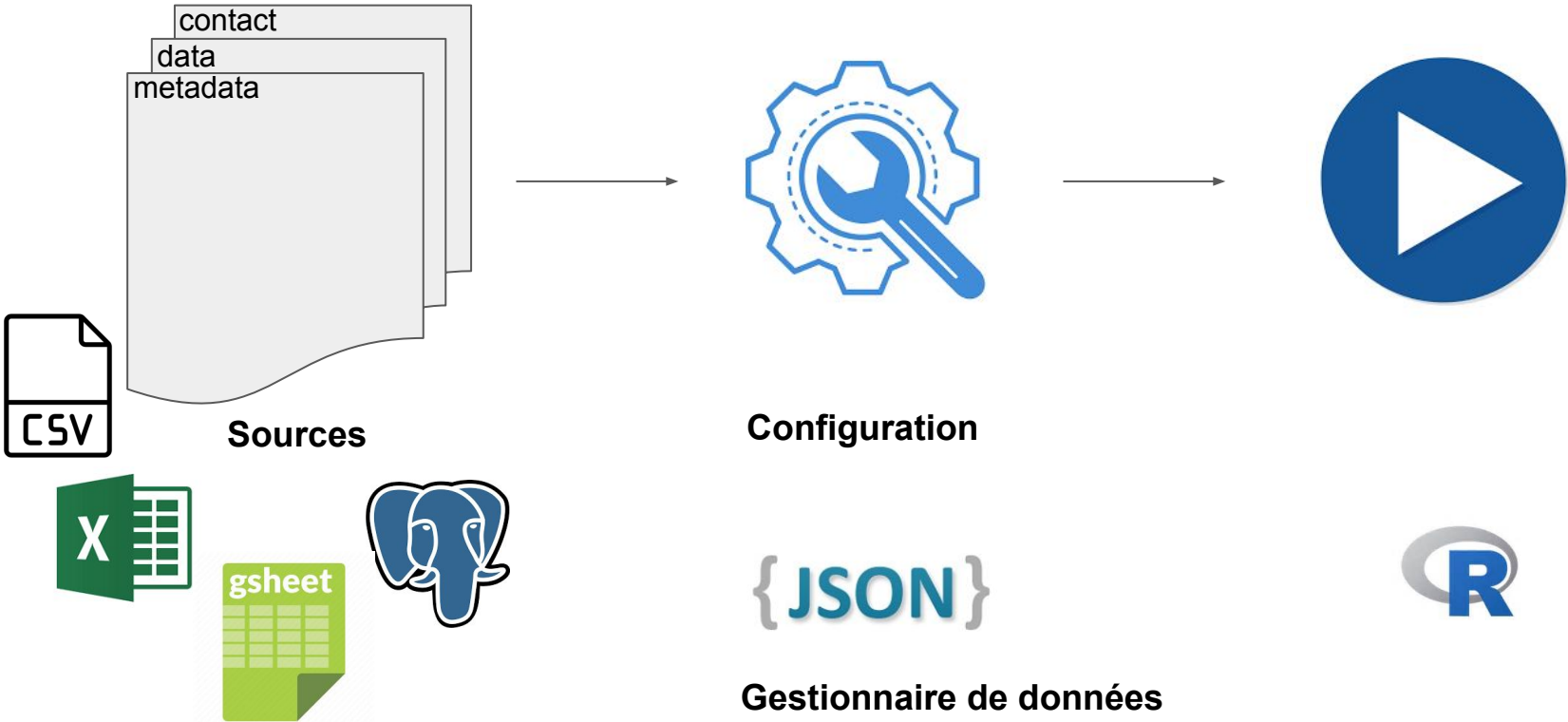


# geoflow : exemples d'actions pour les workflows

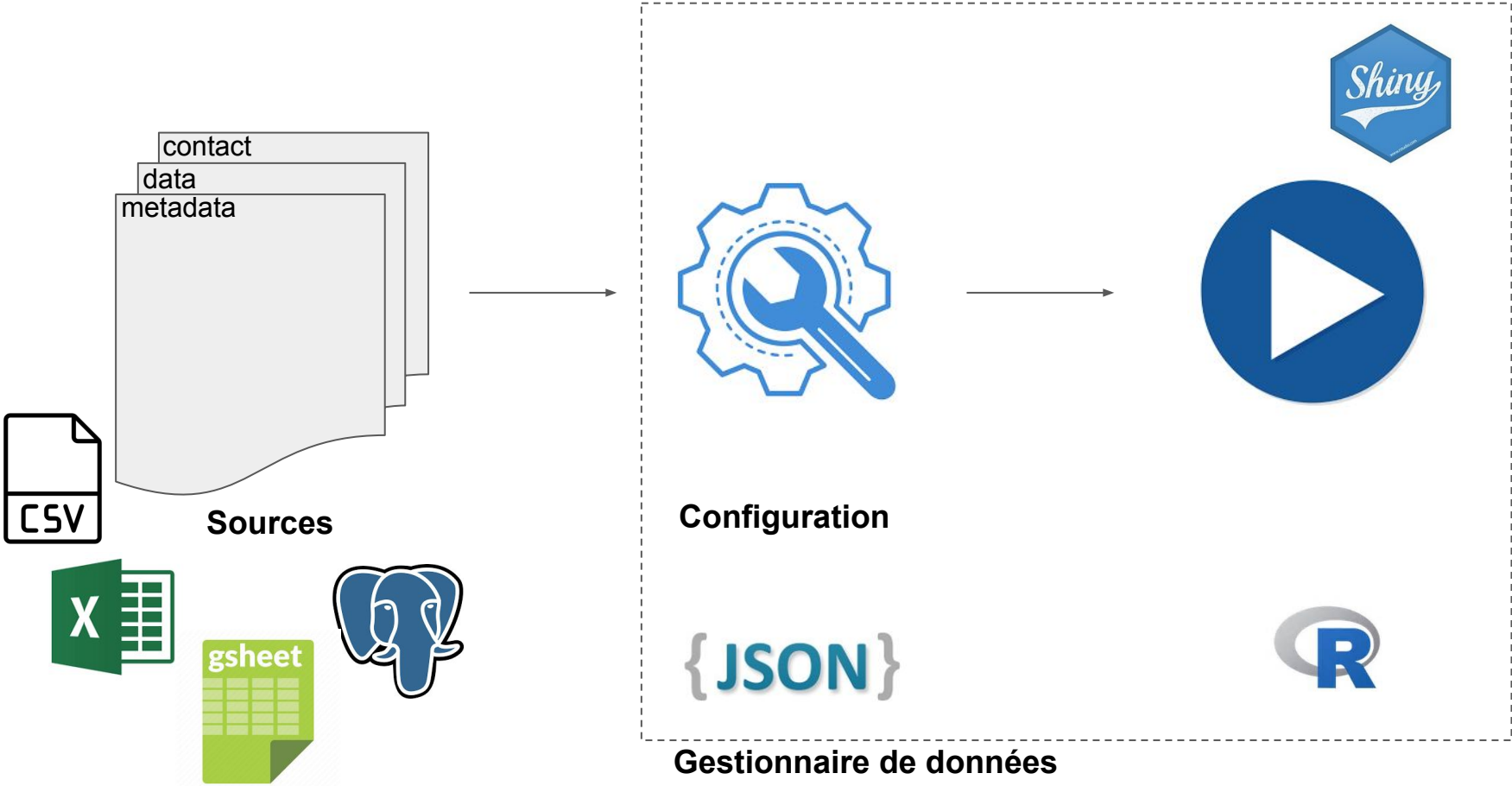




# geoflow : matériel à préparer pour lancer un workflow



# geoflow : matériel à préparer pour lancer un workflow



# Les concepts de geoflow : métadonnée & modèle pivot

Un modèle de métadonnée pivot avec valuation contrôlée, 16 éléments (11+3+1+1) et deux tableaux pour décrire les entités:

- **Métadonnées de découverte:**
  - 11/15 DCMI elements: Identifier, Title, Description, Subject, Date, Format, Language, Relation, Rights, Source, Type.
  - 3 éléments DCMI terms :
    - *SpatialCoverage* et *TemporalCoverage* spécialisent coverage (implicite, 1/15 DCMI)
    - *Provenance* + typage des valeurs
  - 1 élément Creator détourné pour gérer tous les contacts:
    - remplace 3/15 DCMI (contributor, publisher, creator) par un typage des valeurs
    - valué par une clé du **tableau qui décrit les contacts** (json): ~ “Contact” 19115 / EML
- **Métadonnées d’usage:** décrit l’accès à la donnée et sa **structure**
  - 1 élément ajouté : **Data** pour décrire l’accès au dataset et une structure basique,
  - le dictionnaire de données peut être décrit dans des tableau(x) dédié(s) référencés dans le json<sup>34</sup>

# Les concepts de geoflow : métadonnée & modèle pivot

Valuation contrôlée de “**Creator**”, typage des valeurs de contacts :

