

# Un workflow accessible pour une gestion simple et durable des données



Julien Barde, Emmanuel Blondel, Wilfried Heintz

<https://tinyurl.com/ye95rloh>  
Journée data2019opensci

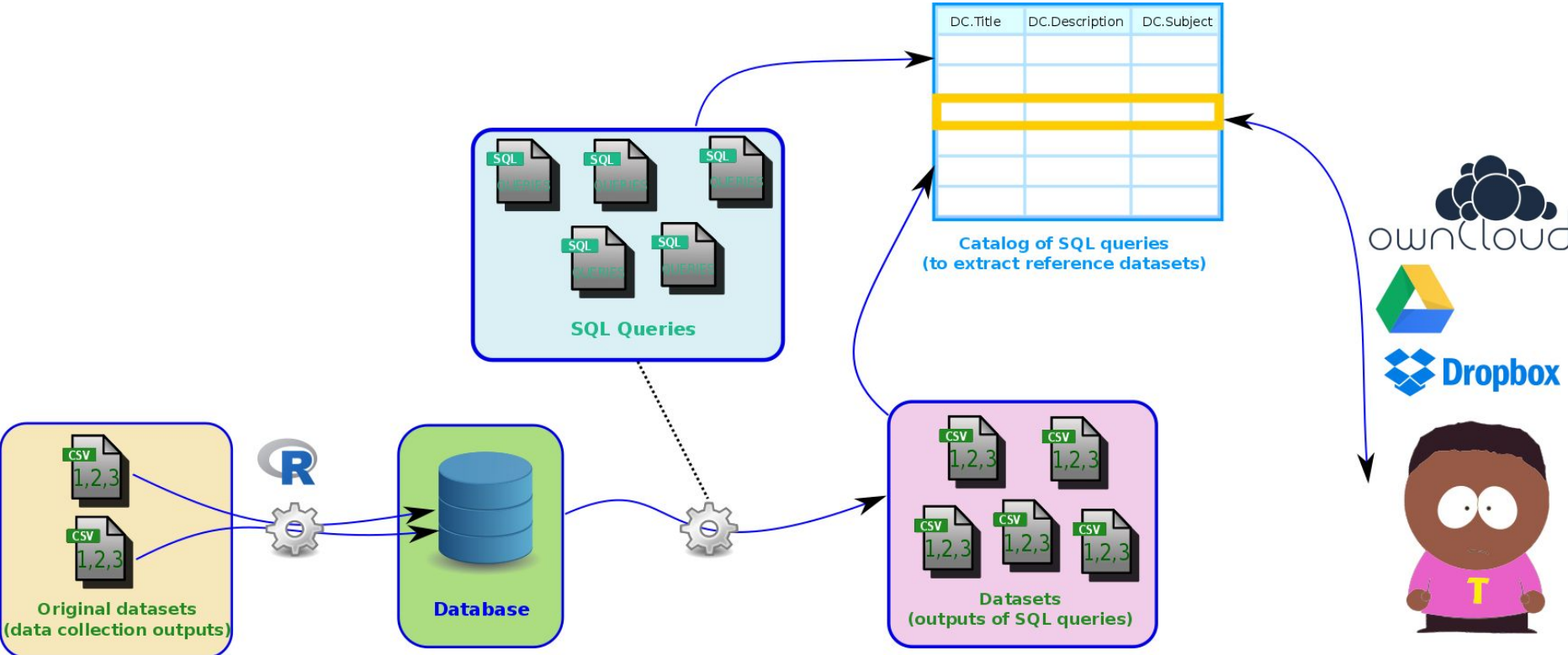
Article dans le numéro spécial du cahier des techniques 2019



# Plan de la présentation

- Données: sources et jeux de données (granularité)
- Gestion de données: points de blocages
- Workflow pour la donnée (spatiale)
  - Objectifs
  - Recommandations / Bonnes pratiques
  - Exemples
  - Vue d'ensemble
- Matériel et méthodes
- [geoflow](#):
  - bonnes pratiques
  - matériel et méthodes
  - modèle pivot de métadonnées
  - exécution et fichier de configuration config.json
  - mapping de métadonnées
  - FAIR workflow

# Données: sources et jeux de données (granularité)



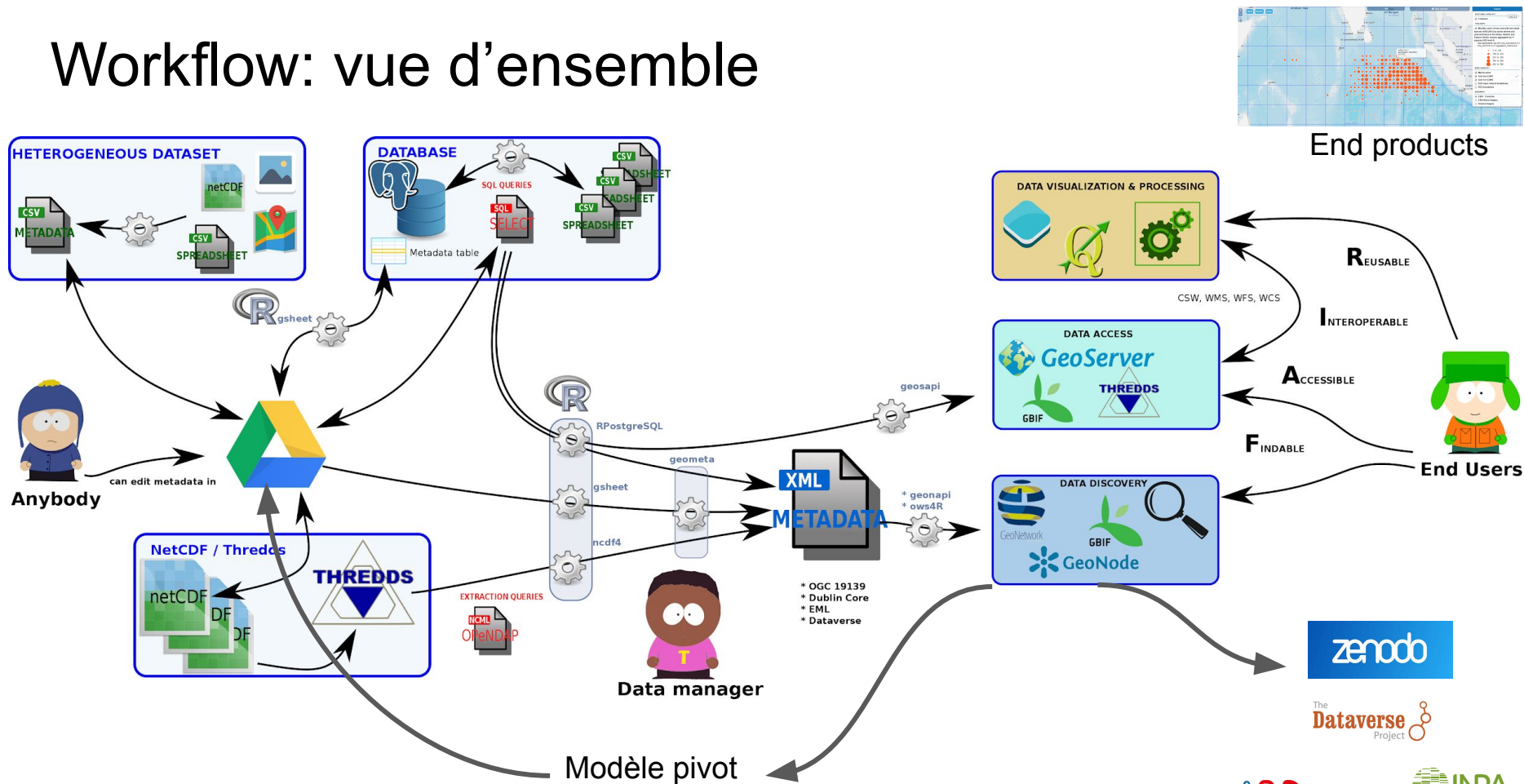
# Gestion de données: points de blocages



