

# Standards de métadonnées, en environnement

Julien Barde

<https://tinyurl.com/y6jn9zwa>



# Plan

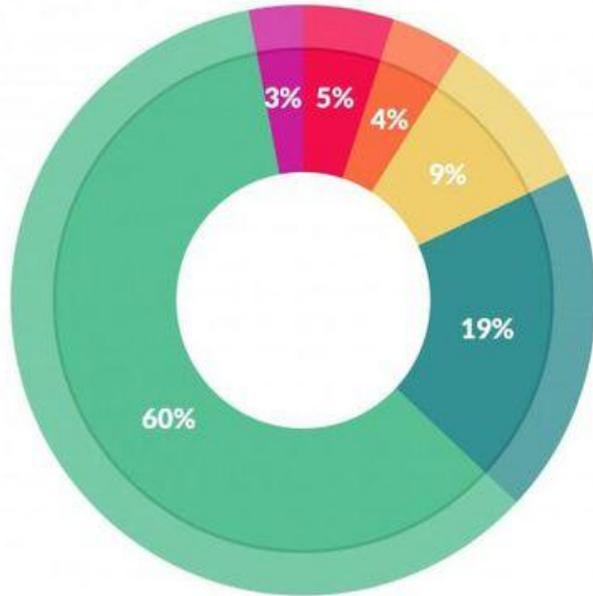
- Terminologie
- Quels types, formats, granularité de données ?
- Objectifs et enjeux liées à la gestion des métadonnées
- Principaux standards de métadonnées en environnement
- Mise en oeuvre (codes et applications)
- Points clés
- Quelques cas d'utilisation

# Terminologie



- Donnée ? Fichiers, bases, entrepôt / data warehouse, cubes...
- Environnement ? Ecologie, biodiversité, paramètres environnementaux..
- Métadonnée = interopérabilité syntaxique et sémantique
  - Eg Harvest (“moissonner”)
- Métadonnée = donnée (dans ou en dehors),
  - métadonnée encapsulée dans la donnée (eg NetCDF, JPEG)
  - métadonnée dans un fichier à part (le fichier “.xml” d’un shapefile)
- Normes et standards (schémas, vocabulaires..):
  - Sont composés d’éléments de métadonnées,
  - La valuation des éléments peut être contrôlée
- Mapping: correspondance entre les éléments de métadonnées des normes

# Data Science & Big Data (Machine / Deep Learning)

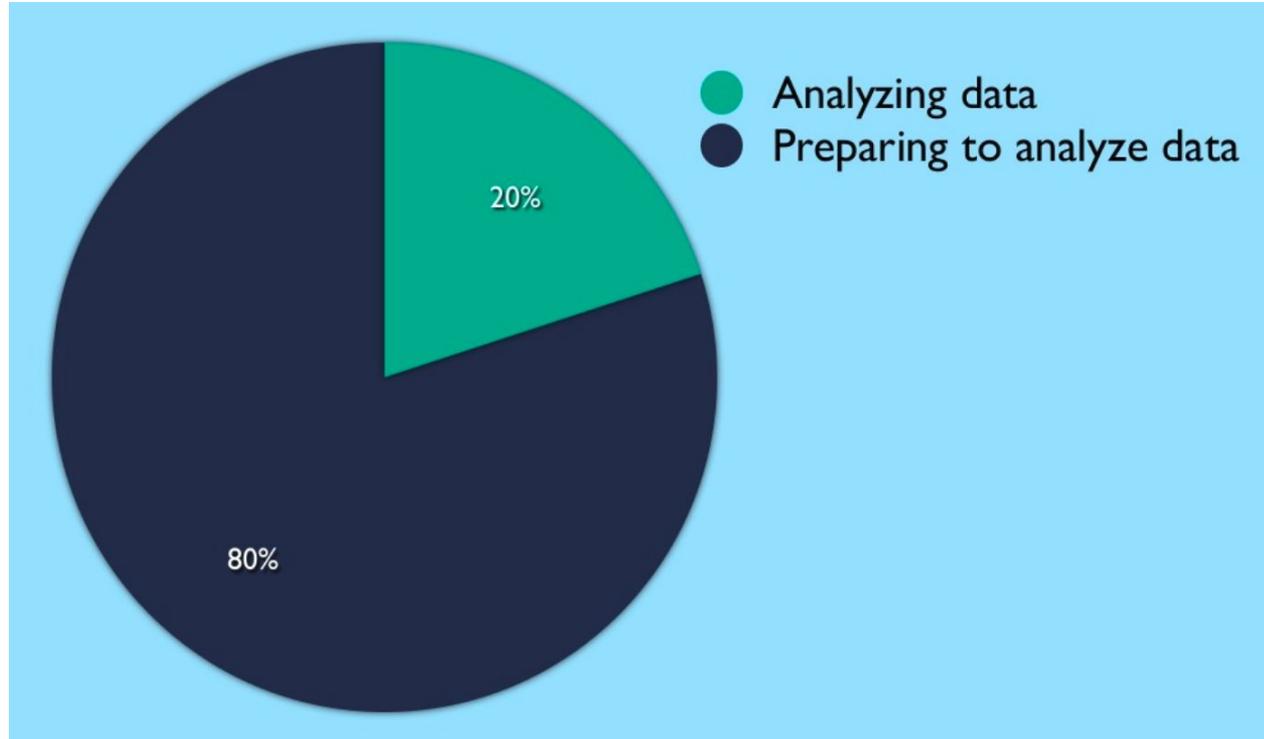


What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets; 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

“76% of data scientists view data preparation as the least enjoyable part of their work” ([Source: Forbes](#), 2016)

# Data Science & Big Data (Machine / Deep Learning, IA)



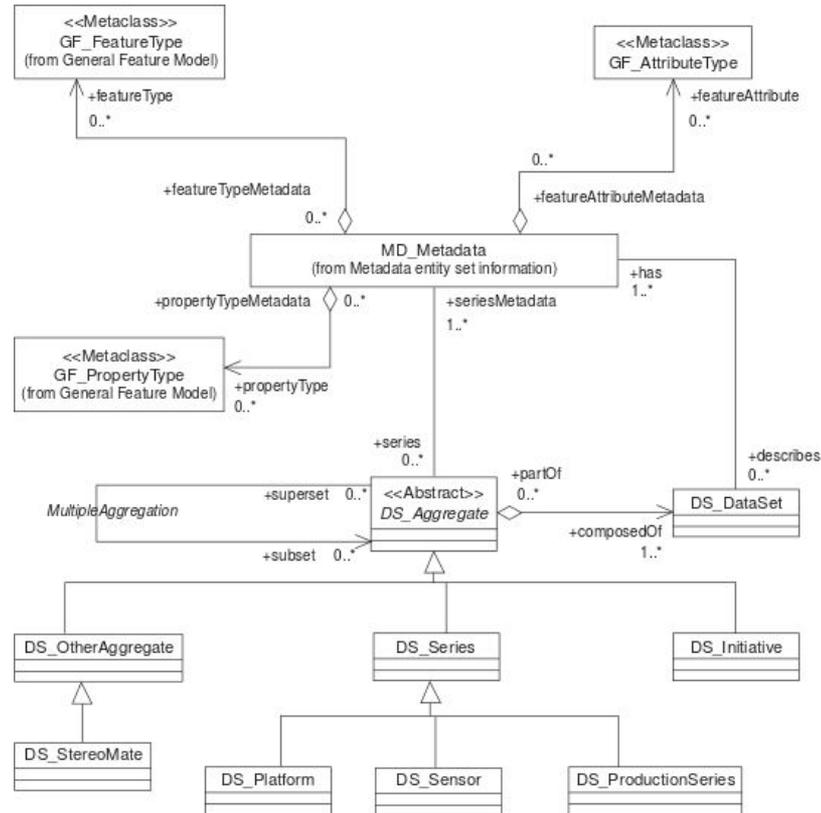
“80% of data scientists time spent on data preparation” ([IBM](#)) => “Open Data ≠ Clean data”.. Need for data catalogs