**owner**:FIRMS-Secretariat@fao.org\_

**pointOfContact**:FIRMS-Secretariat@fao.org\_

**publisher**:FIRMS-Secretariat@fao.org\_

**dsd**:FIRMS-Secretariat@fao.org\_

**principalInvestigator**:ird@ird.fr\_

**metadata**:paul.taconet@ird.fr,julien.barde@ird.fr,emmanuel.blondel@fao.org

# Les concepts de geoflow : métadonnée & modèle pivot

## Ne saisir que le strict nécessaire

- **Métadonnées statiques:**
  - saisies par les humains / non automatisables (ex: titre, résumé, certains mots-clés..)
- **Métadonnées dynamiques** : peuvent être calculées en lisant la donnée
  - ex: spatial, temporal, taxonomic extent...
  - DOI=Datacite=DCMI
  - ORCID / LDAP = métadonnées de Contact
- **Données statiques (csv..) vs données dynamiques (SQL):** colonne “Data”

Beaucoup d'éléments sont ainsi gérés à partir d'une saisie synthétique

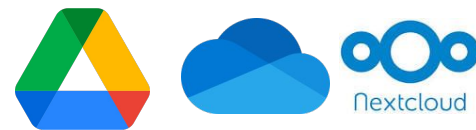
- 16 éléments + tableaux (Contact & dictionnaires de données)
- Syntaxe contrôlée avec typage des valeurs
- Métadonnées dynamiques / déduites de la donnée (statique ou dynamique)

# Les concepts de geoflow : format & outils basiques

## Gestionnaires et utilisateurs collaborent sur un tableau

- \* one row = description of one dataset
- \* one column = one metadata element

TITLE	ABSTRACT		...	AUTHORS
Data source 1				
...				
Data source X	Blabla			John Smith



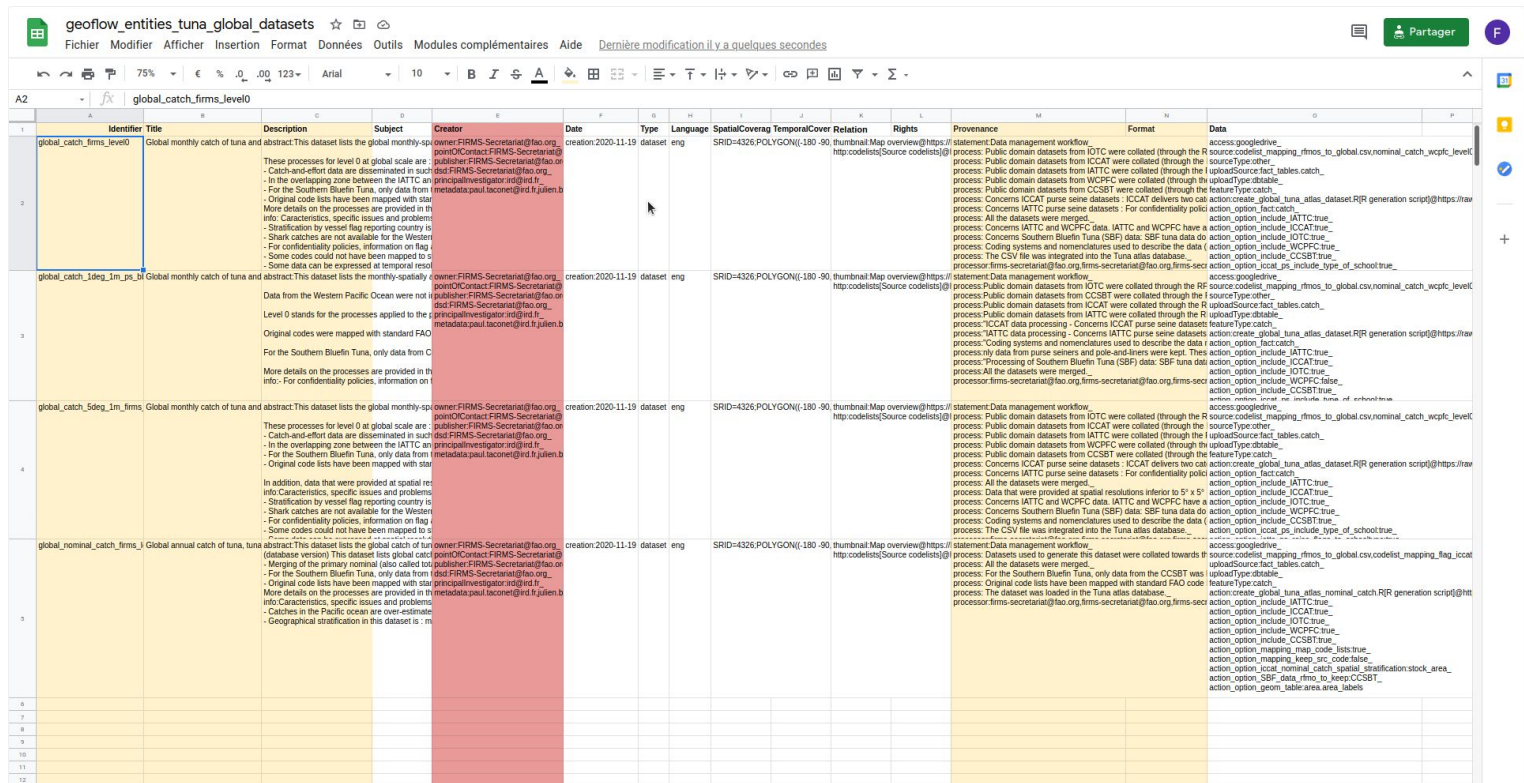
+

Edition du tableau dans un environnement collaboratif (optionnel)