Simplifier la gestion des données:

- Absence / perte de métadonnées,
- **Edition**: écriture / saisie de métadonnées:
  - Création et mises à jour de métadonnées (en impliquant les producteurs et les développeurs),
  - **Mapping** de métadonnées: publication “simultanée” dans différents catalogues pour éviter les saisies multiples: DCMI, OGC, EML, datacite ([SWORD](#) API pour Dataverse, Zenodo API..),
- Cibler en priorité les **métadonnées principales** (normes trop complexes),
- **Shunter les interfaces graphiques** complexes: Geonetwork, Geoserver...avec une prise en main qui fait fuir les utilisateurs,
- **Automatiser** des éléments de métadonnées en lisant la donnée (si disponible),
- Validation (ex: INSPIRE)
- Mise en oeuvre de **Plans de Gestion de Données**: comment faire ?

# Workflow: vue d'ensemble

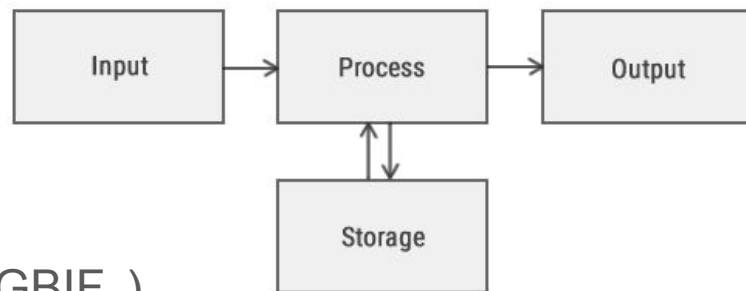


# Workflow: objectifs de notre chaîne de traitements

**Automatisation** (ici en R) de tâches chronophages et récurrentes:

- **Traitements, exemples d'actions:**

- Création & Publication de **métadonnées** (OGC..),
- **Mapping** entre standards de métadonnées,
- Publication de (**flux de**) **données** (eg WMS, WFS..),
- Dépôts de (méta-)données avec attribution de **DOIs**



- **Flux** : du labo aux infrastructures..(INSPIRE, GBIF..)

- Exécution en ligne dans un **VRE** (D4Science / EOSC) composé de PostGIS, Geonetwork, Geoserver, RStudio, Jupyter Notebook...

- **Orchestration** des traitements (“actions”) avec différents packages (ex: R):

- OGC: geometa, ows4R et associés (geonapi, geosapi, geoserver, Rpostgresql, ncdf4..)
- Zenodo (zen4R), GBIF (eml, rgbif), Dataverse (dataverse)
- En discussion: rorcid, LDAP, gsheet / googledrive, owncloud (ie B2DROP), redland / rrdf, exifr

# Workflow: recommandations

Se baser en amont sur des bonnes pratiques:

- Embarquer la **métadonnée** dans la donnée (si possible => producteurs..)
- Fournir un **dictionnaire de données** & un **catalogue de requêtes**
  - Adopter des structures de données standardisées pour les cas classiques: [Darwin Core](#) / [Measurement and facts](#): occurrences d'espèces et biométrie, O&M, OBOE, Conventions-CF
  - Exemple du NetCDF (ou NCML pour gérer la métadonnée à part ou des fichiers virtuels)
    - conformité avec les conventions CF pour les métadonnées et la structure des données (ce que demande thredds, gdal et donc qgis..),
    - serveur OPeNDAP (Thredds) => multiples protocoles d'accès (si conformité CF),
- **Interopérabilité sémantique**: URI & vocabulaires contrôlés,
- **Répliquabilité & collaboration**: environnements mutualisés (serveurs / VRE) pour exécuter les workflows ou collaborer sur la saisie.

# Workflow: exemples pour la donnée spatiale

- “[FAIRification](#)” des données basée sur les standards OGC:
  - publication de **métadonnées & flux de données** (dans SDI)
    - Définition d’un modèle pivot tabulaire (basé sur DCMI) vers OGC, Datacite, EML, knitr..
    - Modèle pivot implémenté en CSV, Google sheet, SQL / Postgis,
  - Attribution de **DOIs** et réinjection des DOIs dans les autres métadonnées
  - Lecture de métadonnées OGC depuis GN ou CSW ou catalogues Thredds (NetCDF-CF ou NCML) pour charger le modèle pivot..
- **Visualisation et extraction** de données à partir des métadonnées: FAIR viewer (CSW+19115+19110+19119)
- **Publier** des “[research objects](#)” sur Zenodo (ex: OpenAIRE-Connect / CTOI),
- Perspectives, exemples:
  - Lire le contenu de l’IPT / GBIF et publier la métadonnée EML dans GN, la donnée dans GS,
  - Descriptions de lots de photos (missions terrain / drones)