# Quels types (et formats) de données ?

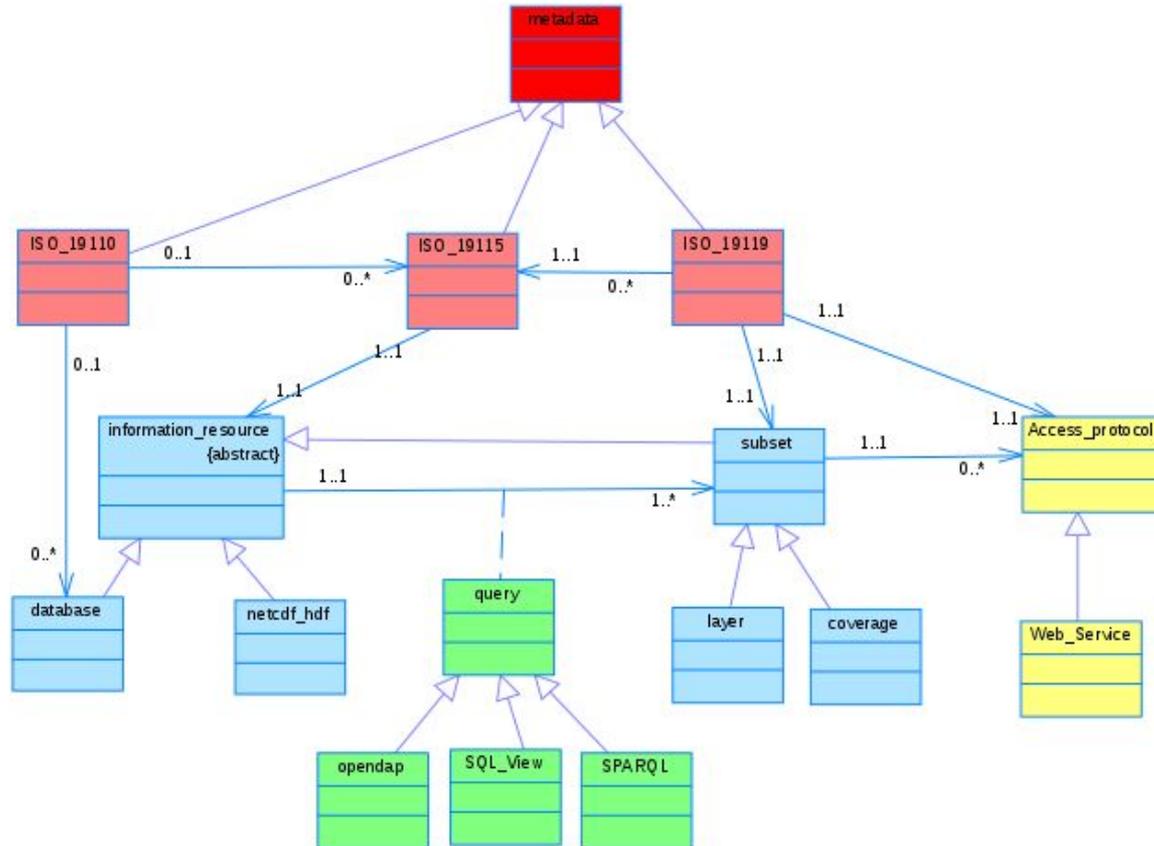
Importance des dimensions temporelle et spatiale pour l'environnement :

- Types de données
  - Observations:
    - Occurrences d'espèces
    - Photos
    - Biométrie
  - Analyses (& protocoles !)
  - Capteurs:
    - In situ
    - Télédétection
  - Modèles / simulation
- Formats clés:
  - tabulaires: CSV & SQL & NetCDF, raster ou vecteur,
  - Différentes structures: Darwin Core, CF conventions..