# Les concepts de geoflow : format & outils basiques

## Gestionnaires et utilisateurs collaborant sur un tableau



Identifier	Title	Description	Subject	Creator	Date	Type	Language	SpatialCoverage	TemporalCover	Relation	Rights	Provenance	Format	Data
global_catch_firms_level0	Global monthly catch of tuna and abstract:	This dataset lists the global monthly-sp... These processes for level 0 at global scale are : - Catch-and-effort data are disseminated in such... - In the overlapping zone between the IATTC an... - For the Southern Bluefin Tuna, only data from... - Original code lists have been mapped with star... More details on the processes are provided in th... info: Characteristics, specific issues and problem... - Stratification by vessel flag reporting country is... - Shark catches are not available for the Wester... - For confidentiality policies, information on flag... - Some codes could not have been mapped to s... - Some data can be expressed at temporal res...		owner:FIRMS-Secretariat@fao.org pointOfContact:FIRMS-Secretariat@fao.org publisher:FIRMS-Secretariat@fao.org principalInvestigator:scj@id.f... metadatas.paul.taconet@id.f... juliem.b...	creation:2020-11-19	dataset	eng	SRID=4326:POLYGON((180 -90, ...		thumbnail Map overview https://... http://codelists/Source/codelists/		statement:Data management workflow... process:Public domain datasets from IOTC were collated (through the R... process:Public domain datasets from ICCAT were collated (through the sourceType:other... process:Public domain datasets from IATTC were collated (through the uploadSource:fact_tables.catch... process:Public domain datasets from WCPFC were collated (through the uploadSource:fact_tables.catch... process:Public domain datasets from CCSBT were collated (through the featureType:catch... process:Concerns IATTC purse seine datasets: For confidentiality policy action_option_include_IATTC.true... process:Concerns IATTC purse seine datasets: ICCAT delivers two call action:create_global_tuna_atlas_dataset.RJR generation script https://raw... process:All the datasets were merged... action_option_include_IATTC.true... process:Concerns IATTC and WCPFC data: IATTC and WCPFC have a action_option_include_ICCAT.true... process:Concerns Southern Bluefin Tuna (SBF) data: SBF tuna data do action_option_include_IOTC.true... process:Coding systems and nomenclatures used to describe the data ( action_option_include_WCPFC.true... process:The CSV file was integrated into the Tuna atlas database... action_option_include_CCSBT.true... processor:firms-secretariat@fao.org,firms-sec... action_option_include_type_of_school>true... access:googledrive...		
global_catch_10deg_1m_ps_M	Global monthly catch of tuna and abstract:	This dataset lists the global monthly-sp... Data from the Western Pacific Ocean were not... Level 0 stands for the processes applied to the... Original codes were mapped with standard FAO... For the Southern Bluefin Tuna, only data from C... More details on the processes are provided in t... info: For confidentiality policies, information on		owner:FIRMS-Secretariat@fao.org pointOfContact:FIRMS-Secretariat@fao.org publisher:FIRMS-Secretariat@fao.org principalInvestigator:scj@id.f... metadatas.paul.taconet@id.f... juliem.b...	creation:2020-11-19	dataset	eng	SRID=4326:POLYGON((180 -90, ...		thumbnail Map overview https://... http://codelists/Source/codelists/		statement:Data management workflow... process:Public domain datasets from IOTC were collated through the R... process:Public domain datasets from ICCAT were collated through the R... process:Public domain datasets from IATTC were collated through the R... process:Public domain datasets from WCPFC were collated through the R... process:Public domain datasets from CCSBT were collated through the R... process:ICCAT data processing - Concerns ICCAT purse seine datasets featureType:catch... process:IATTC data processing - Concerns IATTC purse seine datasets action:create_global_tuna_atlas_dataset.RJR generation script https://raw... process:Coding systems and nomenclatures used to describe the data ( action_option_include_IOTC.true... process:Only data from purse seiners and pole-and-trawlers were kept. These action_option_include_IATTC.true... process:Processing of Southern Bluefin Tuna (SBF) data: SBF tuna data do action_option_include_ICCAT.true... process:All the datasets were merged... action_option_include_WCPFC.false... processor:firms-secretariat@fao.org,firms-sec... action_option_include_CCSBT.true... action_option_include_type_of_school>true... access:googledrive...		
global_catch_5deg_1m_firms	Global monthly catch of tuna and abstract:	This dataset lists the global monthly-sp... These processes for level 0 at global scale are : - Catch-and-effort data are disseminated in such... - In the overlapping zone between the IATTC an... - For the Southern Bluefin Tuna, only data from... - Original code lists have been mapped with star... In addition, data that were provided at spatial r... info:Characteristics, specific issues and problems... - Stratification by vessel flag reporting country is... - Shark catches are not available for the Wester... - For confidentiality policies, information on flag... - Some codes could not have been mapped to s... - Some data could not have been mapped to s...		owner:FIRMS-Secretariat@fao.org pointOfContact:FIRMS-Secretariat@fao.org publisher:FIRMS-Secretariat@fao.org principalInvestigator:scj@id.f... metadatas.paul.taconet@id.f... juliem.b...	creation:2020-11-19	dataset	eng	SRID=4326:POLYGON((180 -90, ...		thumbnail Map overview https://... http://codelists/Source/codelists/		statement:Data management workflow... process:Public domain datasets from IOTC were collated (through the R... process:Public domain datasets from ICCAT were collated (through the sourceType:other... process:Public domain datasets from IATTC were collated (through the uploadSource:fact_tables.catch... process:Public domain datasets from WCPFC were collated (through the uploadSource:fact_tables.catch... process:Public domain datasets from CCSBT were collated (through the featureType:catch... process:Concerns IATTC purse seine datasets: For confidentiality policy action_option_include_IATTC.true... process:Concerns IATTC and WCPFC data: IATTC and WCPFC have a action_option_include_ICCAT.true... process:Concerns Southern Bluefin Tuna (SBF) data: SBF tuna data do action_option_include_IOTC.true... process:Concerns IATTC purse seine datasets: ICCAT delivers two call action:create_global_tuna_atlas_dataset.RJR generation script https://raw... process:All the datasets were merged... action_option_include_IATTC.true... process:Data that were provided at spatial resolutions inferior to 5° x... action_option_include_ICCAT.true... process:Concerns IATTC and WCPFC data: IATTC and WCPFC have a action_option_include_IOTC.true... process:Concerns Southern Bluefin Tuna (SBF) data: SBF tuna data do action_option_include_IOTC.true... process:Coding systems and nomenclatures used to describe the data ( action_option_include_CCSBT.true... process:The CSV file was integrated into the Tuna atlas database... action_option_include_type_of_school>true... processor:firms-secretariat@fao.org,firms-sec... access:googledrive...		
global_nominal_catch_firms_j	Global annual catch of tuna, tuna and abstract:	This dataset lists the global catch of tuna... - Merging of the primary nominal (also called to... - For the Southern Bluefin Tuna, only data from... - Original code lists have been mapped with star... More details on the processes are provided in th... info: Characteristics, specific issues and problem... - Catches in the Pacific ocean are over-estim... - Geographical stratification in the dataset is: m...		owner:FIRMS-Secretariat@fao.org pointOfContact:FIRMS-Secretariat@fao.org publisher:FIRMS-Secretariat@fao.org principalInvestigator:scj@id.f... metadatas.paul.taconet@id.f... juliem.b...	creation:2020-11-19	dataset	eng	SRID=4326:POLYGON((180 -90, ...		thumbnail Map overview https://... http://codelists/Source/codelists/		statement:Data management workflow... process:Public domain datasets from IOTC were collated towards th... process:Public domain datasets from ICCAT were collated towards th... process:All the datasets were merged... process:For the Southern Bluefin Tuna, only data from the CCSBT was... process:Original code lists have been mapped with standard FAO code... process:The dataset was loaded in the Tuna atlas database... processor:firms-secretariat@fao.org,firms-sec... action_option_include_IATTC.true... action_option_include_ICCAT.true... action_option_include_IOTC.true... action_option_include_WCPFC.true... action_option_include_CCSBT.true... action_option_include_type_of_school>true... action_option_include_ICCAT.true... action_option_include_IOTC.true... action_option_include_WCPFC.true... action_option_include_CCSBT.true... action_option_mapping_map_code_lists>true... action_option_mapping_keep_src_code>false... action_option_nominal_catch_spatial_stratification_stock_area... action_option_SBF_data_remove_to_keep_CCSBT... action_option_geom_table_area_area_labels		



# Les concepts de geoflow : format & outils basiques

La construction du tableau répond à différents cas d'utilisation:

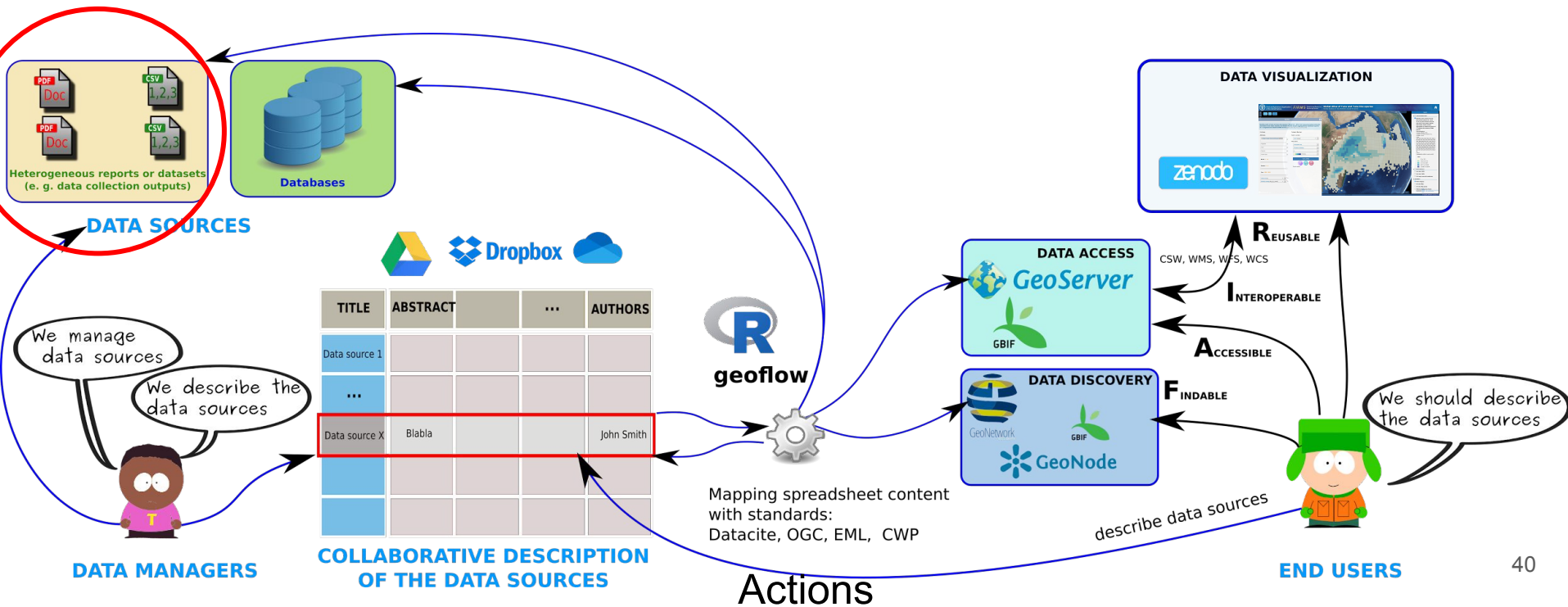
- Une ligne => une entité (métadonnée + donnée)
- Les lignes (entités) peuvent être
  - totalement indépendantes les unes des autres,
  - Liées entre elles, ex:
    - tableau = **catalogue de requêtes d'une même base de données** (même source),
    - chaque ligne représente un nouvel état de la ligne précédente, ex:
      - Ligne 1: un dataset csv avec colonne lat / lon transformé (output d'une action locale) en geopackage (~shapefile),
      - Ligne 2: le dataset geopackage chargé dans une table d'une base de données
      - Ligne 3: le dataset requête SQL query transformé (output d'une action locale) en DwC,
  - un mélange d'entités liées ou indépendantes...