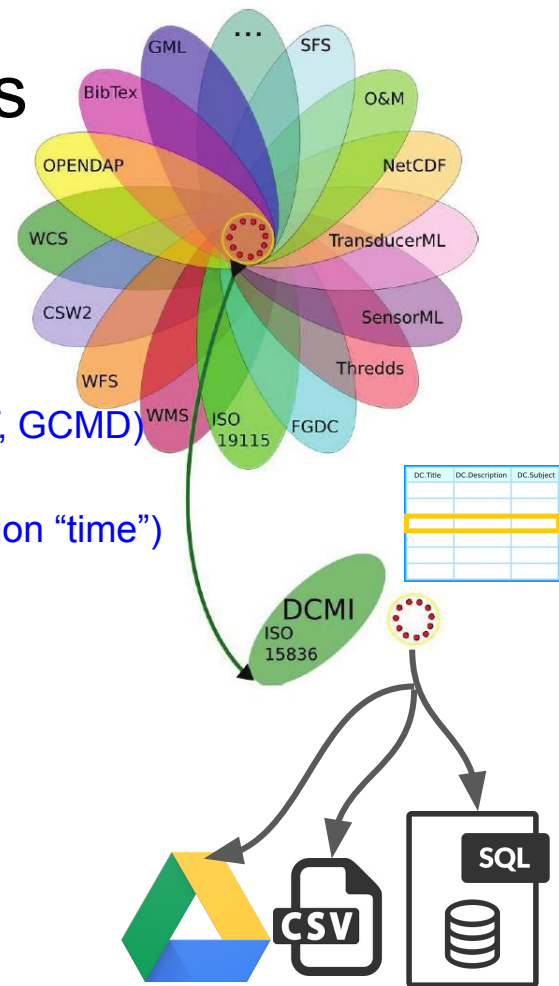


# geoflow: matériel et méthodes

- Modèle pivot de métadonnées basé sur Dublin Core
- Approche mixte en fonction des profils utilisateurs:
  - producteurs de données: GUIs minimalistes et collaboratives pour que les métadonnées soient éditées facilement,
  - gestionnaires de données: exécution des traitements (actions) en R.
- Viser les sources de données classiques:
  - **Semi-structurées**: Shapefiles / CSV..
  - SQL/SFS: ajouter une table “metadata” au modèle,
  - NetCDF (codes à intégrer): conventions CF dans le NetCDF ou en dehors (NCML)
- Privilégier les applications standards:
  - CSW-T en priorité ou APIs natives de geonetwork (rq: problème avec geonode), WFS-T..
- Mise en oeuvre en R: le package geoflow orchestre des actions en coordonnant les actions déléguées à différents packages (dépendances),
- Travail en ligne (vs “Desktop”): VREs de Blue Cloud (EOSC)

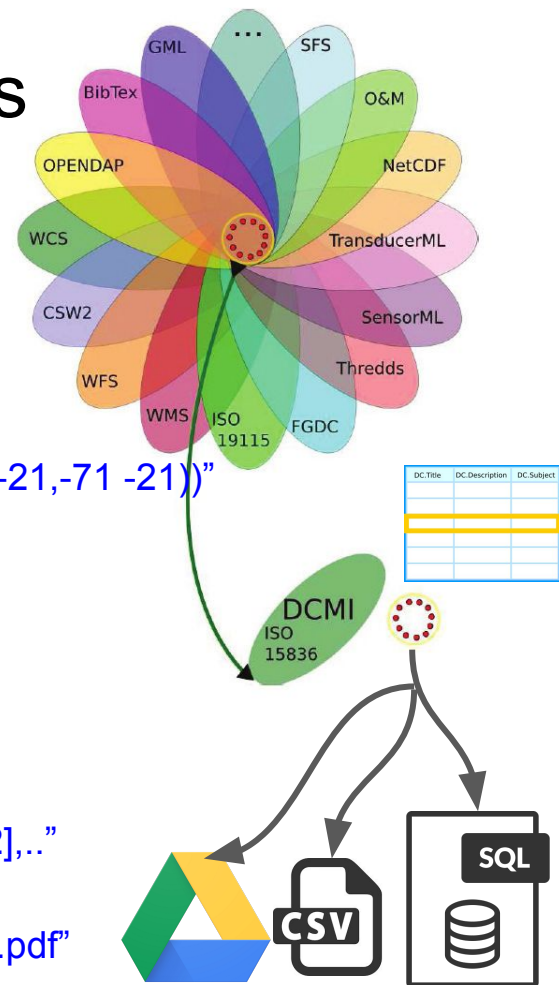
# geoflow: modèle pivot de métadonnées

- **Identifier**: “string\_identifier” + [DOIs](#), [URI](#), [URNs](#)..
- **Title**: “free text”
- **Description**: “prefix:free text” (abstract, purpose, additional info..)
- **Creator**: “role:person” / “role:email” + [UID](#), [ORCID](#) / [FOAF](#)
- **Subject**: “thesaurus:keywords” or [controlled vocabularies](#) (eg [GEMET](#), [GCMD](#))
- **SpatialCoverage**: [eWKT](#) (dynamique si colonne “Data”)
- **TemporalCoverage**: [ISO](#) (dynamique si colonne “Data” avec dimension “time”)
- **Date**: [controlled syntax](#)
- **Type**: “free text”
- **Format**: “free text”
- **Language**: “code langue” (norme [ISO 639](#): fre, eng..)
- **Relation**: “type:relation” or [URLs](#), [URIs](#)...
- **Provenance**: “prefix:free text”
- **Rights**: “prefix:free text” or [Creative Commons](#)....
- **Data**: “prefix:URL” Rules to attach data

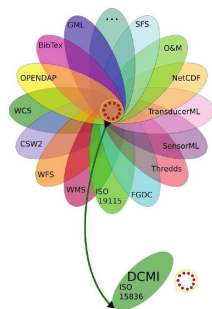


# geoflow: modèle pivot de métadonnées

- **Identifier:** “my-geoflow-record0”
- **Title:** “My Title 0 - metadata only (no data associated)”
- **Description :** “abstract:My metadata description 0\_”
- **Creator:** “metadata:emmanuel.blondel1@gmail.com\_”
- **Subject:** “TH2:kwd1,kwd2,kwd3\_”
- **SpatialCoverage:** “SRID=4326;POLYGON((-71 -21,-71 28,14 28,14 -21,-71 -21))”
- **TemporalCoverage:** “2007/2010”
- **Date:** “2019-02-14”
- **Type:** “dataset”
- **Format:** DOIs
- **Language :** “eng”
- **Relation :** “http:website@http://somelink/website”
- **Provenance:** “process:rationale1[description1],rationale2[description2],..”
- **Rights:** “useConstraint:copyright\_”
- **Data:** “source:cwp\_report.pdf@http://www.fao.org/3/i7805en/I7805EN.pdf”



