



RÉPUBLIQUE  
FRANÇAISE

*Liberté  
Égalité  
Fraternité*



Geosciences pour une terre durable

**brgm**

# BRGM Metadata experience on the French Geocatalogue and Research data metadata (organization & Dataterra RI)

Sylvain Grellet

06/02/2025 - Joint EuroSDR and EuroGeographics Workshop  
Geometadata Production and Usage



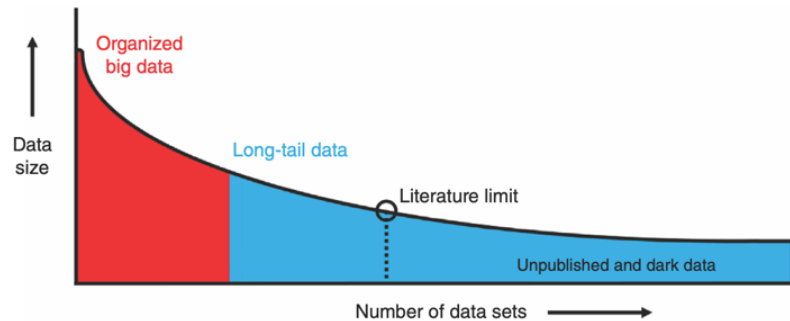


# Context

- Geocatalogue
  - French national reference geo catalogue (INSPIRE, HVD)
- Research data metadata
  - Longtail data => BRGM catalogue & repository
  - Data Terra French Research infrastructure, data centers

- The national portal for geo data
  - <https://www.geocatalogue.fr/fr>
  - Harvesting all the regional and domain nodes catalogues of France
  - Supporting INSPIRE reporting
  - As with all equivalent catalogues : curation, removing duplicates
  - Worked on indexing its content via crawlers : sitemap, JSON-LD + schema (mapping feedback to Geonetwork dev team)
- Core lesson (at least from me 😊 )  
after more than a decade on an ad'hoc solution  
=> too much technical debt, dependency on contractor/dev team  
=> back to a solution supported by a community and that last : Geonetwork

- BRGM metadata catalogue & research data repository
  - <https://data.geoscience.fr/entrepot/public/home>
- Targets to fulfill BRGM's
  - Open Data obligations => Re, INSPIRE, HVD, “classical” Geo data catalogue
  - AND Open Science obligations => Digital Object Identifiers (DOI) for data citation, support also the cataloguing of the long tail of research data



<https://www.nature.com/articles/nn.3838>

- Baselines
    - Semantics : ISO 19115:2014
    - Technical : Geonetwork, in a combo with a dedicated front/back, S3 storage
  
  - Lesson learned
    - Achieving both INSPIRE guidelines and Open Science (ticking FAIR principles) using 19115-2014 is an exercise by itself
- => Could we have EU guidelines on metadata updated to the latest ISO 19115 version ?

```
<mdb:metadataIdentifier>
  <mcc:MD_Identifier>
    <mcc:code>
      <gco:CharacterString>2fbf05ae-6692-4a98-a0ce-5bd443cfe33b</gco:CharacterString>
    </mcc:code>
    <mcc:codeSpace>
      <gco:CharacterString>urn:uuid</gco:CharacterString>
    </mcc:codeSpace>
  </mcc:MD_Identifier>
</mdb:metadataIdentifier>
```



- F1: (Meta) data are assigned globally unique and persistent identifiers

```
<mdb:metadataLinkage>
  <cit:CI_OnlineResource>
    <cit:linkage xsi:type="lan:PT_FreeText_PropertyType">
      <gco:CharacterString>https://data.geoscience.fr/metadataRecord/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b</gco:CharacterString>
    </cit:linkage>
    <lan:PT_FreeText>
      <lan:textGroup>
        <lan:LocalisedCharacterString locale="#EN">https://data.geoscience.fr/metadataRecord/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b</lan:LocalisedCharacterString>
      </lan:textGroup>
    </lan:PT_FreeText>
  </cit:linkage>
  <cit:description xsi:type="lan:PT_FreeText_PropertyType">
    <gco:CharacterString>Uniform Resource Identifier (URI)</gco:CharacterString>
    <lan:PT_FreeText>
      <lan:textGroup>
        <lan:LocalisedCharacterString locale="#EN">Uniform Resource Identifier (URI)</lan:LocalisedCharacterString>
      </lan:textGroup>
    </lan:PT_FreeText>
  </cit:description>
```