one row = description of one dataset  
one column = one metadata element

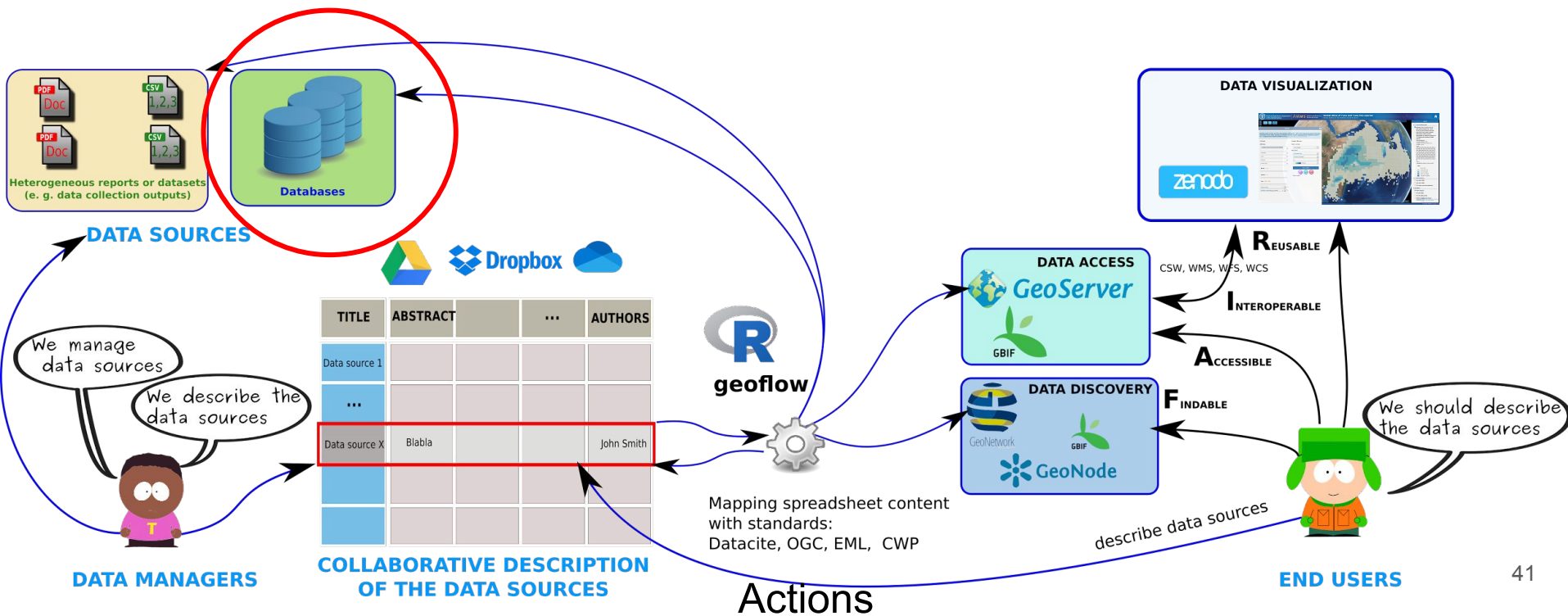
TITLE	ABSTRACT	...	AUTHORS
Data source 1			
...			
Data source 1	Baidu		John Smith



# Les concepts de geoflow : format & outils basiques



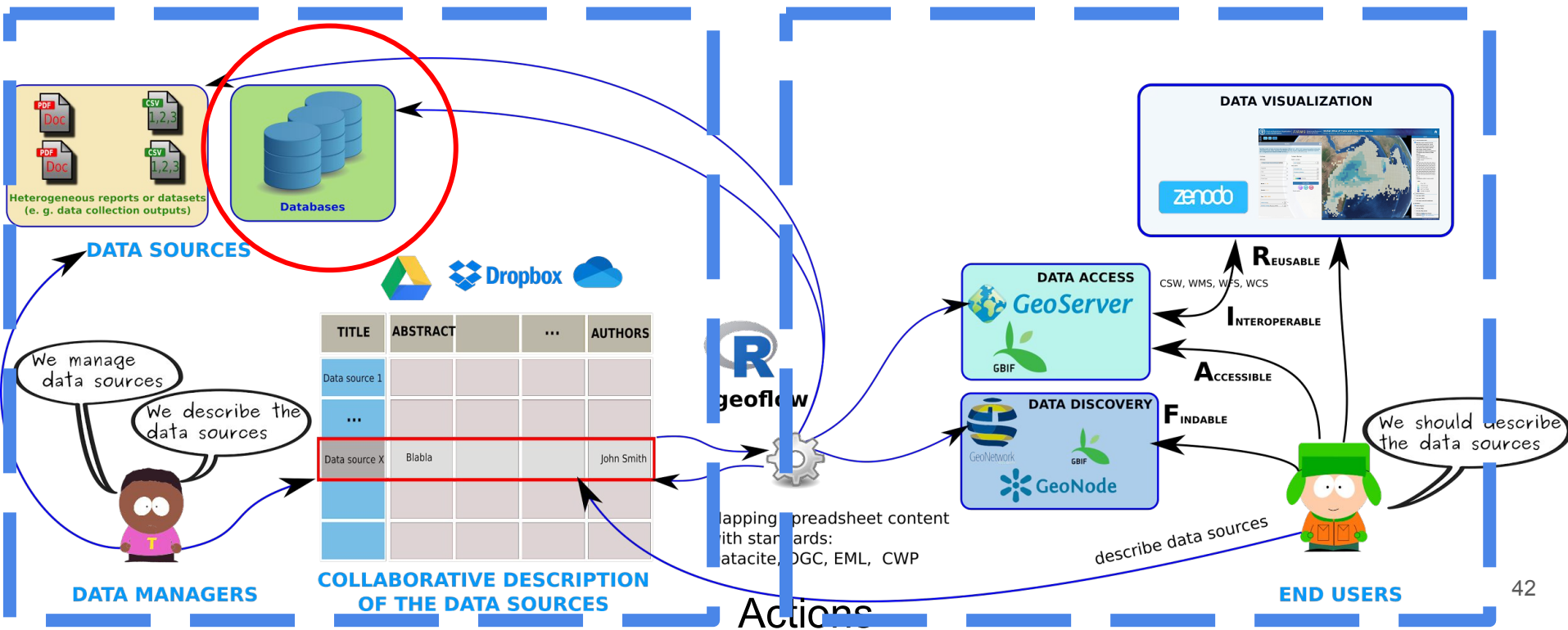
# Les concepts de geoflow : format & outils basiques



# Les concepts de geoflow : format & outils basiques

## Workflow ETL / Step 1

## Workflow FAIR / Step 2



# Les concepts de geoflow : exécution du workflow

En pratique, une commande suffit :

```
geoflow::executeWorkflow("config.json")
```

Le fichier ["config.json"](#) centralise toutes les informations nécessaires à l'exécution.

Le partage de ce fichier avec un collègue lui permet d'exécuter le workflow (~).

# Les concepts de geoflow : exécution du workflow

Un workflow peut ainsi être décomposé en plusieurs étapes:

```
geoflow::executeWorkflow("config_step1.json")
```

```
geoflow::executeWorkflow("config_step2.json")
```

```
geoflow::executeWorkflow("config_step3.json")
```

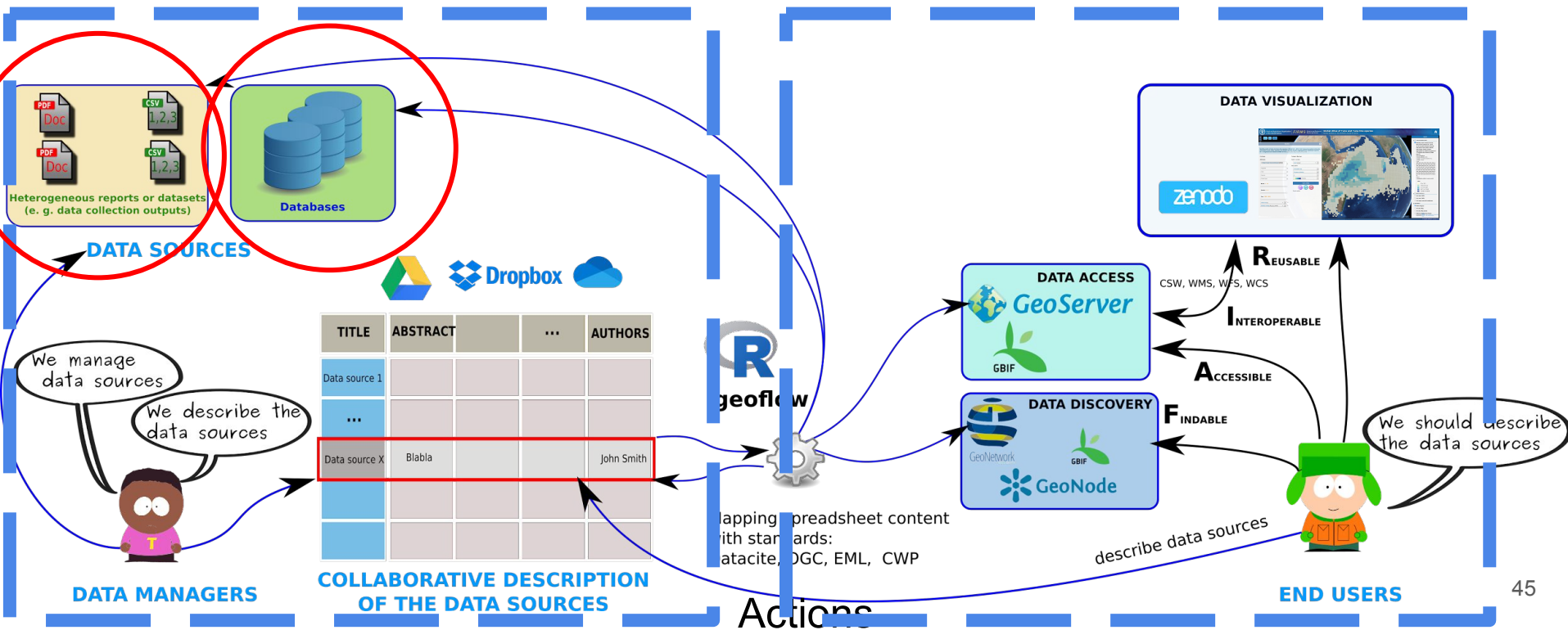
```
geoflow::executeWorkflow("config_final_step.json")
```