# geoflow: modèle pivot de métadonnées (CSV..)



geoflow\_entities ☆

File Edit View Insert Format Data Tools Add-ons Help Last edit was made yesterday at 8:19 PM by Emmanuel Blondel

100% Arial 10 B I U

Identifier	A	B	C	D	E	F	G	H	
Identifier	Title	Description	Subject	Creator	Date	Type	Language	SpatialCoverage	
1	my-geoflow-record0	My Title 0 - metadata ont	abstract:My metadata descrip TH1@http://somevocab?thesaurus=th1:kw1@http://somevocab?keyword=kw purpose:Some purpose for this TH2:kw1, kw2, kw3 info:some supplemental info 0 TH3:kw1, keyword 2@http://someret/kw2, keyword 3	metadatas:emmanuel.blondel1@gmail.com_ owner:emmanuel.blondel1@gmail.com_ pointOfContact:julien.barde@ird.fr,wilfried.heintz@inra.fr	2019-02-14	dataset	eng	SRID=4326,POLYGON(	
2									
3	my-geoflow-record1	My Title 1 - with a shapef	abstract:My metadata descrip TH1@http://somevocab?thesaurus=th1:kw1@http://somevocab?keyword=kw purpose:Some purpose for this TH2:kw1, kw2, kw3 info:some supplemental info 1 TH3:kw1, keyword 2@http://someret/kw2, keyword 3	metadatas:emmanuel.blondel1@gmail.com_ owner:emmanuel.blondel1@gmail.com_ pointOfContact:julien.barde@ird.fr,wilfried.heintz@inra.fr	2019-02-14	dataset	eng	SRID=4326,POLYGON(	
4	my-geoflow-record2	My Title 2 - with a shapef	abstract:My metadata descrip TH1@http://somevocab?thesaurus=th1:kw1@http://somevocab?keyword=kw purpose:Some purpose for this TH2:kw1, kw2, kw3 info:some supplemental info 2 TH3:kw1, keyword 2@http://someret/kw2, keyword 3	metadatas:emmanuel.blondel1@gmail.com_ owner:emmanuel.blondel1@gmail.com_ pointOfContact:julien.barde@ird.fr,wilfried.heintz@inra.fr	2019-02-14	dataset	eng	SRID=4326,POLYGON(	
5	my-geoflow-record3	My Title 3	abstract:My metadata descrip TH1@http://somevocab?thesaurus=th1:kw1@http://somevocab?keyword=kw purpose:Some purpose for this TH2:kw1, kw2, kw3 info:some supplemental info 3 TH3:kw1, keyword 2@http://someret/kw2, keyword 3	metadatas:emmanuel.blondel1@gmail.com_ owner:emmanuel.blondel1@gmail.com_ pointOfContact:julien.barde@ird.fr,wilfried.heintz@inra.fr	2019-03-07	dataset	eng		
6	my-geoflow-record4	My title 4	abstract:My metadata descrip TH1@http://somevocab?thesaurus=th1:kw1@http://somevocab?keyword=kw purpose:Some purpose for this TH2:kw1, kw2, kw3 info:some supplemental info 4 TH3:kw1, keyword 2@http://someret/kw2, keyword 3	metadatas:emmanuel.blondel1@gmail.com_ owner:emmanuel.blondel1@gmail.com_ pointOfContact:julien.barde@ird.fr,wilfried.heintz@inra.fr	2019-02-14	dataset	eng	SRID=4326,POLYGON	
7	my-geoflow-record5	My title 5	abstract:My metadata descrip TH1@http://somevocab?thesaurus=th1:kw1@http://somevocab?keyword=kw purpose:Some purpose for this TH2:kw1, kw2, kw3 info:some supplemental info 5 TH3:kw1, keyword 2@http://someret/kw2, keyword 3	metadatas:emmanuel.blondel1@gmail.com_ owner:emmanuel.blondel1@gmail.com_ pointOfContact:julien.barde@ird.fr,wilfried.heintz@inra.fr	2019-02-14	dataset	eng		
8									
9									
14									
15									

example definition Sheet3

Explore

DC.Title	DC.Description	DC.Subject

# geoflow, exécution: `executeWorkflow("config.json")`

Un fichier de configuration "config.json" à adapter selon les besoins:

- Deux fichiers essentiels:
  - Métadonnées: données décrites avec un modèle pivot (table de 15 colonnes DCMI avec des conventions de syntaxe ). Stockées en CSV ou SQL, fichier, [Google Spreadsheet](#) et PostgreSQL (implicitement table "metadata")
  - un [fichier pour gérer les contacts](#) (basé sur 19115, ~ EML)
- Lister les logiciels ou les infrastructures ciblés:
  - Les connexions possibles:
    - Outils: Google Drive, PostgreSQL/PostGIS, Geonetwork, Geoserver (*vecteur*), Zenodo
    - Avec des protocoles standardisés (ex: OGC WFS et CSW)
  - "A migrer": Dataverse, Thredds, autres??
  - "Perspectives": OGC WCS, GeoServer (*raster*), LDAP (source de contacts),
  - Problème avec Geonode (CSW-T)



# geoflow: fichier de configuration config.json

```
"metadata": {  
  
  "entities": {  
  
    "handler": "gsheet",  
  
    "source": "https://docs.google.com/spreadsheets/d/1iG7i3CE0W9zVM3QxWfCjoYbj1dQvKsMnER6kqwDiqM/edit?usp=sharing"  
  
  },  
  
  "contacts" : {  
  
    "handler": "gsheet",  
  
    "source": "https://docs.google.com/spreadsheets/d/144NmGsidIRE578IN0McK9uZEUHZdBuZcGy1pJS6nAg/edit?usp=sharing"  
  
  }  
  
}
```



## geoflow: fichier de configuration config.json

```
"metadata": {  
  
  "entities": {  
  
    "handler": "csv",  
  
    "source": "~/metadata.csv"  
  
  },  
  
  "contacts" : {  
  
    "handler": "csv",  
  
    "source": "~/contacts.csv"  
  
  }  
  
}
```