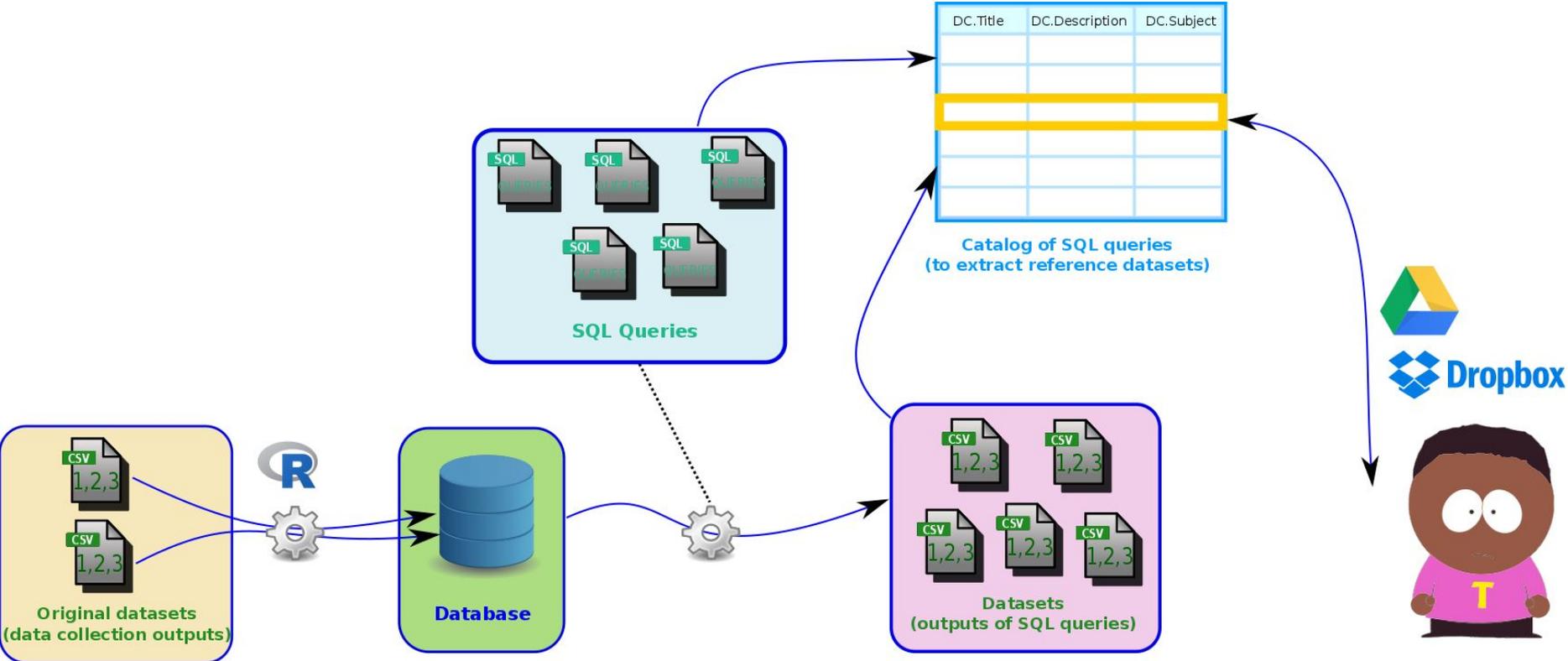
# Sources de données et jeux de données



# Sources de données et jeux de données



# Bases de données et jeux de données



# Découvrir un jeu de données ?

## Exemple d'une base de données SQL

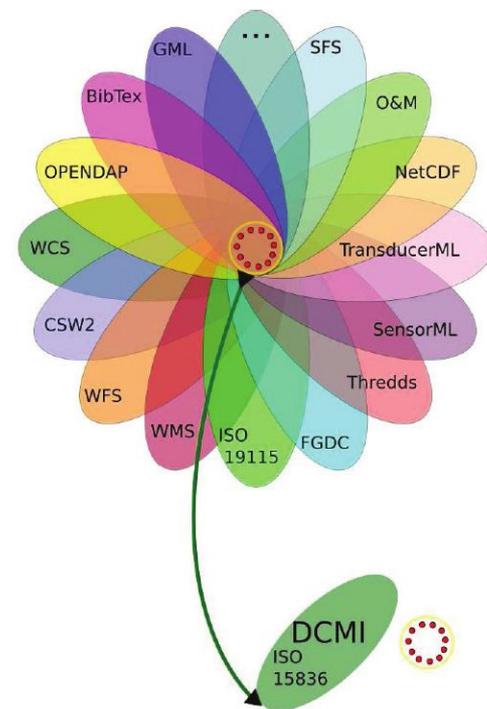
- Dans son unité:
  - Culture de la transmission orale (désintérêt pour la métadonnée),
  - Yvan Bédard: “Les métadonnées ne remplaceront jamais la machine à café”,
- Découverte en ligne par les métadonnées :
  - Grands catalogues: Google (data search), Zenodo, Datasearch, GBIF...
  - Petits catalogues: comment savoir qu'ils existent s'ils ne sont pas moissonnés par les autres ?
- Une fois qu'on sait qu'elle existe on souhaite généralement :
  - Avoir le modèle et le dictionnaire de données
  - Avoir des requêtes SQL de référence: catalogues de requêtes
  - Savoir qui l'administre et gère les accès,
  - Connaître la provenance: protocole et pré-traitements,
  - Connaître l'ouverture ou les restrictions d'accès....

# Objectifs des métadonnées

Des normes avec un **noyau commun** et des spécificités

Selon la richesse des éléments de métadonnées:

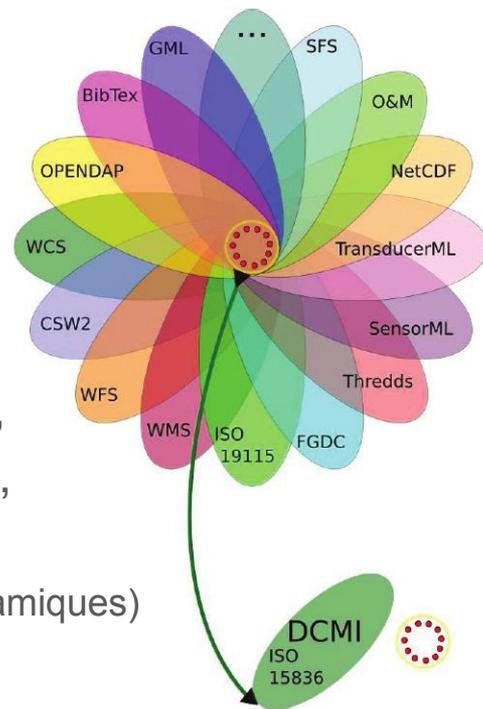
- Découverte,
- Visualisation,
- Accès,
- Extraction / Requêtes:
  - => description de la **structure des données**
- Citation, License.,
- Interopérabilité

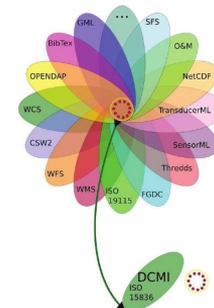


Les métadonnées sont le point clé pour les Plans de Gestion de Données.

# Enjeux liés aux métadonnées

- Qui doit décrire ? => le(s) producteur(s)
- Garantir un minimum d'éléments bien décrits:
  - Qui, quoi, quand, où, comment...
- Répliquer les métadonnées dans différents catalogues,
- Aller jusqu'à la description de la structure des données,
- Mise en place de workflows pour les mises à jour
  - les données évoluent et certaines métadonnées avec (dites dynamiques)
  - Automatiser les mappings:
    - vers une saisie unique
    - Synchronisation de la publication dans différents systèmes
- DOIs pour être cité... idéalement via un Data paper.

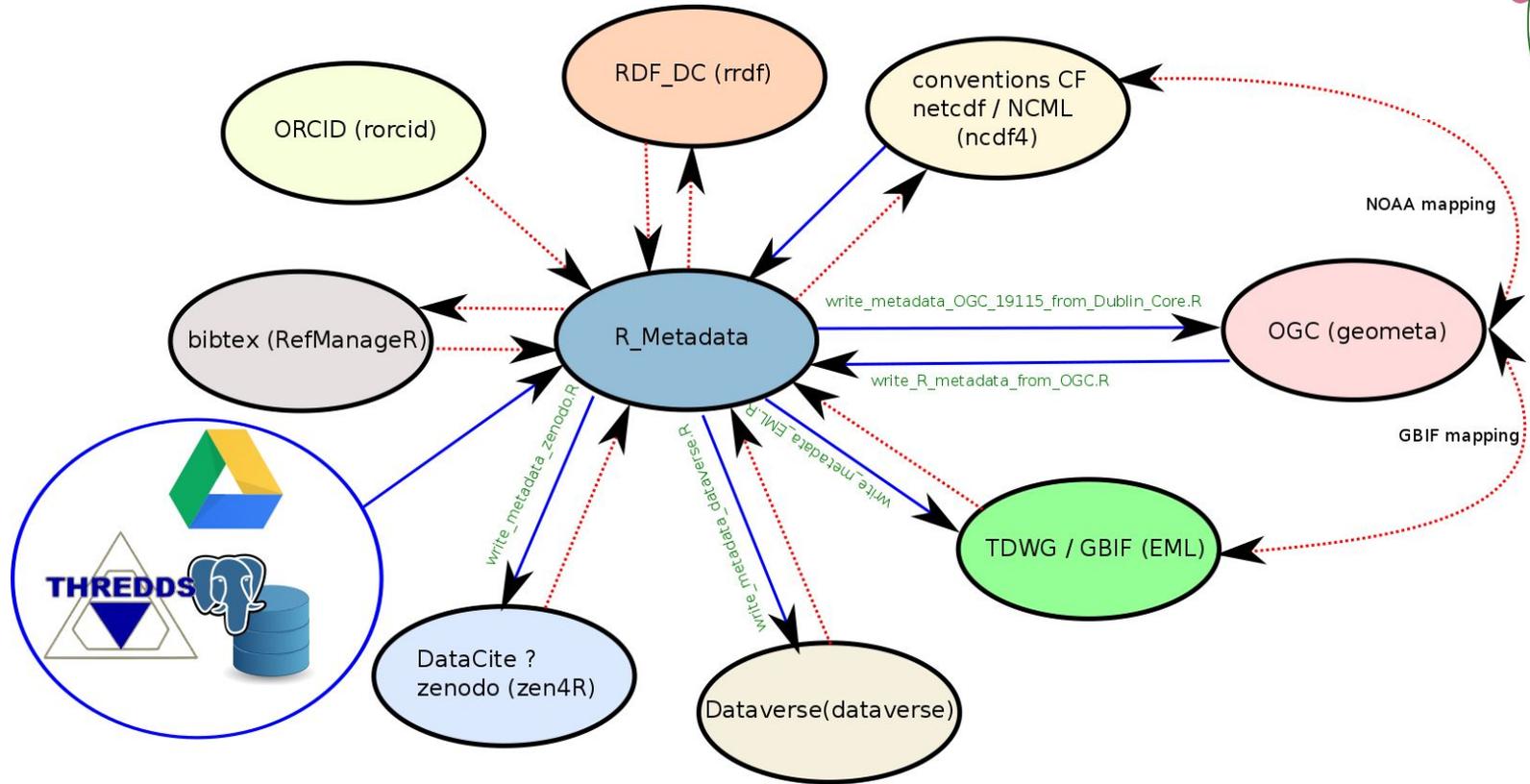
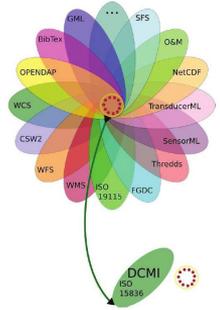