Le contenu des fichiers ".json" diffère pour organiser la séquence...

# Les concepts de geoflow : format & outils basiques

## Workflow ETL / Step 1

## Workflow FAIR / Step 2



# Les concepts de geoflow : exécution du workflow

Un workflow décomposé en trois étapes (ETL + FAIR):

```
geoflow::executeWorkflow("transform_files.json")
```

```
geoflow::executeWorkflow("load_database.json")
```

```
geoflow::executeWorkflow("FAIR_services.json")
```

# Les concepts de geoflow : fichier json de configuration

"config.json" liste et décrit les objets du workflow:

```
{  
  
  "id": "my-workflow",  
  
  "mode": "entity",  
  
  "metadata": { <metadata sources defined here> },  
  
  "software": [ <pieces of software defined here> ],  
  
  "actions": [ <actions defined here> ],  
  
  "profile": { <global profile (metadata) defined here> },  
  
  "options": { <global options defined here> },  
  
}
```

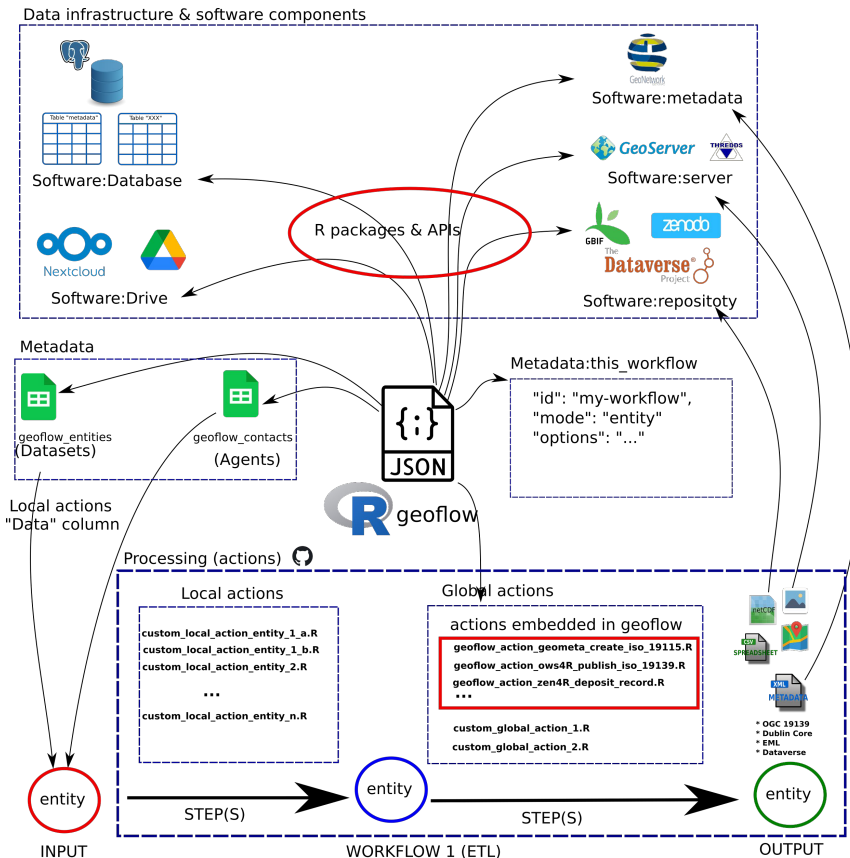


# Workflow : le fichier json de configuration

```

{
  "id": "my-workflow",
  "mode": "entity",
  "profile": { <global profile (metadata) defined here> },
  "options": { <global options defined here> },
  "metadata": { <metadata sources defined here> },
  "software": [ <pieces of software defined here> ],
  "actions": [ <actions defined here> ]
}

```



# Workflow : le fichier json avec Shiny

Geoflow UI

Configuration

Configuration editor

The geoflow configuration editor allows users to create a geoflow data flow configuration file in an interactive user-friendly manner. The user will be able to load an existing configuration file. Once the configuration file created/edited, the user will be able to execute it workflow interactively.

Load configuration file?

Choose **Json File**

Browse... No file selected

Load

Profile Metadata Software Actions

Workflow Identifier

Workflow mode

entity

Project name

Organization

Logos

Geoflow UI

Configuration

Configuration editor

The geoflow configuration editor allows users to create a geoflow data flow configuration file in an interactive user-friendly manner. The user will be able to load an existing configuration file. Once the configuration file created/edited, the user will be able to execute it workflow interactively.

Load configuration file?

Choose **Json File**

Browse... config.json

Upload complete Valid JSON

Load

Profile Metadata Software Actions

Contacts Entities

Add a new contact source Modify a new contact source Delete a contact source

Handler	Source
gsheet	<a href="https://docs.google.com/spreadsheets/d/144NmGskldiRE57BIN0mck9uZEUH2dBUzCcy1pJ56nAg/edit#gid=0">https://docs.google.com/spreadsheets/d/144NmGskldiRE57BIN0mck9uZEUH2dBUzCcy1pJ56nAg/edit#gid=0</a>
gsheet	<a href="https://docs.google.com/spreadsheets/d/1qGUHFnH3d-DucDphL2MSCE6pXYSaSNJu0vJBK4T2iw/edit?usp=sharing">https://docs.google.com/spreadsheets/d/1qGUHFnH3d-DucDphL2MSCE6pXYSaSNJu0vJBK4T2iw/edit?usp=sharing</a>

Showing 1 to 2 of 2 entries

Geoflow UI

Configuration

Configuration editor

The geoflow configuration editor allows users to create a geoflow data flow configuration file in an interactive user-friendly manner. The user will be able to load an existing configuration file. Once the configuration file created/edited, the user will be able to execute it workflow interactively.

Load configuration file?

Choose **Json File**

Browse... config.json

Upload complete Valid JSON

Load

Profile Metadata Software Actions

Add a new action Modify an action Delete an action

Identifier	Run?	Action Type	Definition
zen4R-deposit-record	true	Data publication	Deposits/Publish data and/or metadata in the Zenodo infrastructure
geometa-create-iso-19115	true	Metadata production	Produce an ISO/OGC 19115/19139 metadata object

Options

- logo: true
- doi: false
- doi\_thumbnail: false
- addfeatures: false

# Les concepts de geoflow : fichier json de configuration

"config.json" liste et décrit les objets du workflow:

- **Chargement du ".json"** : `config <- geoflow::initWorkflow("this_file.json")`
- **Métadonnées:**
  - Info générales sur le **workflow** : `config$id + config$mode + config$profile`
  - **Données / entités** à traiter par le workflow (lien vers le tableau): `config$metadata`
    - liste des objets (ex: datasets /DOIs): `config$getEntities()`
    - dictionnaire des données: `config$getDictionary()`
  - **Contacts**: personnes (ex: ORCID) liées aux datasets: `config$getContacts()`
- **Logiciels** : `config$software` VS `geoflow::list_software()`
- **Actions (Traitements):**
  - **Globales** :
    - natives: `config$actions` in `geoflow::list_actions()`
    - Custom / spécifique: `config$actions` (not in `geoflow::list_actions()`)
  - **Locales** : `config$getEntities()[[1]]$data$actions[[1]]`

# Les concepts de geoflow : actions globales & locales

**Action** = traitement appliqué à une **entité**:

- **Action = source** du script (local, github...) + **paramètres** pour cette entité
- **Deux types d'actions:**
  - **Actions globales:**
    - s'appliquent à toutes les entités,
    - directement déclarées dans le json
  - **Actions locales:**
    - s'appliquent à une seule entité
    - forcément déclarées dans la colonne "Data"

# Les concepts de geoflow : actions globales (json)

Les **actions globales** s'appliquent à tous les jeux de données:

- **actions globales natives:** existent par défaut dans geoflow pour gérer des actions récurrentes dans tous les workflow. `geoflow::list_actions()`:
  - "geometa-create-iso-19115"
  - "geometa-create-iso-19110"
  - "geosapi-publish-ogc-services"
  - "ows4R-publish-iso-19139"
  - ...
- **actions globales spécifiques:** intérêt limité à un / qqs workflow(s)

# Les concepts de geoflow : actions locales (pas json)

Les actions locales sont listées dans la colonne Data:

- Plusieurs actions locales sont possibles pour une même entité:
  - Ex: première action locale appliquée à la première entité:
    - `config$getEntities()[[1]]$data$actions[[1]]`
- **Exemples d'actions locales :**
  - Changer le référentiel espèce d'un jeu de données..
  - Spatialiser un jeu de donnée
  - Charger un jeu de données dans une base de données
- La donnée de l'entité peut être produite à la volée / dynamiquement:
  - la donnée de l'entité décrite sera le résultat (output) d'une action locale (scripts R, SQL..)
  - si pas d'actions locales pour une entité alors la donnée existe déjà (donnée statique: csv,..)