- F1: (Meta) data are assigned globally unique and persistent identifiers




```
<mdb:metadataLinkage>
<cit:CI_OnlineResource>
  <cit:linkage xsi:type="lan:PT_FreeText_PropertyType">
    <gco:CharacterString>https://data.geoscience.fr/metadataRecord/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b</gco:CharacterString>
    <lan:PT_FreeText>
      <lan:textGroup>
        <lan:LocalisedCharacterString locale="#EN">https://data.geoscience.fr/metadataRecord/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b</lan:LocalisedCharacterString>
      </lan:textGroup>
    </lan:PT_FreeText>
  </cit:linkage>
  <cit:description xsi:type="lan:PT_FreeText_PropertyType">
    <gco:CharacterString>Uniform Resource Identifier (URI)</gco:CharacterString>
    <lan:PT_FreeText>
      <lan:textGroup>
        <lan:LocalisedCharacterString locale="#EN">Uniform Resource Identifier (URI)</lan:LocalisedCharacterString>
      </lan:textGroup>
    </lan:PT_FreeText>
  </cit:description>
</mdb:metadataLinkage>
```



F1: (Meta) data are assigned globally unique and persistent identifiers

Lithium content of minerals from the Buntsandstein, Permian and Hercynien granites from the Upper Rh...

Voir plus

Citation [html](#) [text](#) [ris](#) [bibtex](#) 

“

LACH Philippe (BRGM), LEROUX  
(BRGM) (2025). Lithium content  
project.  
<https://doi.org/10.18144/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b>

 DOI: <https://doi.org/10.18144/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b>

```
<cit:identifier>
  <mcc:MD_Identifier>
    <mcc:code>
      <gco:Anchor xlink:href="https://doi.org/10.18144/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b">0.18144/2fbf05ae-6692-4a98-a0ce-5bd443cfe33b</gco:Anchor>
    </mcc:code>
    <mcc:codeSpace>
      <gco:CharacterString>doi.org</gco:CharacterString>
    </mcc:codeSpace>
    <mcc:description gco:nilReason="missing" xsi:type="lan:PT_FreeText_PropertyType">
      <gco:CharacterString>Digital Object Identifier (DOI)</gco:CharacterString>
    </mcc:description>
    <lan:PT_FreeText>
      <lan:textGroup>
        <lan:LocalisedCharacterString locale="#EN">Digital Object Identifier (DOI)</lan:LocalisedCharacterString>
      </lan:textGroup>
    </lan:PT_FreeText>
  </mcc:MD_Identifier>
</cit:identifier>
```

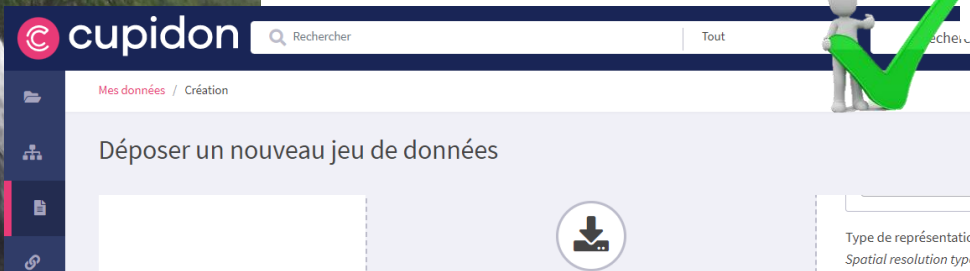


F1: data are assigned globally unique and persistent identifiers



F3 : Metadata clearly and explicitly include the identifier of the data they describe

Even if we can argue the fact that the DOI points to the landing page (thus metadataRecord) and not the resource itself it seems that in the research community that distinction is not (yet) made about the DOI