# Métadonnées & valeurs essentielles

- **Identifiant**: free text or [DOIs](#), [URI](#), [URNs](#)..
- **Title | Description** : free text
- **Creator**: free text ou [ORCID](#) / [FOAF](#)
- **Subject**: free text or [controlled vocabularies](#) (eg [GEMET](#), [GCMD](#))
- **Spatial and temporal coverage**: [controlled syntax](#) (dynamiques)
- **Date**: [controlled syntax](#)
- **Type**: free text
- **Format**: free text
- **Language** : norme [ISO 639](#)
- **Relation** : free text or [URLs](#), [URIs](#)...
- **Provenance**: free text
- **Rights**: free text or [Creative Commons](#)....

# Mapping de métadonnées (eg packages R)



# Principaux standards de métadonnées

Names	Related Standards	Scope	R Implementation
<a href="#">Dublin Core</a> & related ( <a href="#">DCAT</a> , Datacite..)	XML & RDF	<b>Any</b> resource	Ongoing with R packages (eg <a href="#">oaih</a> , rrdf, redland)
<a href="#">Datacite</a>	XML, OAI-PMH	<b>Any</b> resource ("discipline-agnostic")	Indirect via zen4R (R binding for Zenodo API)
<a href="#">ISO TC 211 &amp; OGC</a> : 19115/119/110, WPS	GML & NetCDF, WMS/WFS/WCS	<b>Spatial</b> data, processes	Ongoing with <a href="#">geometa</a> R package (along with <a href="#">geonapi</a> & <a href="#">geosapi</a> )
<a href="#">TDWG</a> : EML	Darwin Core	<b>Biodiversity</b> data (occurrences, specimens, observations and measures..)	Ongoing with <a href="#">EML R package</a>
<a href="#">CF Conventions</a>	NetCDF, OPeNDAP, NCML	<b>Ocean and climate</b> : in situ sensors, remote sensing, model outputs. Parameters: physical, chemical...	Ongoing with <a href="#">ncdf4 R package</a> to handle NetCDF ( <a href="#">metadata</a> embedded) with OPeNDAP
<a href="#">FAST</a>	NetCDF	<b>Acoustic data</b>	Not for now..will be done with <a href="#">ncdf4 R package</a>
FIRMS	CSV	<b>Fisheries data</b>	<a href="#">rtunaatlas</a>

# Dublin Core Metadata Initiative



Schéma générique:

- Standard historique ([norme ISO 15836](#)),
- 15 éléments de métadonnées basiques, tous optionnels et répétables (0..\*)  
=> toujours conforme !
- Des extensions qui rendent la norme beaucoup plus complète: [DCMI terms](#)
- Différentes versions:
  - Schéma XML
  - [Schéma RDF](#) => SPARQL endpoint comme protocole d'accès
- Protocole d'accès:
  - SPARQL, CSW..
- Outils: beaucoup de logiciels et applications
- Exemple: <https://zenodo.org/record/1145786/export/xd>

# DataCite (Consortium) Metadata Schema



Schéma générique:

- Métadonnées associées aux DOIs  $\Leftrightarrow$  DOI 10.5281/zenodo.1145786
- Pour décrire des objets digitaux (“research objects”):
  - Documents
  - Données
  - Codes
  - Images..
- Liens officiels avec les autres initiatives: DCMI, ORCID..
- Outils associés: <https://search.datacite.org>, Zenodo, Pangaea, GBIF..
- Exemple:
  - <https://schema.datacite.org/meta/kernel-4.3/example/datacite-example-full-v4.xml>
  - <https://zenodo.org/record/1145786/export/dcite4>

# ISO TC 211 & standards OGC



Normes pour la géomatique et la donnée spatiale:

- Norme pertinente pour l'environnement:
  - transversale ("cross-domain"),
  - très complète avec une version minimale proche du DCMI
- Des formats et des protocoles d'accès
  - [norme ISO 19115](#),
  - 19110: dictionnaire de données,
  - 19119: services sur les données,
  - SWE pour les capteurs (SensorML, SOS..) => observatoires
- Formats de données concernés: shapefile, WKT, KML, geojson...NetCDF
- Protocoles d'accès : CSW / WMS / WFS / WCS
- Outils associés: geonetwork, geonode...INSPIRE => obligatoire (cf [Annexes](#))





# NetCDF (standard OGC) & conventions CF

Paramètres environnementaux mesurés (in situ et télédétection) ou simulés:

- Métadonnées encapsulées dans le fichier (par le producteur):
  - Attributs globaux
  - Attributs de variables: description de la structure des données
- NCML: pour gérer les métadonnées à part et création de fichiers virtuels
- Protocole d'accès: OPeNDAP
- Termes contrôlés pour les variables
- Formats associés: HDF, zarr..
- Exemple: modèle océanographique

Format important pour caractériser le contexte environnemental d'une observation en écologie. Format clé pour les observatoires.

# EML & Darwin Core ([standards TDWG](#))



## Normes pour la biodiversité & écologie

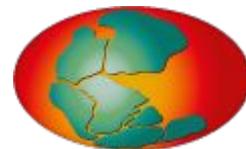
Biodiversity  
Information  
Standards  
T D W G

- Occurrences d'espèces
- Collections
- Structure de données associée: Darwin Core (schéma XML)
- Il faut faire un mapping de la structure des données pour publier sur le GBIF
- Protocole d'accès ? DOIs
- Outils associés:
  - IPT => pas de synchronisation entre données et métadonnées!
  - OBIS, eml (CRAN), Datasearch
- Exemple: [CSIRO](#)

# Mise en oeuvre

## Agents humains et logiciels

- Humains: le plus compliqué (PGD, gouvernance..)
- Logiciels:
  - Interfaces graphiques:
    - OGC: Geonetwork, Geonode, Geoserver, Constellation
    - TDWG / GBIF: IPT
  - API & Langages de programmation:
    - Python: owslib, pycsw..
    - Java: Apache SIS, Geotools, Geotoolkit..
    - R: geometa, geonapi, geosapi, eml, ncdf4
  - Approches mixtes
  - Infrastructures de données:
    - DOIs: <https://www.pangaea.de>, <https://zenodo.org/>, <https://dataverse.ird.fr>
    - SDI (eg INSPIRE) : Postgis, Thredds, Geoserver, Geonode...