# Les concepts de geoflow : les options

Différents types d'options:

- Options du **workflow** (clé "options" du json): `config$options`
  - `"line_separator": ";\n"`,
  - `"skipFileDownload": true`
  - `"skipDynamicBbox": true`
- Options / paramètres des **actions globales**:
  - `geoflow::list_action_options()`
  - ex: `geoflow::list_action_options(id="geometa-create-iso-19115")`
- Options / paramètres des **actions locales**:
  - `config$getEntities()[[1]]$data$actions[[1]]$options`

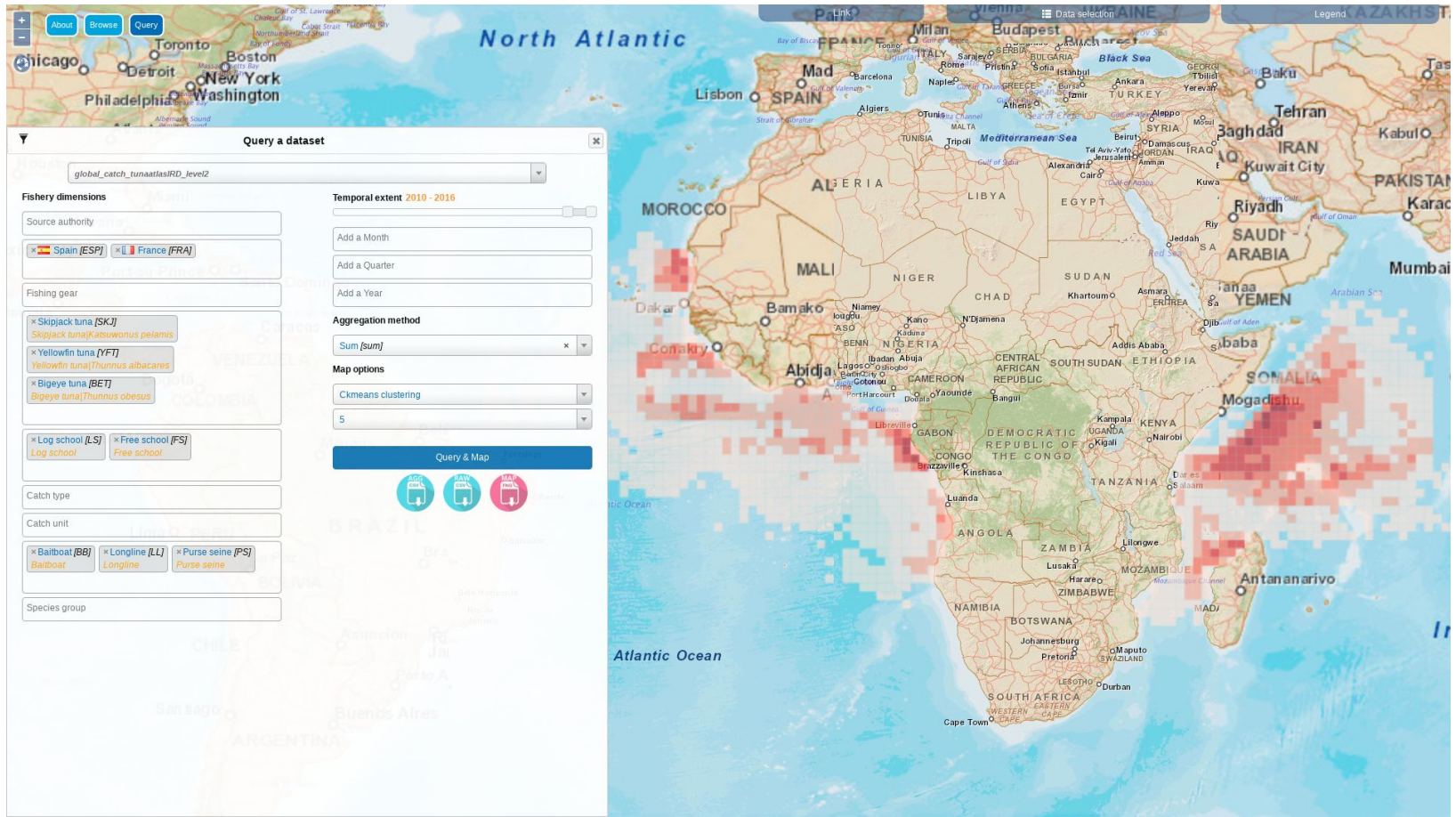
# geoflow : exemple de FAIR Workflow

## Implémentation de FAIR DMP: un cas générique et fédérateur

- problématique partagée par les gestionnaires de données :
  - Publier les mêmes métadonnées et données dans différents catalogues et serveurs:
    - Geonetwork / Geoserver ET Zenodo ou Dataverse
  - Volonté de standardisation: interopérabilité (DCMI / Datacite / OGC)
- Exemples en cours:
  - tableau (modèle pivot) dans un Drive (gsheet) utilisée pour publier des jeux de données dans GN/GS/Zenodo/OFV
    - Données de projets hétérogènes: listées dans un tableau et publiées sur Geonetwork / Geoserver / Dataverse / Zenodo (ex: COI, CTOI..)
  - une table “metadata” (modèle pivot) dans une base SQL (Postgis) est utilisée pour publier un jeu de requêtes dans GN/GS/Zenodo/OFV:
    - [Tuna fisheries global atlas](#) (FAO / IRD):
    - Base de données séries d’images sous-marines ou aériennes (IRD / Ifremer)
  - Dynafor (INRAE..)



# geoflow : exemple de FAIR Workflow (Tuna atlas)



# geoflow : exemple de FAIR Workflow ([Seatizen](#))

The screenshot displays the Seatizen web application interface. The central map shows a coastal area with green flow lines representing movement paths. The interface includes several panels:

- Access Panel (Left):** Contains the title "session of the 2017 11 04 kite Le Morne" and various filters and options for data visualization.
- Attributes Panel (Left):** Lists attributes such as "GPSTime" and "DateTimeOriginal".
- Map options Panel (Left):** Includes "Choropleth map", "Okmeans clustering", and "5" (likely a zoom level or scale).
- Export options Panel (Left):** Provides icons for exporting data.
- Data Popup (Center):** Displays detailed information for a selected point, including coordinates, session ID, and date/time.
- Legend Panel (Right):** Shows the current map layers, including "session of the 2017 11 04 kite Le Morne" and "Geographic Area of Competence of North Pacific Fisheries Commission (NPFC)".
- Base Overlays Panel (Right):** Includes options for "Exclusive Economic Zones" and "FAO major areas & breakdown".
- Base Maps Panel (Right):** Lists available map styles such as "World Imagery", "OpenStreetMaps", "UN Clear Map", "UN Clear Map (Dark)", and "EMCOmet Bathymetry World baselayer".

The bottom status bar shows the coordinates "Lat: -20.47 Lon: 97.31" and the application version "OpenStreetMaps (version 2.2.1)".