F2 : Data are described with rich metadata

```
<mdb:defaultLocale>
  <lan:PT_Locale id="FR">
    <lan:language>
      <lan:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="fre"/>
    </lan:language>
    <lan:characterEncoding>
      <lan:MD_CharacterSetCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_CharacterSetCode" codeListValue="utf8"/>
    </lan:characterEncoding>
  </lan:PT_Locale>
</mdb:defaultLocale>
<mdb:parentMetadata uuidref="2d213ad0-0ea4-4c9b-8058-2a7cd5fc9d1c"/>
<mdb:metadataScope>
  <mdb:MD_MetadataScope>
    <mdb:resourceScope>
      <mcc:MD_ScopeCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_ScopeCode" codeListValue="dataset"/>
    </mdb:resourceScope>
    <mdb:name gco:nilReason="missing" xsi:type="lan:PT_FreeText_PropertyType">
      <gco:CharacterString>Dataset</gco:CharacterString>
    </mdb:FreeText>
  </mdb:MD_MetadataScope>
</mdb:metadataScope>
```

ISO 19115







 **cupidon**  Tout

Mes données / Création

## Déposer un nouveau jeu de données



Taille limite par fichier: 1000 Mo

Aucun fichier choisi

Type de représentation  
Spatial resolution type  
Résolution spatiale / :

### R1.1: (Meta)data are released with a clear and accessible data usage license

 Date de création / *Creation*

 Date de révision / *Revision date*

 Contraintes d'utilisation / *Legal constraints* \*

Embargo sur la donnée / *Data embargo* \*

Droits d'accès / *Access rights*

Étendue temporelle / *Time extent* \*

Début / *Beginning* \*

Fin / *Ending* \*

Licence Ouverte 2.0 (Etalab)

Licence Ouverte 2.0 (Etalab)

CC BY 4.0

CC BY-SA 4.0

CC BY-NC 4.0

CC BY-NC-SA 4.0








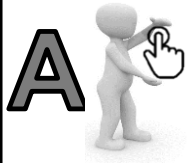














CC BY-ND 4.0

CC BY-NC-ND 4.0

 Traçabilité (origine) / *Origin* \*





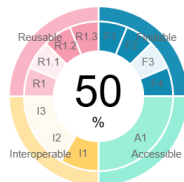
|  cupidon - VS FAIR Principles |  | Data  | Metadata  |
|--|--|---|---|
| <b>FINDABLE</b><br>         | F1 : (Meta) data are assigned globally unique and persistent identifiers                                       |   |     |
|  | F2 : Data are described with rich metadata   |   |    |
|  | F3 : Metadata clearly and explicitly include the identifier of the data they describe                          |   |    |
|  | F4 : (Meta)data are registered or indexed in a searchable resource   |   |    |
| <b>ACCESSIBLE</b><br>       | A1 : (Meta)data are retrievable by their identifier using a standardised communication protocol                |   |    |
|  | A1.1 : The protocol is open, free and universally implementable  |   |    |
|  | A1.2 :The protocol allows for an authentication and authorisation procedure where necessary                    |   |    |
|  | A2 : Metadata should be accessible even when the data is no longer available                                   |   |    |
| <b>INTEROPERABLE</b><br>    | I1 : (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation |   |    |
|  | I2 : (Meta)data use vocabularies that follow the FAIR principles   |   |    |
|  | I3 : (Meta)data include qualified references to other (meta)data   |   |    |
| <b>REUSABLE</b><br>        | R1 : (Meta)data are richly described with a plurality of accurate and relevant attributes                      |   |    |
|  | R1.1 : (Meta)data are released with a clear and accessible data usage license                                  |  |    |
|  | R1.2: (Meta)data are associated with detailed provenance   |   |    |
|  | R1.3 (Meta)data meet domain-relevant community standards   |   |  |



# cupidon - VS FAIR Principles

## Summary:

- Still, tools like F-UJI (<https://www.f-ujl.net/>) complain while the information is there ☹



|                | Score earned: | Fair level:           |
|----------------|---------------|-----------------------|
| Findable:      | 6 of 7        | <span>moderate</span> |
| Accessible:    | 1 of 3        | <span>initial</span>  |
| Interoperable: | 2 of 4        | <span>initial</span>  |
| Reusable:      | 3 of 10       | <span>initial</span>  |

FsF-F3-01M - Metadata includes the identifier of the data it describes.