# Points clés



En complément des standards

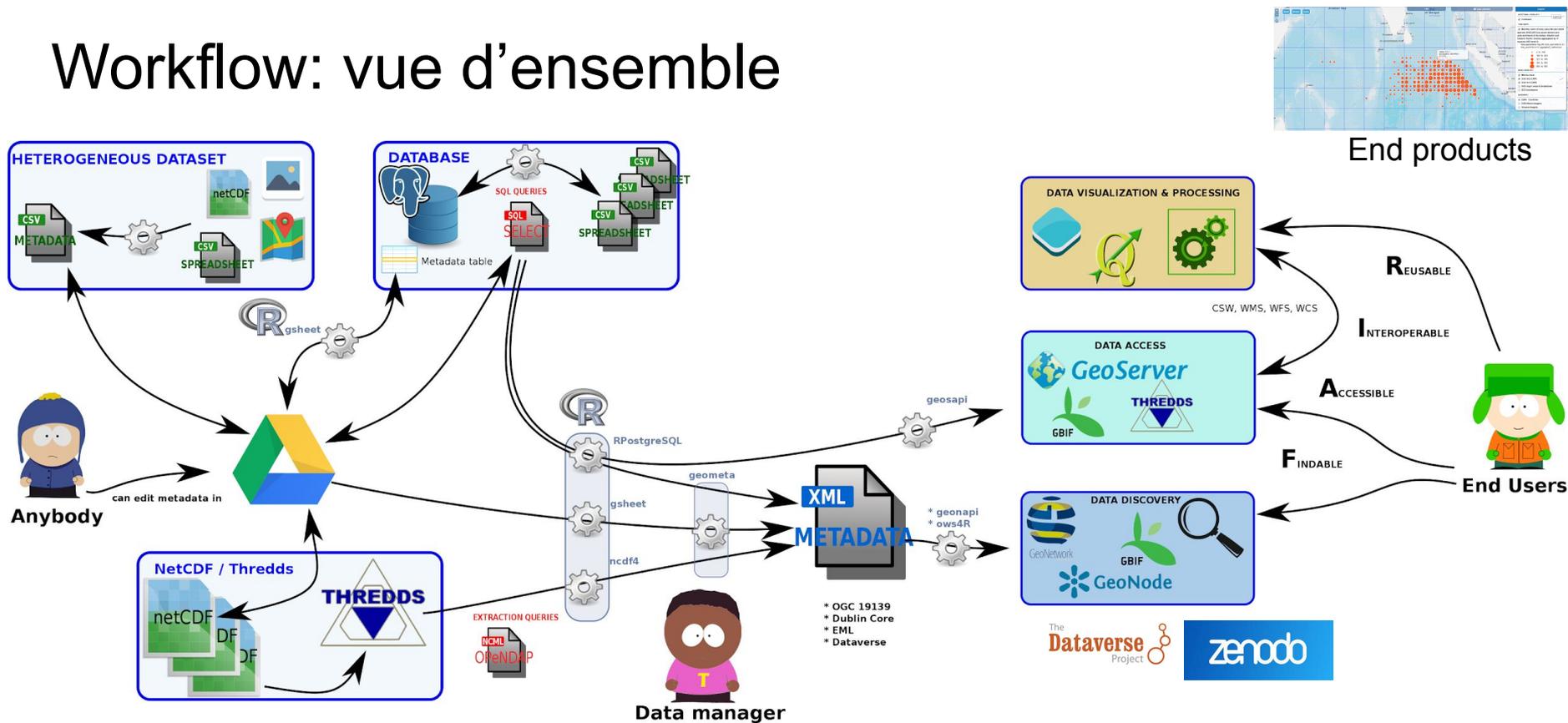
- Interopérabilité sémantique => vocabulaires contrôlés
  - Dictionnaire
  - Thésaurus
  - Ontologies
  - Annuaire: ORCID, LDAP..
  - Mieux vaut moins d'éléments de métadonnées avec plus de vocabulaires contrôlés
- Des environnements de saisie adaptés aux utilisateurs :
  - Producteurs : ergonomie des interfaces
  - Gestionnaire de données : reproductibilité et automatisation des mises à jour => workflows

# Quelques cas d'utilisation

## Quelques exemples

- Métadonnées pour la découverte
- Mapping
- Métadonnées pour l'accès et le requêtage de données

# Workflow: vue d'ensemble



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



URD Catalogue de métadonnées de l'UMR MARBEC / OSU-OREME  Rechercher  Visualiser 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'hypersédimentation 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

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

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. 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

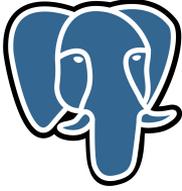
### Versions

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 ??

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



Activités | Navigateur Web Chromium | mar. 13 5:8

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 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.

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.

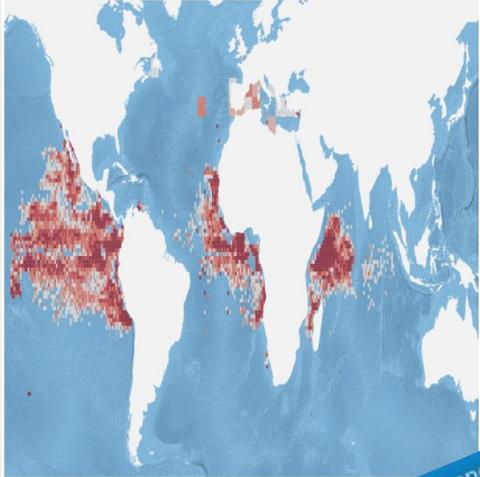
More details on the processes are provided in the lineage section.

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

# Dataverse (eg INRA, IRD...)

 **Dataverse** 🔍 User Guide Support Julien Barde 

Julien Barde Dataverse (MARBEC) Unpublished

Entrepôt IRD > **Julien Barde Dataverse**

✉ Contact 📄 Share 📄 Publish ✎ Edit

Search this dataverse... 🔍 Find Advanced Search + Add Data

**Dataverses (0)**

**Datasets (5)**

Files (0)

**Publication Status**

Draft (5)

Unpublished (5)

**Author Name**

arvam@arvam.com (5)

info@mncs-ngo.org (5)

julien.barde@ird.fr (5)

progeco@coi-loc.org (5)

secretariat@coi-loc.org (5)

More...

**Keyword Term**

Flic-en-Flac (5)

Harmony (5)

Ile Maurice (5)

Le Morne (5)

Maurice (5)

More...

**1 to 5 of 5 Results** 🔼 Sort

fr@Carte de synthèse de la valeur patrimoniale de chacun des habitats Carte de synthèse de la valeur patrimoniale des habitats marins du Sud-Ouest de l'île Maurice Draft Unpublished

29 juin 2018

 arvam@arvam.com, info@mncs-ngo.org, spersand@gmail.com, progeco@coi-loc.org, secretariat@coi-loc.org, alexandre.noel@coi-loc.org, julien.barde@ird.fr, 2018, "fr@Carte de synthèse de la valeur patrimoniale de chacun des habitats Carte de synthèse de la valeur patrimoniale des habitats marins du Sud-Ouest de l'île Maurice", <https://doi.org/10.5072/FK2:FPOENZ>, Entrepôt IRD, DRAFT VERSION

fr@La classification patrimoniale des habitats marins de la côte Sud-Ouest de l'île Maurice entre Le Morne et Flic-en-Flac. Elle a été réalisée lors de l'étude de faisabilité pour la mise en place d'une ou plusieurs aires marines protégées sur la côte sud-ouest de Maurice. La Mau...

fr@Etat de santé des récifs de la pente externe du Sud-Ouest de Maurice Draft Unpublished

29 juin 2018

 arvam@arvam.com, info@mncs-ngo.org, spersand@gmail.com, progeco@coi-loc.org, secretariat@coi-loc.org, julien.barde@ird.fr, 2018, "fr@Etat de santé des récifs de la pente externe du Sud-Ouest de Maurice", <https://doi.org/10.5072/FK2:TZGQDN>, Entrepôt IRD, DRAFT VERSION

fr@Ce jeu de données correspond à l'état de santé des récifs de la pente externe entre Le Morne et Flic-en-Flac à Maurice. Elle a été réalisée lors de l'étude de faisabilité pour la mise en place d'une ou plusieurs aires marines protégées sur la côte sud-ouest de Maurice. La Mau...