# Workflow : exemple de cas spécifique (ETL)

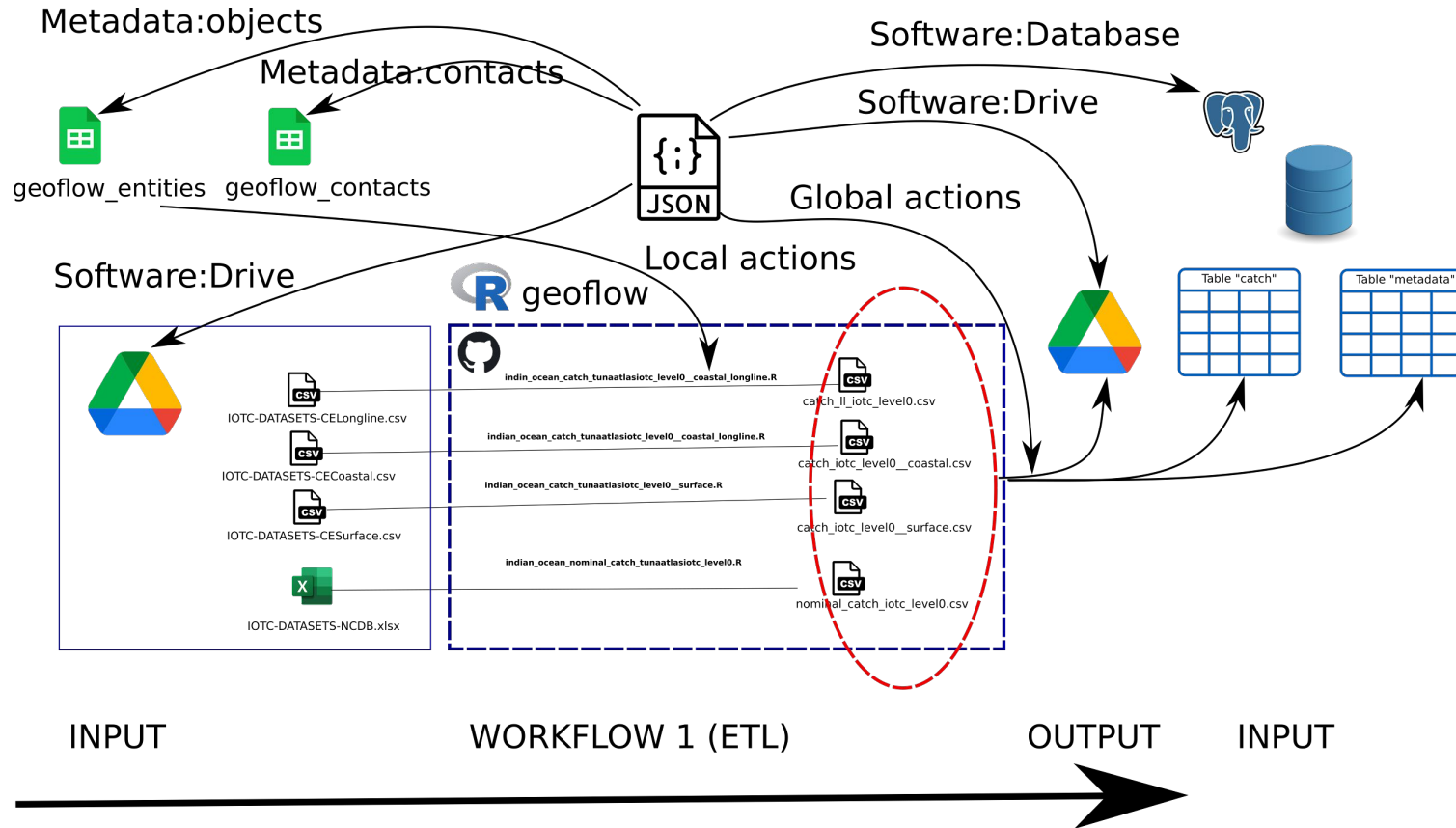
Chargement d'une base de données à partir de multiples fichiers :

- Tuna atlas (en cours): base postgres chargée à partir de CSVs hétérogènes.

Cas complexe:

- Transformation de formats / structure de données,
- Conversion d'unités,
- Changement de référentiels / Codelists,
- Création de métadonnées riches avec historique des traitements dans "Provenance"
- Chargement dans la base
- Base de métadonnées images (à venir):
  - Extraction de métadonnées exif,
  - Ecriture de métadonnées (modèle pivot)
  - Chargement des métadonnées et données dans la base
  - ...puis le workflow pour les services FAIR

# Workflow : exemple de cas spécifique (ETL)



# Geoflow : proposition d'atelier pour fin 2021

Sur le même principe qu'en 2018 (Agropolis, Montpellier), avec 3 journées fin 2021 (Toulouse) pour 3 profils d'utilisateurs / niveaux d'implication :

- **Jour 1 : cas classiques (services FAIR)**
  - Prises en main d'exemples de codes geoflow fonctionnels,
  - Public: bases en R + gestionnaires de données
- **Jour 2 : cas avancés type ETL (charger et FAIRiser une base de données)**
  - Adapter les codes de geoflow à ses outils (bases et actions spécifiques),
  - Public: développeurs (R basique) & gestionnaires de données
- **Jour 3 : développeurs (debugger & contribuer) (et parler de la diapo 50)**
  - Debugger et contribuer aux codes de geoflow avec de nouvelles actions,
  - Public: développeurs (R avancé) & gestionnaires de données

# Geoflow: évolution & Perspectives

Quelques idées en cours de discussion:

- Gestion de formats standardisés en entrée:
  - Pêche: CWP
  - DwC => intégration par défaut du [dictionnaire de données](#)
- Gestion raster (métadonnées, publication geoserver)
- Interaction avec de nouveaux software:
  - Annuaire de contacts existants: LDAP, ORCID ?
  - Entrepôts de données Nextcloud, Owncloud?
- Désactivation locale des actions globales

# Geoflow: évolution & Perspectives

Les projets (et financements) en cours:

- BlueCloud: Tuna Atlas
- INTERREG La Réunion (vecteur & raster)
- Seatizen (données raster)
- Projet BED
- Autres à ajouter..

# Geoflow: perspectives de workflow “faciles” à réaliser

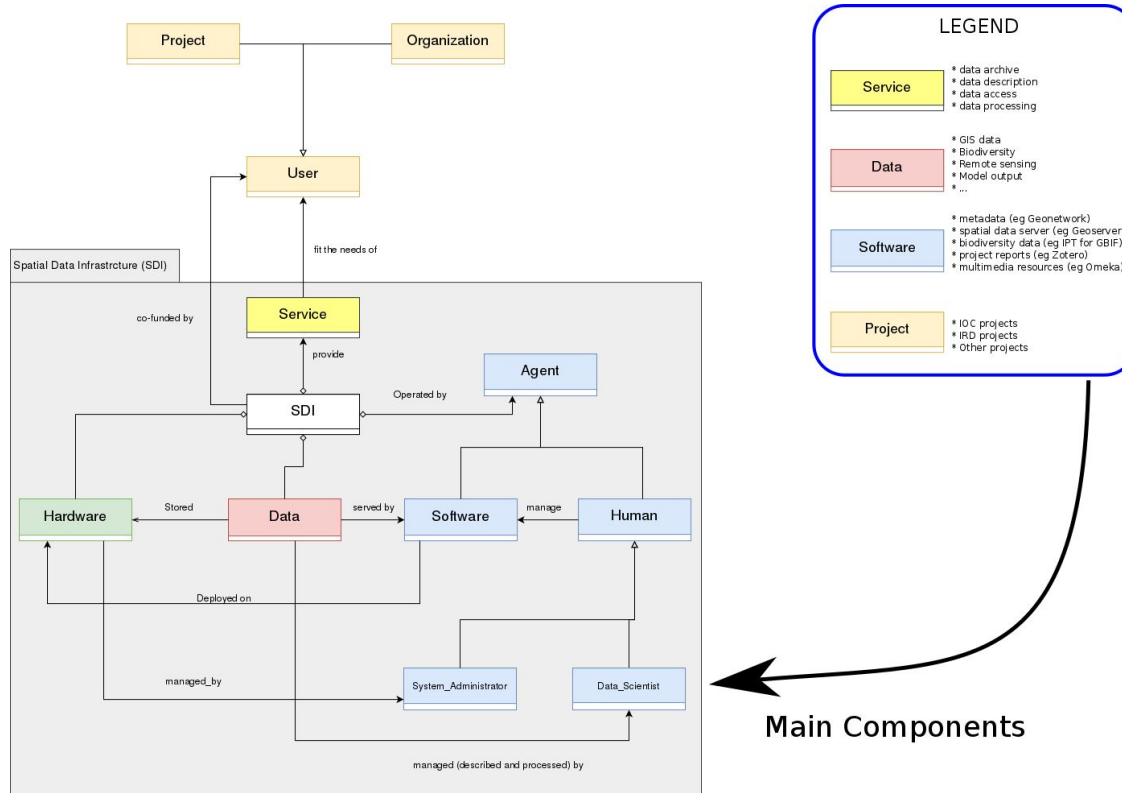
Des exemples de workflows déjà discutés (codes à migrer ou à créer):

- FAIR DMP OGC (CSW/WMS/WFS) + DOIs
- CSW/19139 to EML/Dataverse
- IPT/EML+DwC to GN+GS+OFV
- Thredds/NetCDF/NCML to GN
- DOIs to GN..
- MD to Rmd