FAIR level: 0 of 3  
Score: 0 of 1  
Output: {  
 "object\_identifier\_included": null,  
 "object\_content\_identifier\_included": []  
}

VS

```

</cit:date>
<cit:identifier>
  <moc:MD_Identifier>
    <moc:Code>
      <gdx:Anchor xlink:href="https://doi.org/10.57932/55b9f0fd-406b-4db8-98c5-128efb9833d9">10.57932/55b9f0fd-406b-4db8-98c5-128efb9833d9</gdx:Anchor>
    </moc:Code>
    <moc:codeSpace>
      <gco:CharacterString>doi.org</gco:CharacterString>
    </moc:codeSpace>
    <moc:description xsi:type="lan:PT_FreeText_PropertyType">
      <gco:CharacterString>Digital Object Identifier (DOI)</gco:CharacterString>
      <lan:PT_FreeText>
        <lan:textGroup>
          <lan:LocalisedCharacterString locale="#EN">Digital Object Identifier (DOI)</lan:LocalisedCharacterString>
        </lan:textGroup>
      </lan:PT_FreeText>
    </moc:description>
  </cit:identifier>
</cit:Citation>

```

FsF-R1.1-01M - Metadata includes license information under which data can be reused.

FAIR level: 0 of 3  
Score: 0 of 2  
Output: []

Metric tests:

| Test:          | Test name:   | Score: | Maturity: | Result: |
|----------------|--|--------|-----------|---------|
| FsF-R1.1-01M-1 | License information is given in an appropriate metadata element        | 0      |           | ?       |
| FsF-R1.1-01M-2 | Recognized licence is valid (community specific or registered at SPDX) | 0      |           | ?       |

VS

Debug messages:

| Level:  | Message:   |
|---------|--|
| WARNING | License information unavailable in metadata  |
| INFO    | Will consider all SPDX licenses as community specific licenses for FsF-R1.1-01M                    |
| WARNING | Skipping SPDX and community license verification since license information unavailable in metadata |

```

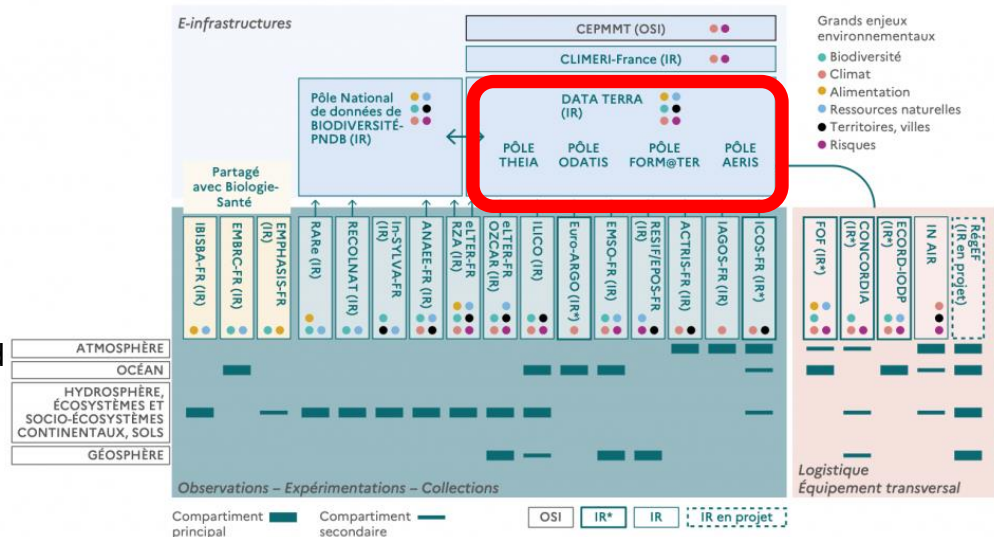
<mri:resourceConstraints>
  <moc:MD_LegalConstraints>
    <moc:accessConstraints>
      <moc:MD_RestrictionCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_RestrictionCode" codeListValue="unrestricted" />
    </moc:accessConstraints>
    <moc:useConstraints>
      <moc:MD_RestrictionCode codeList="http://standards.iso.org/iso/19115/resources/Codelists/cat/codelists.xml#MD_RestrictionCode" codeListValue="license" />
    </moc:useConstraints>
    <moc:otherConstraints xsi:type="lan:PT_FreeText_PropertyType">
      <gdx:Anchor xlink:href="https://spdx.org/licenses/CC-BY-NC-4.0.html">CC-BY-NC-4.0</gdx:Anchor>
      <lan:PT_FreeText>
        <lan:textGroup>
          <lan:LocalisedCharacterString locale="#EN">CC-BY-NC-4.0</lan:LocalisedCharacterString>
        </lan:textGroup>
        <lan:LocalisedCharacterString locale="#FR">CC-BY-NC-4.0</lan:LocalisedCharacterString>
      </lan:PT_FreeText>
    </moc:otherConstraints>
  </mri:resourceConstraints>