fr@Recouvrement relatif des habitats biotiques à fort intérêt patrimoniale du Sud Ouest de Maurice Draft Unpublished

29 juin 2018

 arvam@arvam.com, info@mncs-ngo.org, spersand@gmail.com, progeco@coi-loc.org, secretariat@coi-loc.org, julien.barde@ird.fr, 2018, "fr@Recouvrement relatif des habitats biotiques à fort intérêt patrimoniale du Sud Ouest de Maurice", <https://doi.org/10.5072/FK2:ZUGJNW>, Entrepôt IRD, DRAFT VERSION

fr@Le recouvrement relatif des habitats biotiques à fort intérêt patrimoniale des zones récifales côtières entre Le Morne et Flic-en-Flac à Maurice. Il a été réalisée lors de l'étude de faisabilité pour la mise en place d'une ou plusieurs aires marines protégées sur la côte sud-ou...

fr@Habitats marins du Sud-Ouest de l'île Maurice Draft Unpublished

29 juin 2018

 arvam@arvam.com, info@mncs-ngo.org, spersand@gmail.com, progeco@coi-loc.org, secretariat@coi-loc.org, alexandre.noel@coi-loc.org, julien.barde@ird.fr, 2018, "fr@Habitats marins du Sud-Ouest de l'île Maurice", <https://doi.org/10.5072/FK2:EBGIZB>, Entrepôt IRD, DRAFT VERSION

fr@La typologie des habitats marins de la côte Sud-Ouest de l'île Maurice entre Le Morne et Flic-en-Flac. Elle a été réalisée lors de l'étude de faisabilité pour la mise en place d'une ou plusieurs aires marines protégées sur la côte sud-ouest de Maurice. La Mauritius Marine Cons...

fr@Stations d'échantillonnage de la côte Sud-Ouest de l'île Maurice Draft Unpublished

29 juin 2018

# Interopérabilité : CKAN (open data US et UK)



Home / Organisations / BlueBRIDGE / Global Tuna Atlas / Annual catch of tuna and ...

Item Groups

## Annual catch of tuna and tuna-like species in the East Pacific Ocean (1918-2015) (IATTC level 0)

This dataset lists catch of tuna and tuna-like species from the Commission for the Inter-American Tropical Tuna Commission (IATTC) from 1918 to 2015. Catches are stratified by year, species, fishing gear, vessel flag reporting country. It corresponds to the IATTC nominal catch (task1) dataset.

This dataset was downloaded on the IATTC website. It was then restructured to be in line with the Sardara database structure (columns normalization, names, etc.). However the content of data and the code lists were kept exactly as their are available in the source dataset.

### Dataset extent

Map data © OpenStreetMap contributors  
Tiles by MapBox

### Tags

Belize Bermuda Canada Carangidae Cayman Is Chile China  
ChinaTaiwan Colombia Coryphaenidae Costa Rica Ecuador  
El Salvador Euthynnus lineatus Feature Catalog Res... Fisheries  
Fr Polynesia Gear nei Gear not known Gillnets and entang...  
Guatemala Handlines and hand... Harpoons Honduras Istiophoridae  
Istiophorus platypt... Japan Katsuwonus pelamis Korea Rep  
Longlines nei Makaira indica Makaira nigricans Mexico Netherlands  
Nicaragua Osteichthyes Other nei Panama Peru Purse seines  
Rajiformes Sarda spp Selachimorpha Pleur... Senegal Spain  
Tetrapturus angusti... Tetrapturus audax Thunnini Thunnus alalunga  
Thunnus albacares Thunnus obesus Thunnus orientalis Trolling lines  
Tuna Tuna atlas USA Unknown Vanuatu Venezuela  
Xinbias oladius detailed data nominal catch series timeseries

Annual catch of tuna and tuna-like species in the East Pacific Ocean (1918-2015) (IATTC level 0)



Show Graphic

Followers

0

Follow

Organisation



### Global Tuna Atlas

The Global Tuna Atlas VRE is a tool to integrate and analyze public domain data from various (Tuna) Regional Fisheries Management Organizations (T-RFMOs). Data include catches,...

[read more](#)

# SDI (Postgis, GN, GS) => bancs de thons



Données Partager Outils Le GBIF

JEU DE DONNÉES D'OCCURRENCES | DATE D'ENREGISTREMENT 22 FÉVRIER 2011

## ecoscope\_observation\_database

Publié par [IRD - Institute of Research for Development](#)  
Julien Barde

ENSEMBLE DE DONNÉES PROJET STATISTIQUES ACTIVITÉ TÉLÉCHARGER PAGE D'ACCUEIL

89 874 OCCURRENCES 19 CITATIONS

This dataset represent a part of the joint database of UMR212-EME (Exploited Marine Ecosystems) research unit (IRD, Ifremer, University of Montpellier 2 and related partner like YugNIRO: Southern Scientific Research of Marine Fisheries and Oceanography, Ukraine). This database is used to manage various dataseries collected for tropical tuna fisheries and research of tuna and associated species (so far mainly YugNIRO and IRD datasets). Biology of top predators in the open ocean ecosystem including distribution, size and weight parameters, trophic ecology, and sexual maturity are covered. In the forthcoming years more data will be available.

**Dernière modification des métadonnées:** 18 août 2016  
**Hébergé par:** IRD - Institute of Research for Development  
**Licence:** CC BY 4.0  
Comment citer [DOI](#) 10.15468/dz1kk0

89 874 Occurrences	97% Avec correspondance de taxon	100% Avec coordonnées	99.9% Avec l'année
-----------------------	-------------------------------------	--------------------------	-----------------------

89 874 ENREGISTREMENTS GÉORÉFÉRÉS

Généré le 11 à 8 heures © OpenStreetMap contributors, © OpenMapTiles, GBIF

N'importe quelle année 1834-2012

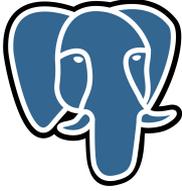
EXPLORER

# SDI (Postgis, GN, GS) => bancs de thons



A screenshot of a GBIF data record page. The page has a green header with navigation links: "Données", "Partager", "Outils", and "Le GBIF". On the right of the header are icons for zoom, search, and a "Connexion" button. Below the header is a secondary navigation bar with links: "ENSEMBLE DE DONNÉES", "PROJET", "STATISTIQUES", "ACTIVITÉ", "TÉLÉCHARGER", and "PAGE D'ACCUEIL". On the right side of this bar, there are two dark grey boxes: "89 874 OCCURRENCES" and "19 CITATIONS". The main content area is divided into two columns. The left column is a sidebar with a vertical list of menu items: "Description", "Couverture temporelle", "Couverture géographique", "Méthodologie", "Bibliographie", "Contacts", "Description des données", "Enregistrement dans le réseau GBIF", and "Citation". The right column contains the main text for each section. The "Description" section has a heading "Description" and a paragraph of text. The "Couverture temporelle" section has a heading "Couverture temporelle" and a bullet point indicating the date range "22 novembre 1961 - 15 novembre 2006". The "Couverture géographique" section has a heading "Couverture géographique" and a paragraph of text. Below the text is a world map with a blue rectangular box highlighting the tropical regions of the Indian and Atlantic Oceans. At the bottom left of the map are zoom in (+) and zoom out (-) buttons. At the bottom of the map is a small copyright notice: "© OpenStreetMap contributors, © OpenMapTiles, GBIF."