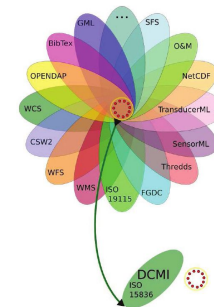


# geoflow: fichier de configuration config.json

```
"metadata": {  
  
  "entities": {  
  
    "handler": "dbi",  
  
    "source": "metadata"  
  
  },  
  
  "contacts" : {  
  
    "handler": "csv",  
  
    "source": "~/contacts.csv"  
  
  }  
  
}
```

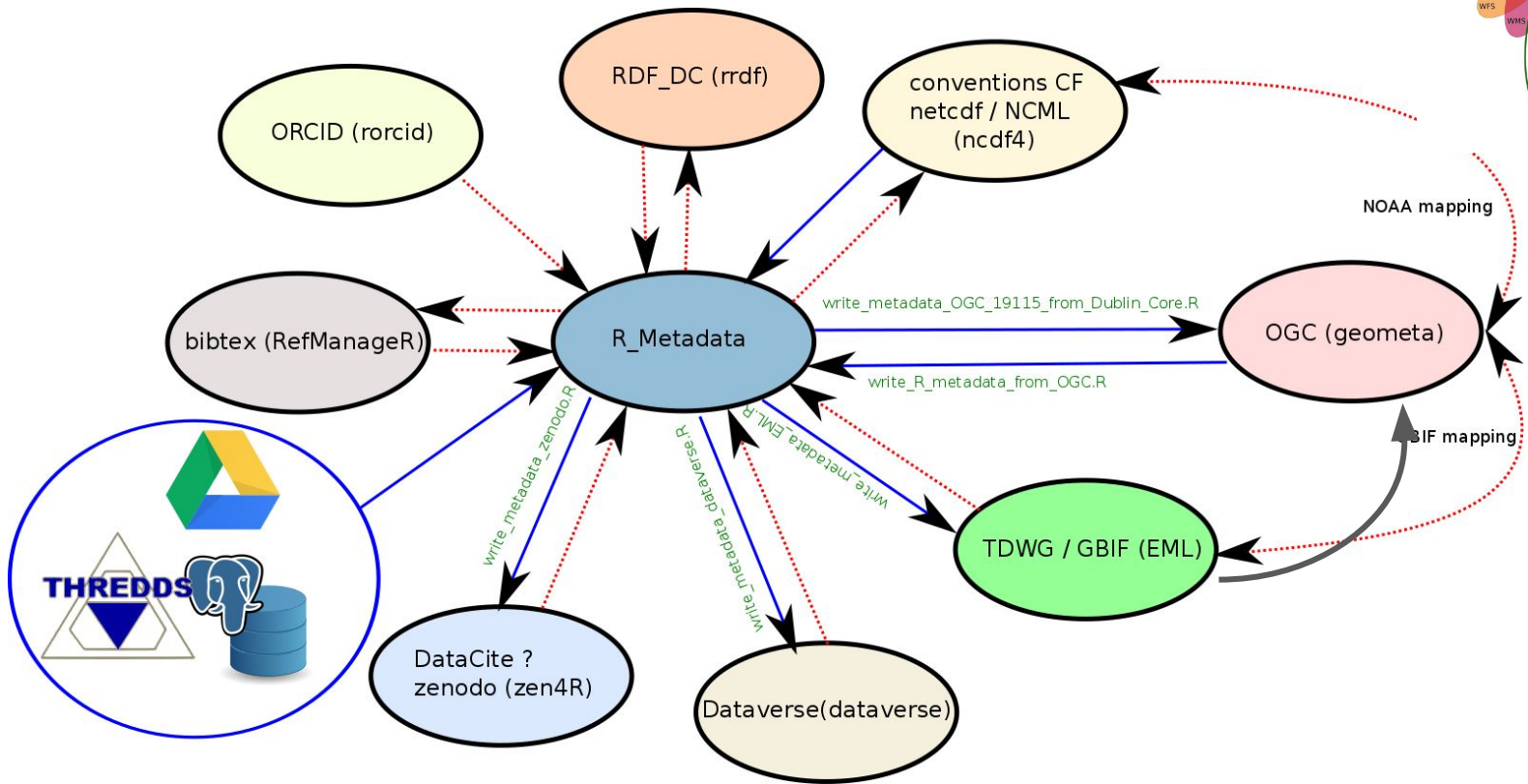
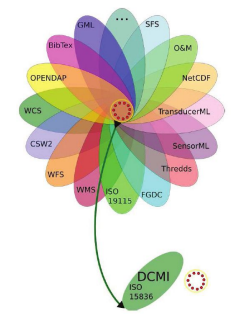


# geoflow: mapping de métadonnées



- Les normes ciblées:
  - DCMI, Datacite (Zenodo API), ORCID, sword (API Dataverse)
  - OGC: Formats (19115/19110/19119) et Protocoles (CSW/WMS/WFS/WCS/OPeNDAP)
  - EML & Darwin Core (structure de données 19110 par défaut)
- Deux approches sont expérimentées en parallèle:
  - Directement **d'une norme à une autre norme** (via geometa):
    - 19115 (geometa) vers EML, ...
  - **Du modèle pivot vers les autres normes** (codes R\_metadata à migrer dans geoflow):
    - Avantages: export vers toutes les normes gérées à partir du modèle pivot
    - Inconvénients: restreint aux éléments gérés par le mapping avec le modèle pivot

# geoflow: mapping de métadonnées



# geoflow (spreadsheet, GN, GS) => fiche simple



URD Catalogue de métadonnées de l'UMR MARBEC / OSU-OREME   S'identifier Français

[Retour à la recherche](#)

## Carte de sensibilité du complexe récifal de Foulpointe (Madagascar)

Mis à jour : il y a 2 ans

La carte de sensibilité écologique récifale de Foulpointe à Madagascar. Elle a été réalisée dans le cadre des programmes environnementaux de la Commission de l'Océan Indien portant sur l'étude cartographique et sédimentologie du récif de Foulpointe.

A Madagascar, comme dans les autres pays de la COI, la politique environnementale revêt un aspect prioritaire par le lien direct qui existe entre la qualité de l'environnement marin et les activités socio-économiques intéressant la bande côtière tourisme , pêche, urbanisation, etc.