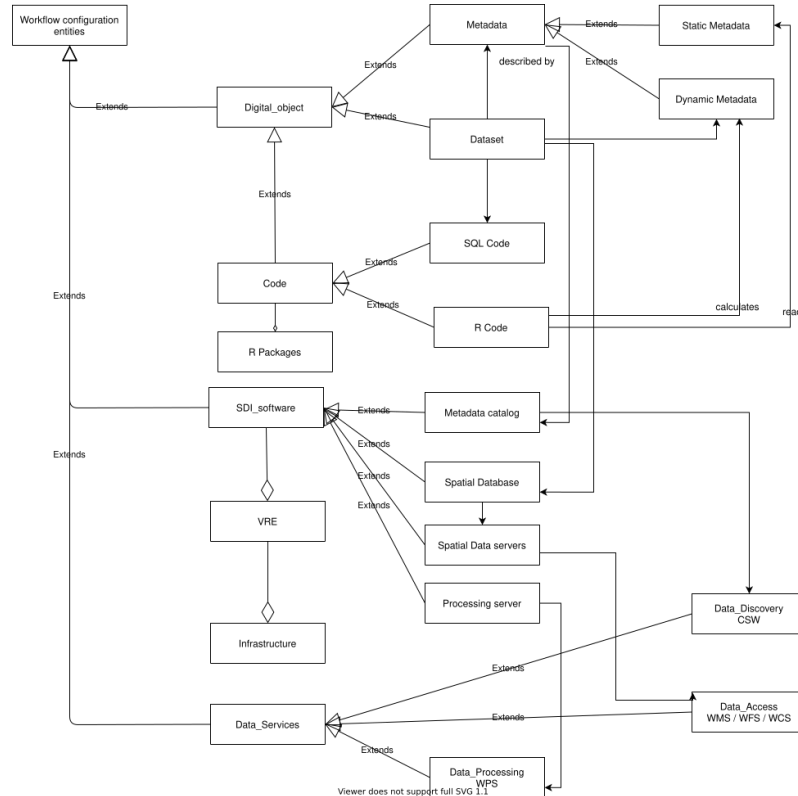


**Diapos en plus...**

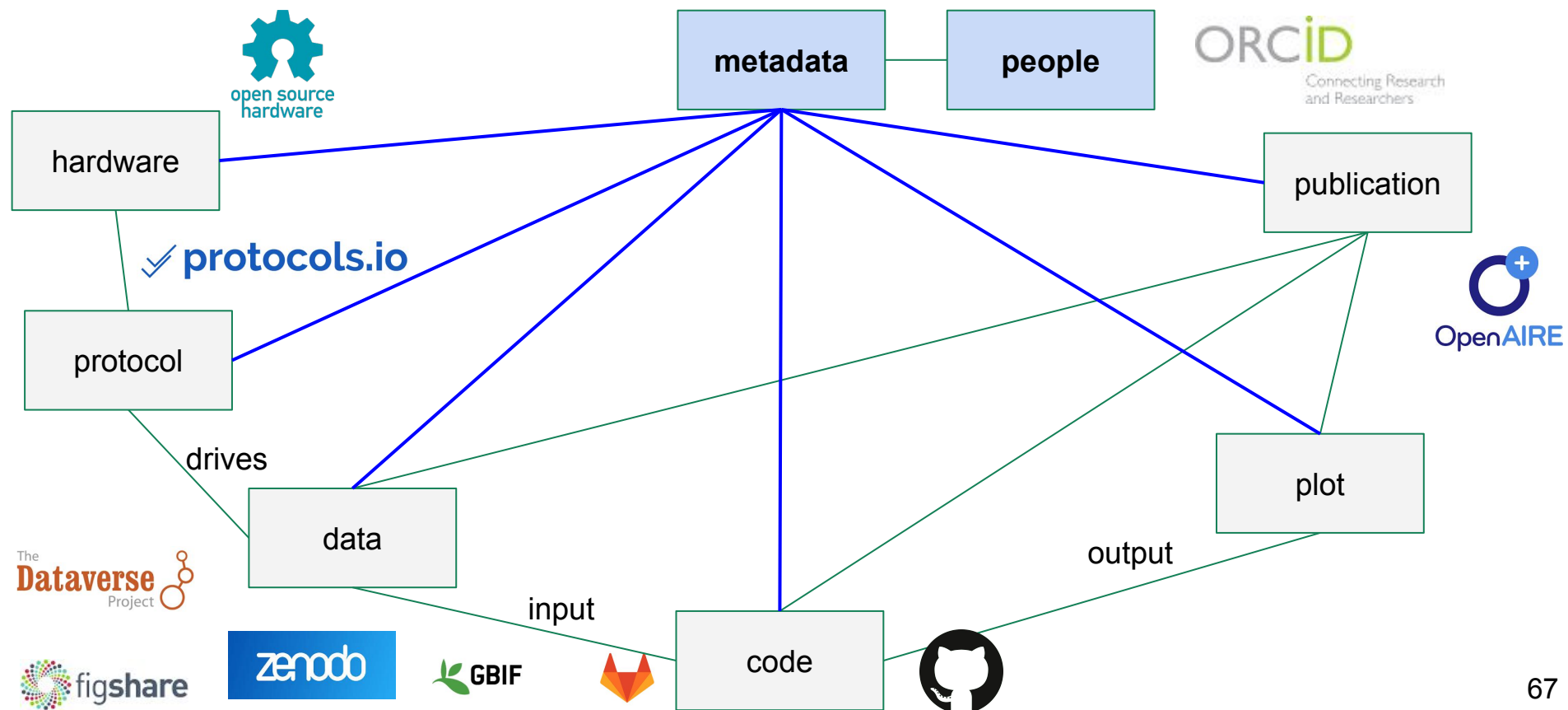
# Spatial Data Infrastructure



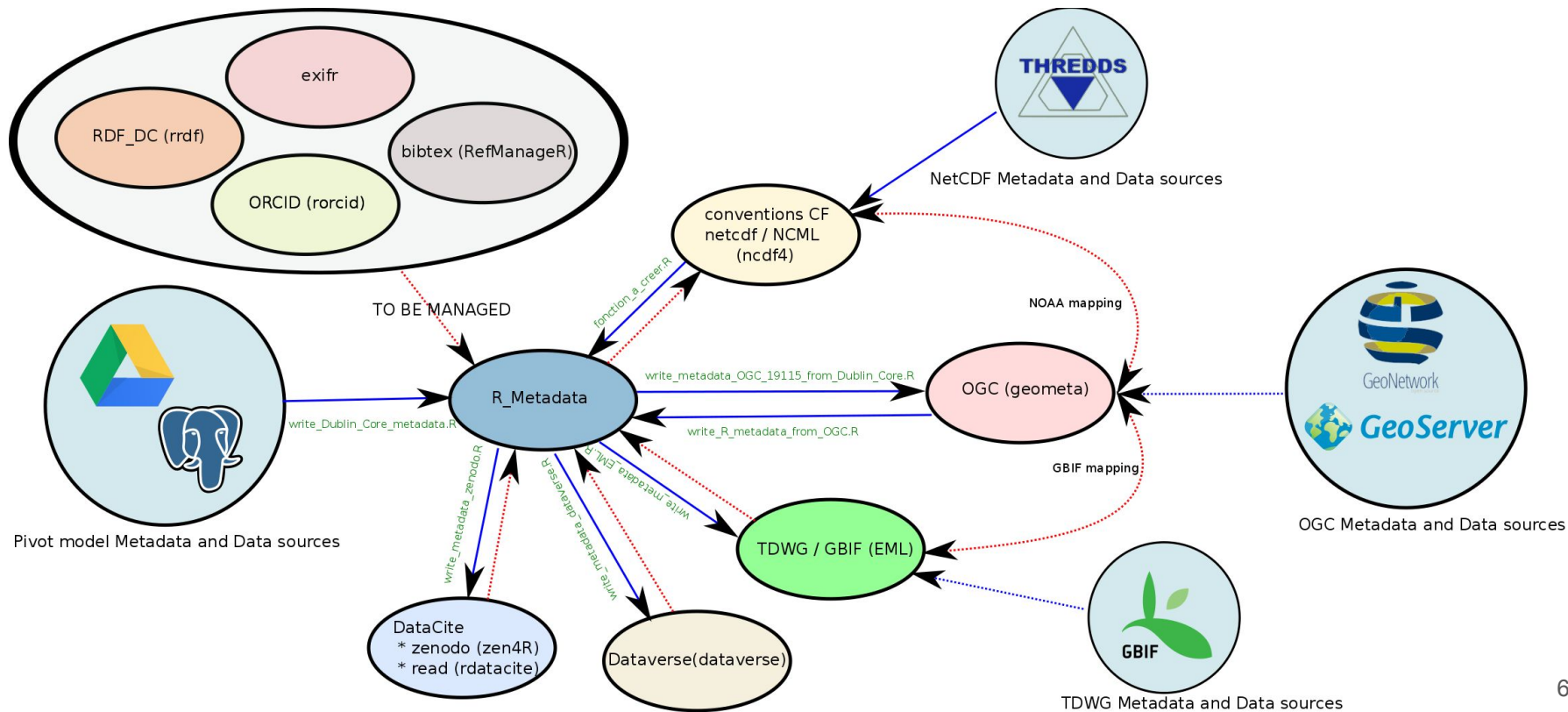
# Workflow (UML diagram)



# Contexte général: Open Science & FAIR



# Ongoing work: overview (see [related google sheet](#))



# Packages R: Mapping de métadonnées