# SDI (Postgis, GN, GS) => marquages de thons



Catalogue de métadonnées de l'UMR MARBEC / OSU-OREME - UMR MARBEC, OSU-OREME - Mozilla Firefox

Catalogue de métadonnées de l'UMR MARBEC / OSU-OREME

thredds.oreme.org:8080/geonetwork/srv/fre/catalog.search#/metadata/db\_RTTP\_CTOI\_19115

Rechercher

Retour à la recherche

## Base de données RTTP CTOI

Mis à jour : il y a 9 mois

Le jeu de données Tagging de la CTOI (Commission Thonière de l'Océan Indien à Victoria aux Seychelles) correspond à des données de marquages et de recaptures (168164 marquages) collectées dans le cadre du projet RTTP (Regional Tuna Taggin Program) supervisé par la CTOI (final beneficiary of the project), mis en place par la Commission de l'Océan Indien (Authorizing officer) à l'aide d'un financement du Fond Européen de Développement (xème FED) de 14M€. Le projet s'est déroulé entre janvier 2005 et décembre 2009. Les marquages ont été réalisés entre XXX et XXX par du personnel qualifié, essentiellement avec des sarmacs Spaghetti (les marques de 14.5 cm de long ont été privilégiées) et les données de recaptures ont été collectées entre XXXX et fin 2016. Les espèces ciblées étaient Thunnus Albacarus (54684 marquages), Skipjack (78332 marquages) et Thunnus obesus (34569 marquages). Tous les poissons abimés ont été rejetés. La vitesse de marquage a été essentielle avec un maximum de 15 secondes hors de l'eau pour un individu marqué. On note aussi la présence de données sur des espèces inconnues (à préciser...). 22858 individus ont été marqués deux fois pour estimer le nombre de pertes de marques (16.6%). 5944 thons (pour les trois espèces) ont été marqués chimiquement (avec de l'OTC : oxytetracycline). En complément, un jeu de données biologiques a été collecté sur les individus restés sur le pont. Chaque thon marqué recapturé par le bateau de marquage a été laissé sur le pont. L'essentiel des prises a été effectué par des canneurs avec des appâts vivants. 27 % des données biologiques proviennent des sémiers contre 1% pour les palangiers. Ces données sont caractérisées par la période importante qui a séparé le marquage de la recapture (Time at liberty, TL), probablement due au choix des zones et des périodes de marquage, en dehors des principales activités de pêches (Tanzanie, Oman...). Des récompenses ont été données pour encourager les retours de marques par les pêcheurs (de 3 à 10 dollars selon le niveau de vie des pays concernés). Le marquage a fait face aux contraintes suivantes qui ont fait douter de sa réussite en début de projet: disponibilité d'appâts vivants pour les canneurs, localisation des câbles (les thons tropicaux étant très mobiles), les conditions météo agitées dans l'océan Indien, disponibilité des ports pour les rotations et ravitaillements (le nombre de ports étant limité), développement de la piraterie, éviter le marquage à proximité de la pêche à la seine pour ne pas avoir de recapture trop précoces, obtenir accès aux ZEE pour la pêche aux thons et aux eaux territoriales pour celles des appâts.

Les données de capture et de recapture sont gérées dans deux bases différentes (Access). Certaines des données collectées contenues dans cette base sont confidentielles (nom des bateaux...) et ne peuvent faire partie des données rendues publiques. Les données brutes ont fait l'objet d'un effort important de nettoyage (corrections documentées) qui a amélioré la qualité générale du jeu de données. Un code qualité a été attribué en fonction de la qualité estimée des données. Les données brutes ont cependant été conservées et restent accessibles. Ces données sont publiques sur demande argumentée faite à la CTOI. Plusieurs thèses ou applications sont associées à ces données. La base access contient, dans ses différentes tables, des données sur les activités des bateaux, la pêche à la canne, l'observation, les captures sur le pont, les données biologiques, les marques électroniques, les recaptures relâchées. Un manuel de marquage décrit en détail tous les données et codes associés.

Jeu de données de marquage: pour chaque thon marqué ont été reportés les observations suivantes: Numéro de la marque, espèce, longueur, poids, cradle, date et position ainsi que des codes de qualité si nécessaire.

Le jeu de données biologiques inclut les longueurs (FL, FLx, CL), le poids, le sexe, la maturité sexuelle, le poids des gonades, le poids du foie, le poids du contenu stomacal (pas pour tous les poissons sur le pont). La base contient également des informations sur les opérations de pêche pour collecter les appâts et les informations sur les captures. Des échantillons d'appâts ont été collectés, les espèces ont été identifiées et les individus mesurés. Seuls ont été conservés la moyenne, le min et le max des longueurs des appâts. Les données sur ces activités et observations ont été complétées par des données environnementales.

Tous les bancs ont été répertoriés ainsi que d'autres éléments associés à ces occurrences (cétacés, objets flottants, FADs, current lines, etc.).

Le jeu de données sur les recaptures est stocké dans une base de données ACCESS différente qui contient les observations suivantes: Numéro de la marque, espèce, longueur, poids, l'a recapturé, où (pays, bateau, conserverie, position) et quand (process, date, bateau, well) a eu lieu la recapture, engine, type de banc (objet, libre, mixed)... ainsi que des codes de qualité lors de la recapture.

**Aperçu**

**Étendue spatiale**

- Les données de localisation sont de qualités variables. Les données de marquage proviennent des positions GPS des bateaux pendant les campagnes de marquage scientifiques. Les données de recaptures sont, quant à elles, de qualités hétérogènes, GPS ou recalculées selon les informations disponibles lors de la recapture (débarquement...) qui nécessite parfois une approximation. La période de marquage des thons s'est déroulée entre XX et XX. Les recaptures continuent encore.

**Téléchargements et liens**

	Liste des documents référencés sur le projet RTTP par le Centre de Documentation de la COI. <small>Site Zotero du Centre de Documentation de la COI</small>	Ouvrir le lien
	Trajectories of scientific cruises <small>Cette donnée est publiée dans le service de visualisation (WMS) disponible à l'adresse <a href="http://vmeosocopest.mpi.ird.fr/wms/?couche=dbTaggingCTOI_cruises_trajectories">http://vmeosocopest.mpi.ird.fr/wms/?couche=dbTaggingCTOI_cruises_trajectories</a>.</small>	Visualiser
	OGC web map service to visualize the location of biological data sampling data (OGC WMS) <small>Cette donnée est publiée dans le service de visualisation (WMS) disponible à l'adresse <a href="http://mdst-macros.ird.fr:8080/constellation/WMS/wms/dbTaggingRTTP?couche=mdst:query_all_schools_biological_dataasampling">http://mdst-macros.ird.fr:8080/constellation/WMS/wms/dbTaggingRTTP?couche=mdst:query_all_schools_biological_dataasampling</a>.</small>	Visualiser
	Web Map Service (OGC WMS standard) to visualize the accurate location of tagged tuna <small>Cette donnée est publiée dans le service de visualisation (WMS) disponible à l'adresse <a href="http://mdst-macros.ird.fr:8080/constellation/WMS/wms/dbTaggingRTTP?couche=mdst:query_all_schools_tagged_individuals">http://mdst-macros.ird.fr:8080/constellation/WMS/wms/dbTaggingRTTP?couche=mdst:query_all_schools_tagged_individuals</a>.</small>	Visualiser

# SDI (mapping Thredds & GN) => télédétection



SDI Lab Catalogue    Rechercher    Visualiser    sign in    Français

Retour à la recherche    < Précédent    Suivant >

## Mozambique, daily Sea Surface Temperature (1 km MODIS SST image serie, composited Level-3 satellite data)

Mis à jour : il y a 4 mois

The Monitoring for Environment and Security in Africa (MESA) project utilized Earth Observation (EO) and satellite technologies to generate products and information services for socio-economic development of African people as well as sustainable environmental management. The main idea to better use of earth observation data for an improved management of marine and coastal areas.