L'étude dont est issue cette carte entre dans cette politique. Son objectif principale est de comprendre les causes du phénomène d'hyperméditerranéisation du récif de Foulpointe en vue de sa gestion durable. Elle a été réalisée pour le compte du Ministère de l'Environnement, des Eaux et Forêt par la Division de la Recherche et d'Application de la Gestion Intégrée des Zones Côtières du Centre National de Recherche sur l'Environnement (CNRE) avec l'appui technique de l'Agence pour la Recherche et la Valorisation Marine (ARVAM). Le financement a été assuré par la COI via les reliquats du PRE/COI-UE.

La présente carte classe les différentes entités du récif selon leur sensibilité écologique qui prend en considération de nombreux paramètres faisant appel à des caractéristiques structurelles et/ou fonctionnelles des écosystèmes.

*Under development*

### Téléchargements et liens

Lien de la référence zotero du document associé

Cartographie et bilan diagnostic du complexe récifal Foulpointe. (Cote Est de Madagascar)

### À propos de cette ressource

Catégories

Mots-clés

- Récifs coralliens
- Récifs
- Sensibilité
- Ecologie
- Océan Indien
- Madagascar
- Foulpointe
- Commission de l'Océan Indien
- Programme Régional Environnement PRE/COI-UE
- CNRE
- ARVAM
- Biota
- Océans
- Environnement

Langue

- Anglais

### Aperçu

# geoflow (Postgis, GN, GS) => Tuna Fisheries

zenodo

Search

q

Upload

Communities

Log in

Sign up

February 1, 2018

Dataset Open Access

## Global monthly catch of tuna, tuna-like and shark species (1950-2015) aggregated by 1° or 5° squares (IRD level 2)

Paul Taconet; Emmanuel Chassot; Julien Barde

### Data collector(s)

Fabio Fiorellato; Carlos Palma; Peter Williams; Nickolas Vogel; Colin Millar

This dataset lists global catch of tuna, tuna-like and shark species from 1950 to 2015. Catches are stratified by month, species, gear, vessel flag reporting country, fishing mode (i.e. type of school used), area (mainly 1° or 5° square) and unit of catch (weight or number). This dataset was computed using public domain catch-and-effort datasets released by the five tuna Regional Fisheries Management Organizations: the Indian Ocean Tuna Commission (IOTC), the International Commission for the Conservation of Atlantic Tunas (ICCAT), the Inter-American Tropical Tuna Commission (IATTC), The Western and Central Pacific Fisheries Commission (WCPFC), the Commission for the Conservation of Southern Bluefin Tuna (CCSBT).

IRD Level 2 stands for the processes applied to the primary datasets by the French National Research Institute for Sustainable Development (IRD) to generate the dataset. These processes for level 2 at global scale are :

- Catch-and-effort data are disseminated in such way that redundancy may exist between the various datasets released, and that dimensions may be split over the datasets for some strata. To cope with these issues and get one single and more complete possible value of catch per stratum (i.e. with all the available dimensions), these datasets had to be merged in specific ways - i.e. not simply merging them but removing the duplicated strata or reassembling the strata with all the available dimensions split over the datasets.
- Values expressed in weight were kept and numbers were converted into weight using simple conversion matrices (A. Fonteneau, pers. com). These conversion factors depend on the species, the gear, the year and the main geographical area (equatorial or tropical). They were computed from the Japanese and Taiwanese size-frequency data as well as from the Japanese total catches and catch-and-effort data. Some data might not be converted at all because no conversion factor exists for the stratum: those data were kept and the unit of catch was set to Number of fishes harvested.
- Geo-referenced catches were raised to the total catches.
- Data located at land or without any spatial information were equally redistributed on data at sea on areas with same characteristics (same year, month, gear, flag, species, type of school).
- In the overlapping zone between the IATTC and the WCPFC, only data from the IATTC was kept (i.e. for this specific zone, data from the WCPFC was removed). IATTC data was kept rather than WCPFC's because the information on vessel flag reporting country is available in IATTC datasets contrary to WCPFC's ones.
- For the Southern Bluefin Tuna, only data from the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) was kept (i.e. data from the other RFMOs for SBF was removed).
- Original code lists have been mapped with standard FAO code lists (for gears, species, flags). These mappings have been done in collaboration with the RFMOs secretariats.

More details on the processes are provided in the lineage section.

Characteristics, specific issues and problems that need to be taken into account :

115

views

121

downloads

[See more details...](#)

Indexed in

OpenAIRE

### Publication date:

February 1, 2018

### DOI:

[10.5281/zenodo.1164128](https://doi.org/10.5281/zenodo.1164128)

### Keyword(s):

[tuna](#) [fisheries](#) [tuna atlas](#) [catch](#) [time series](#)  
[geographic information system](#) [log school](#)

### Related identifiers:

Identical to

[https://tunaatlas.d4science.org/geonetwork/srv/eng/catalog\\_search#/metadata/global\\_catch\\_tunaatlasIRD\\_Level2](https://tunaatlas.d4science.org/geonetwork/srv/eng/catalog_search#/metadata/global_catch_tunaatlasIRD_Level2)  
[https://tunaatlas.d4science.org/tunaatlas/index.html?dataset=global\\_catch\\_tunaatlasIRD\\_Level2](https://tunaatlas.d4science.org/tunaatlas/index.html?dataset=global_catch_tunaatlasIRD_Level2)

### Communities:

[Fisheries and aquaculture](#)

### License (for files):

[CC](#) Creative Commons Attribution Non Commercial 4.0 International

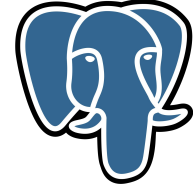
### Versions

Version 1	Feb 1, 2018
<a href="https://doi.org/10.5281/zenodo.1164128">10.5281/zenodo.1164128</a>	

Cite all versions? You can cite all versions by using the DOI

## ... & Dataverse IRD ??

# geoflow (Postgis, GN, GS) => Tuna Fisheries



Activités | Navigateur Web Chromium | mar. 13 5:08

Catalogue de méta- x | thredds.oreme.org:8080/geonetwork/srv/eng/catalog.search#/metadata/origin\_institution

Catalogue de métadonnées de IUMR MARBEC / OSU-OREME | Search | Map | Sign in | English