```



- The French Research e-Infrastructure focused on the Earth system and environment
  - <https://www.data-terra.org/en/>
  - Within the national roadmap for research infrastructures
  - Recently accepted as new EOSC national node (2025)

Organized around  
thematic  
reference centers



Research e-  
infrastructures

Research  
infrastructures

Figure 1 : OSI/IR\*/IR du domaine SST & ENV par grands types (observations – expérimentaux – et e-infrastructures), par grands compartiments du système Terre (atmosphère, océan, hydr, socio-écosystèmes continentaux – sols et géosphère) et par grands enjeux environnementaux (alimentation, ressources naturelles, territoires-villes, risques). En jaune figurent trois infrastructures avec le domaine Biologie – Santé.

- Those French Research Infrastructures are connected to their EU counterparts

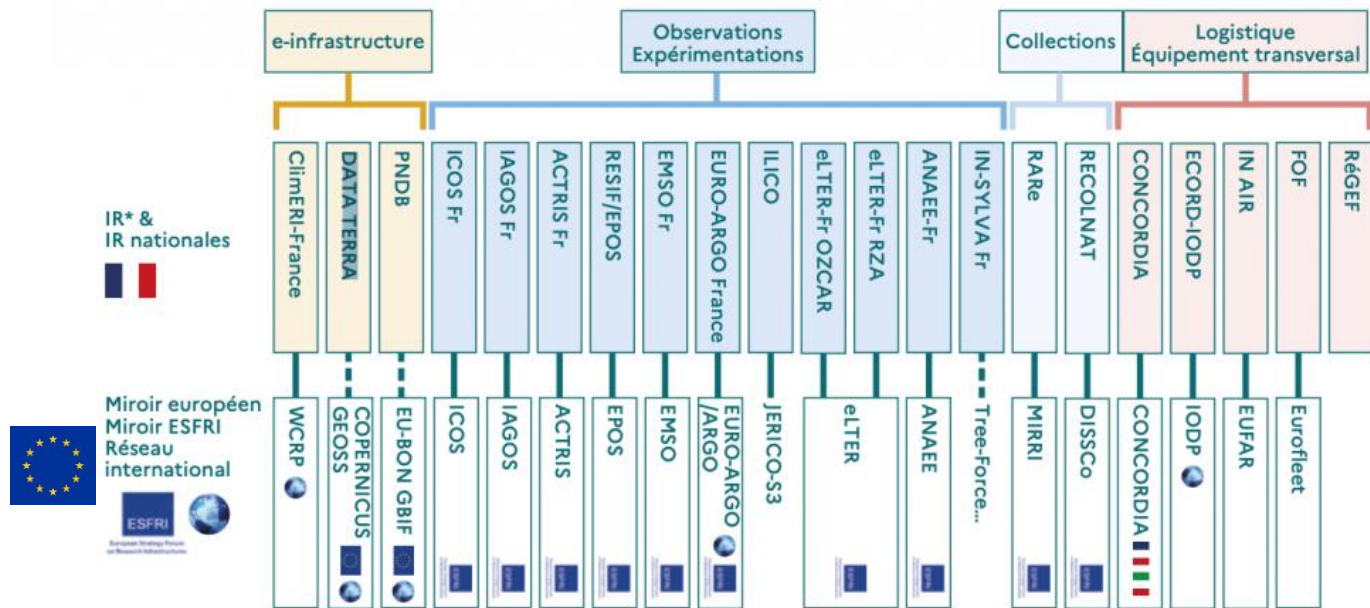


Figure 2 : Cohérence entre les infrastructures françaises et européennes ou internationales. Ne sont pas prises en compte ici les infrastructures communes avec le domaine Biologie-Santé ni l'OSI CEPMMT.