### Téléchargements et liens

	Get more information on MESA project from MOI Website MESA project on MOI Website	Ouvrir le lien
	Thredds metadata sheet Thredds metadata sheet	Ouvrir le lien
	Visualize maps from WMS (Web Map Service) - see service/operation metadata for guidance to use query parameters Cette donnée est publiée dans le service de visualisation (WMS) disponible à l'adresse <a href="http://mdst-macrocs.ird.fr:8080/thredds/wms/BlueBridge/MOI/SST/SST_1km_daily/Mozambique_SST_1km_daily-MODIS.ncml?">http://mdst-macrocs.ird.fr:8080/thredds/wms/BlueBridge/MOI/SST/SST_1km_daily/Mozambique_SST_1km_daily-MODIS.ncml?</a> , couche sst.	Visualiser
	Visualize metadata HTML view in Geonetwork OGC metadata: HTML view in Geonetwork	Ouvrir le lien
	Visualize metadata XML from CSW (Catalogue Service for the Web) OGC metadata: XML from CSW	Ouvrir le lien

### À propos de cette ressource

Catégories

Mots-clés

- Projet MESA
- COI
- Sea Surface Temperature

Aperçu

# Thredds : GUIs de requête NetCDF/OPeNDAP



## OPeNDAP Dataset Access Form

Tested on Netscape 4.61 and Internet Explorer 5.00.

Action:

Data URL:

Global Attributes:

Variables:  **lat: Array of 32 bit Reals [lat = 0..1800]**

lat:

**lon: Array of 32 bit Reals [lon = 0..3599]**

lon:

**time: Array of 32 bit Integers [time = 0..2525]**

time:

**analysed\_sst: Grid**

time:  lat:  lon:

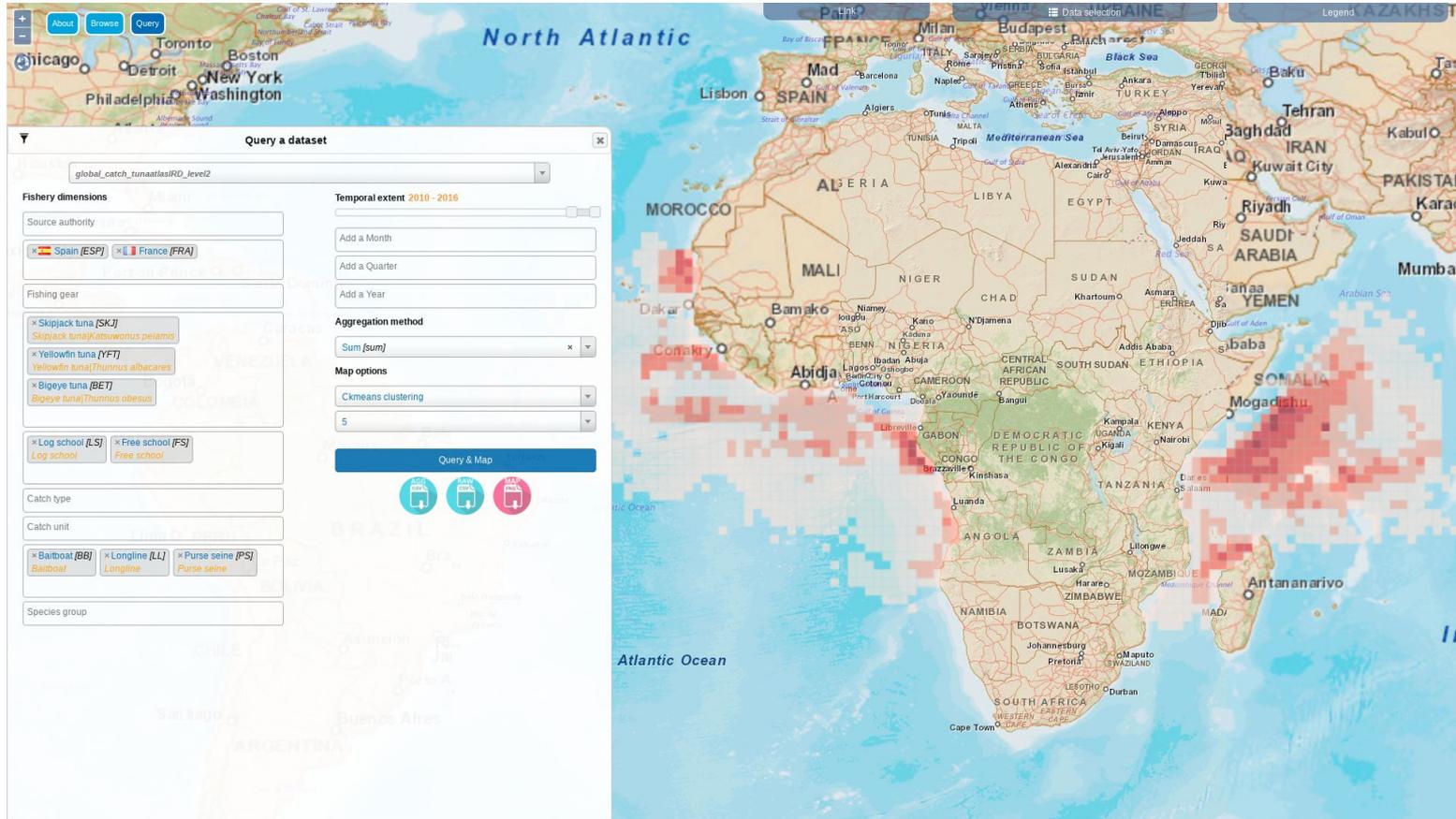
**analysis\_error: Grid**

time:  lat:  lon:

**mask: Grid**

time:  lat:  lon:

# Requêter la donnée : Tuna Atlas example



# Data paper = métadonnées

- Cherchez l'erreur : personne ne veut produire les métadonnées mais tout le monde veut faire des Data papers.
- Data paper = métadonnées littéraire avec un DOI pour le papier et un DOI pour la donnée (et ses métadonnées).
- R: export de la métadonnée en knitr / Rmarkdown (Latex, html..)

Exemple en [écologie marine](#), métadonnées et data paper

- Données de différents projets stockées dans une base de données
- 1 jeu de données = 1 requête SQL,
- Multiple requêtes SQL = multiple fiches de métadonnées
- 1 métadonnée = 0..1 data paper



# Conclusion

- “Choisir” les normes et automatiser le mapping entre ces normes:
  - DOIs => métadonnées et données (aussi pour les personnes, les protocoles..),
  - Dimension spatiale (généralement mal affichée par les viewers)
- Des métadonnées de base si possible directement insérée dans la donnée:
  - SQL => catalogues de requêtes directement stocké dans la base, dictionnaire de données ?
  - NetCDF => nativement intégrée dans le fichier
  - => travail en amont avec les producteurs (PGD..)
- Méthodes de production: GUIs et/ou programmation (R, Python, Java..):
  - Point essentiel pour les données régulièrement mises à jour,
- Quels outils pour la gestion et/ou la dissémination ?
  - Scripts ou serveurs classiques : pérennité ?
  - Infrastructures: Zenodo, Pangaea, GBIF, dataverse..

# Recommandations

- Pas de métadonnées, pas de DMP, pas de data paper...
- Faire participer les producteurs de données : inventaire des données à décrire dans un environnement d'édition collaboratif (eg Drive..)
- Bien décrire les éléments essentiels : mieux que la bounding box, mots clés..
- Contrôlez la valuation,
- SQL => pour chaque base de données fournir un dictionnaire de données et un catalogue de requêtes
- NetCDF => adopter les conventions CF
- DOIs et métadonnées synchronisées => GBIF.. et services institutionnels
- Valoriser la métadonnée par un Data paper.