Back to search | < Previous | Next >

## Monthly catch of tuna, tuna-like and shark species (1950-2015) by purse seiners and pole-and-liners in the Indian, Atlantic and Eastern Pacific Oceans aggregated by 1° squares (IRD level 2)

Updated: 19 days ago

This dataset lists catch of tuna, tuna-like and shark species by purse seiners and pole-and-liners from 1950 to 2015 in the Indian, Atlantic and Eastern Pacific Oceans. Catches are stratified by month, species, gear, vessel flag reporting country, fishing mode (i.e. type of school used), area (1° square) and unit of catch (weight or number). This dataset was computed using public domain catch-and-effort datasets released by four of the five Tuna Regional Fisheries Management Organizations: the Indian Ocean Tuna Commission (IOTC), the International Commission for the Conservation of Atlantic Tunas (ICCAT), the Inter-American Tropical Tuna Commission (IATTC), the Commission for the Conservation of Southern Bluefin Tuna (CCSBT). Data from the Western Pacific Ocean were not included because WCPFC provides purse seine and pole-and-line catch at 5° square spatial resolution.

IRD level 2 stands for the processes applied to the primary datasets by the French National Research Institute for Sustainable Development (IRD) to generate the dataset. These processes for level 2 at global scale are :

- Catch-and-effort data are disseminated in such way that redundancy may exist between the various datasets released, or that dimensions may be split over the datasets for some strata. To cope with these issues and get one single and more complete possible value of catch per stratum (i.e. with all the available dimensions), these datasets had to be merged in specific ways - i.e. not simply merging them but removing the duplicated strata or reassembling the strata with all the available dimensions split over the datasets.
- Values expressed in weight were kept and numbers were converted into weight using simple conversion matrices (A. Fonteneau, pers. com). These conversion factors depend on the species, the gear, the year and the main geographical area (equatorial or tropical). They were computed from the Japanese and Taiwanese size-frequency data as well as from the Japanese total catches and catch-and-effort data. Some data might not be converted at all because no conversion factor exists for the stratum: those data were kept and the unit of catch was set to Number of fishes harvested. Some data might not be converted at all because no conversion factor exists for the stratum: those data were kept and the unit of catch was set to Number of fishes harvested.
- Geo-referenced catches were raised to the total catches.
- Data located at land or without any spatial information were equally redistributed on data at sea on areas with same characteristics (same year, month, gear, flag, species, type of school).
- For the Southern Bluefin Tuna, only data from the Commission for the Conservation of Southern Bluefin Tuna (CCSBT) was kept (i.e. data from the other RFMOs for SBT was removed).
- Original code lists have been mapped with standard FAO code lists (for gears, species, flags). These mappings have been done in collaboration with the RFMOs secretariats.

In addition, data that were provided at resolutions superior to 1° x 1° were disaggregated to the corresponding 1° x 1° quadrants by dividing the catch equally on the overlappings 1° x 1° quadrants.

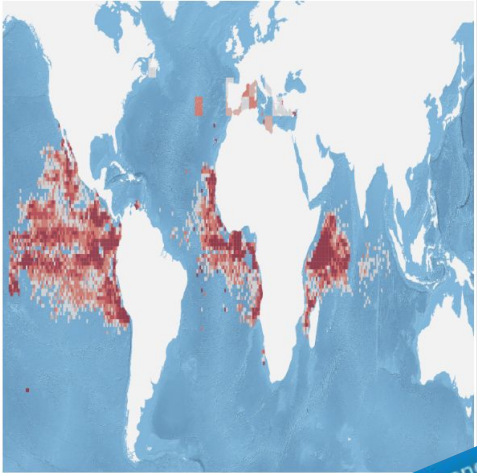
More details on the processes are provided in the lineage section.

Characteristics, specific issues and problems that need to be taken into account :


In addition to the respective specific issues of all the tuna RFMOs datasets, the following issues might be taken into account :

- For confidentiality policies, information on flag and school type for the geo-referenced catches is available in separate files for East Pacific Ocean (IATTC) Purse seine datasets. For each stratum, the catch from the flag-detailed dataset was raised to the catch from the school type-detailed dataset to get an estimation of the catches by flag and school type in each stratum.
- Data provided in number of fishes harvested for the Southern Bluefin Tuna (SBF) were not converted into weight of fishes, because no factors of conversion are available for this species. This might represent a great amount of the data for SBF.
- Geo-referenced catches were raised to the total catches for all RFMOs. Depending on the availability of the flag dimension (currently not available for the geo-referenced catch-and-effort dataset from the Western-Central Pacific Ocean), the dimensions used for the raising are either (Flag, Species, Year, Gear) or (Species, Year, Gear). Some catches cannot be raised because the combination (Flag, Species, Year, Gear) (resp. (Species, Year, Gear)) does exist in the geo-referenced catches but the same combination does not exist in the total catches. In this case, non-raised catch data were kept. Most catch-and-effort data have catches inferior to the catch available in the nominal catch dataset for a given stratum. However, in some cases the value of catch in the catch-and-effort data can be greater than the one in the nominal catch. In this case, the catch was 'downgraded' to the nominal catch one.
- Some codes could not have been mapped to standard code lists, for some RFMOs own-defined codes that usually are aggregation of existing codes (e.g. flag IDPH - Indonesia and Philippines for WCPFC; species Otun - other tuna for ICCAT). In those cases, the code for species was set to UNK (Unknown). However, these codes have been mapped with more aggregated code lists i.e. group of species.
- Some data can be expressed at temporal resolutions greater than 1 month.