- GaiaData project : <https://www.gaia-data.org/>
- Helps DataTerra, the Research Infrastructures on biodiversity (PNDB) and Climate (Climeri) work together in describing their resources, feed Virtual Research Environments and share calculation capacity
- 8 years, 21+ organizations, 62 m€ project

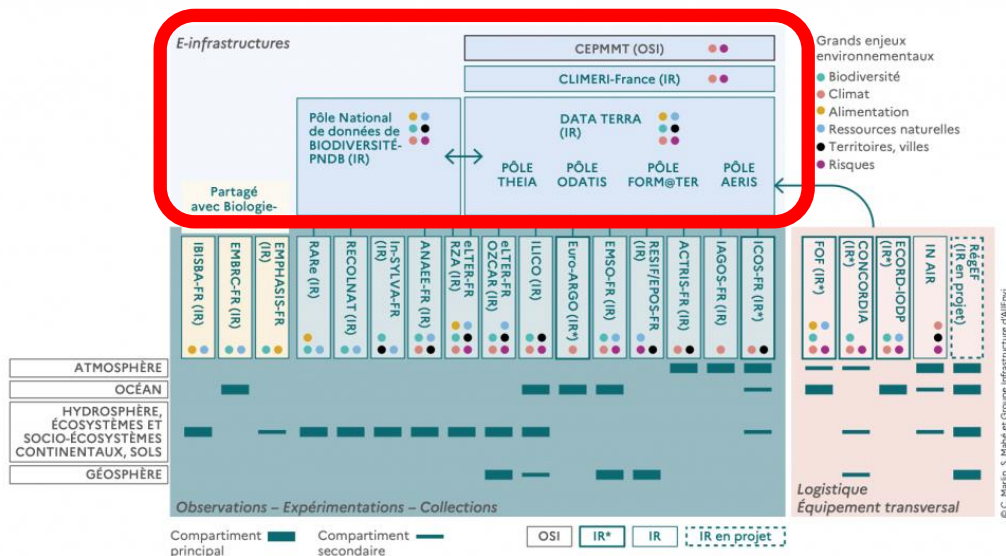


Figure 1: OSI/IR\*IR du domaine SST & ENV par grands types (observations – expérimentaux – collections, logistiques et e-infrastructures), par grands compartiments du système Terre (atmosphère, océan, hydrosphère-écosystèmes – socio-écosystèmes continentaux – sols et géosphère) et par grands enjeux environnementaux (biodiversité, climat, alimentation, ressources naturelles, territoires-villes, risques). En jaune figurent trois infrastructures partagées avec le domaine Biologie – Santé.



- Overview of the meta catalogue part



Various thematic reference centers running for decades for some  
=> Various practices regarding Metadata (métadonnées)