Overview



Spatial extent

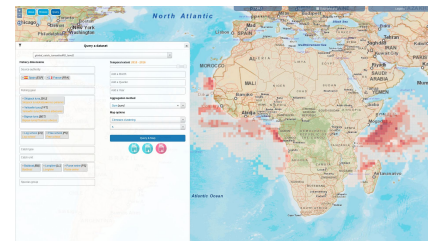


DOI 10.5281/zenodo.1145786

tentative



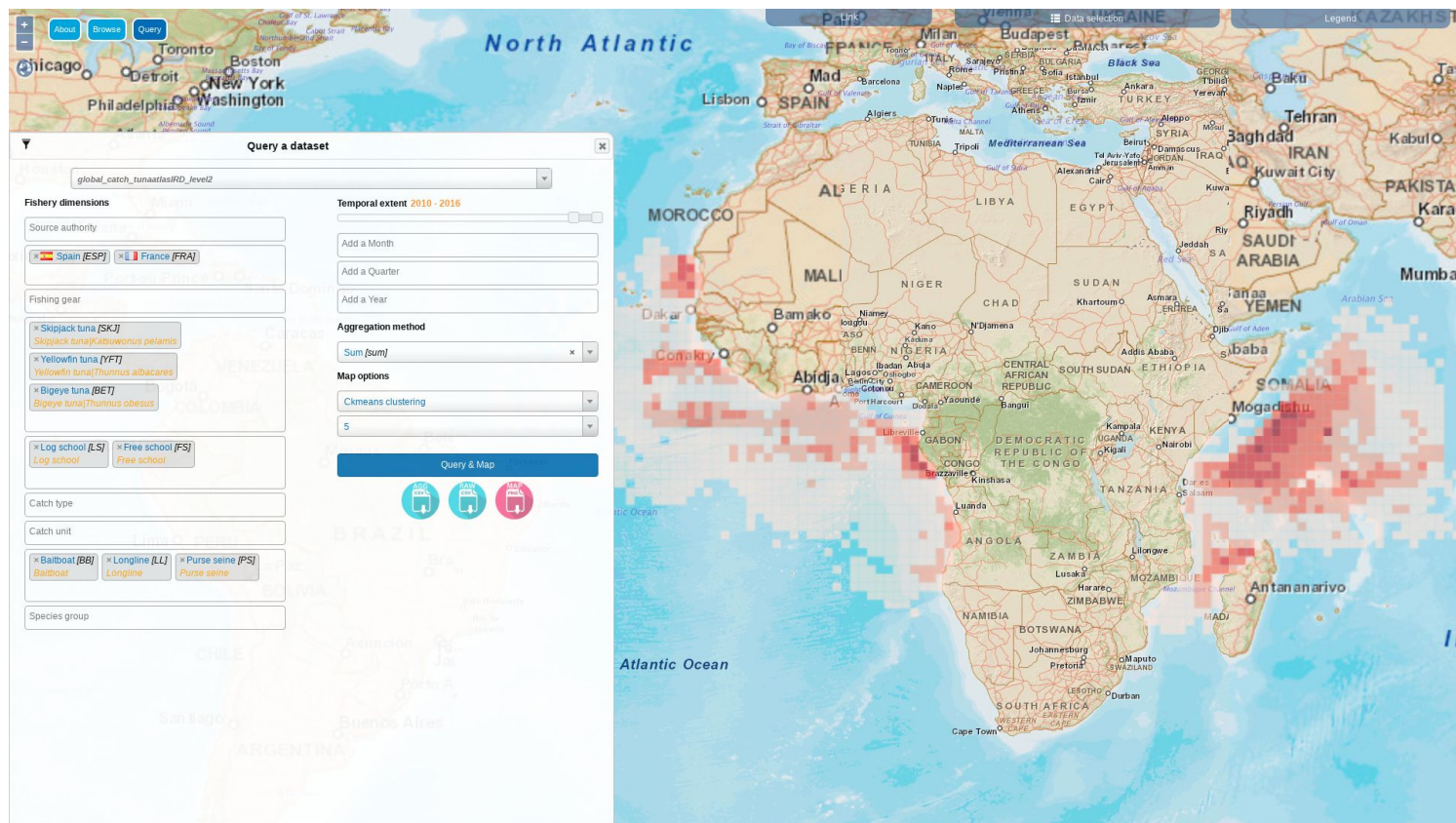
# geoflow: FAIR workflow



Assurer les principaux services à partir de “métadonnées riches”:

- Production de métadonnées conformes aux standards OGC data & services
  - 19115 (modèle pivot)
  - 19110 (autre modèle pivot dans un CSV / google doc)
  - 19119 (publiés dans Geoserver)
- Validation du contenu des métadonnées: un viewer FAIR ([OpenFairViewer](#))
  - Métadonnées avec DOI,
  - Accès aux données,
  - Visualisation cartographique,
  - Requête / Filtres sur les données sur des dimensions ou variables communes à différentes structures de données.

# geoflow: FAIR workflow (exemple du Tuna atlas)



# Conclusion

- **Input:** méthode générique et indépendante des standards basée sur les métadonnées et mise en oeuvre avec des outils simples:
  - **Tableaux de saisie** simples (pour les utilisateurs):
    - **Métadonnées principales:** modèle pivot (DCMI) et mapping (19115, EML...),
    - **Structure de données:** basée sur 19110, CF-conventions (CSV, shapefiles, SQL, nc..),
    - **Contacts**
  - **Codes R** (pour les développeurs) avec automatisation si lecture possible de la donnée
- **Output:** FAIR DMP en privilégiant les standards OGC (Réutilisable = effort sur la structuration des données),
- **Replicabilité:** gestion des configurations et logs, exécution en ligne...
- **Upgrade** (releases) prévus (Blue Cloud..):
  - table générique pour la structure des données (utilisée par geometa pour produire du 19110)
  - Dataverse / Thredds / IPT
- Workflow **customisable** pour gérer au cas par cas



# Perspectives

- Interface **Shiny** pour l'exécution du workflow sans toucher le code R, Json..
- Appuyer l'**interopérabilité sémantique**: vocabulaires contrôlés (SKOS..),
- Validation & Publication des métadonnées **INSPIRE**
- Gestion de (méta-)données **raster**
- Gérer plus facilement les contacts avec des annuaires: LDAP & ORCID
- EML & Darwin Core vers OGC
  - Lire les archives sur un IPT
  - Mapping direct ou indirect (via modèle pivot)
  - Produire une version 19110 du Darwin Core
- Interagir avec des plateformes **cloud** (eg nextcloud, owncloud): eg [B2DROP](#),
- Organiser un nouveau **workshop** dans le cadre du projet H2020 Blue Cloud / EOSC avec des **VRE** D4Science (RStudio serveur, Jupyter Notebook)
- Alimenter d'autres documents "officiels" : PGD, Outils de suivi qualité ...

# Remerciements

- [Blue Cloud H2020 project](#) (grant agreement ID: 862409, Funded under: H2020-EU.3.2.5.1.),
- [BlueBridge](#) H2020 project (grant No. 675680): e-infrastructure for marine domain
- [OpenAIRE-Connect](#) H2020 project (grant No. 731011): Open Science, Fisheries
- RDA:
  - [RDA EU 4.0](#)
  - [Fisheries Data Interoperability WG](#)