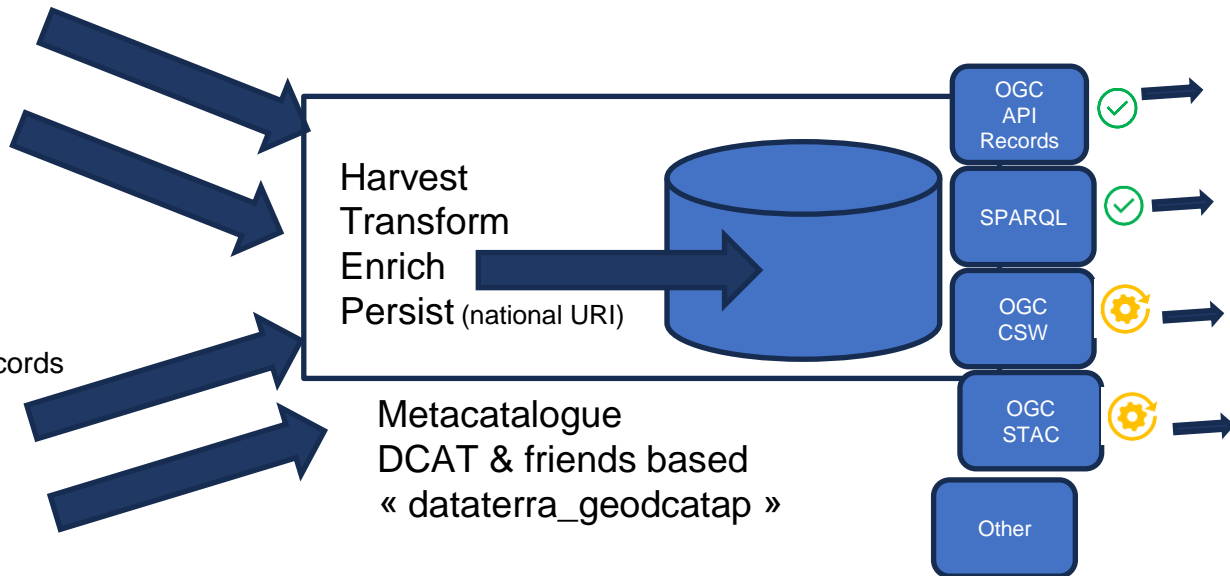
- Federating various communities with various practices/views on metadata

- Semantics

- 19115-2003, 19115-2014
    - DCAT
    - DCAT + others (ex : OBOE, SOSA)
    - Ad'hoc
    - EML
    - STAC
    - ...

- APIs

- CSW
    - OGC API Records
    - SPARQL
    - Ad'hoc
    - STAC
    - ....







## ● Atmosphere

### Données sol

- En général, découpage par **instrument** et **niveau** de données (contrôle qualité, traitement, etc.)  
⇒ Une fiche catalogue par **produit**
- Mais on peut aussi choisir de regrouper tous les instruments identiques d'un **réseau** (ex : Stations MF)
- Et si un instrument est déployé dans le cadre de **campagnes** de mesure, on découpe aussi par campagne
- Un instrument peut être un seul capteur ou un ensemble de capteur (ex : station météo) et donc mesurer 1 ou plusieurs **variables**
- Structuration classique : [observatoire /] plateforme / instrument / **produit**

P20A  
Pic du Midi  
Lannemezan (CRA)  
Weather Station  
Sky Imager  
Sky Imager - P20A-CRA station (LIA)  
Sky Imager - P20A-CRA station - ELIFAN Cloud Fraction - Hourly means (L2A)  
Radar VHF

- Pour les plateformes, on peut aussi parler de site ou de station
- Liens dans les métadonnées pour un **accès plus fin à la donnée**  
Ex : Serveur HTTP ou FTP, API DAALIA, STAC, Services Thredds, etc.
- Tous les produits disponibles dans les CDS ne sont pas forcément décrits dans le catalogue  
Ex : données brutes

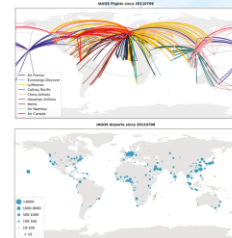


Various groupings  
- monitoring  
- Network  
- product  
- Campaign  
- variable

### Données aéroportées

Grouping by  
monitoring Campaign, sensor  
Notion of cruise

- Données
  - Campagnes utilisant les avions de recherche (principalement)
  - Instrumentation embarquée sur des avions commerciaux (IR IAGOS)
  - Ballons troposphériques et stratosphériques (HEMERA)
- Granularité des fiches dans le catalogue AERIS
  - Tous les vols d'une campagne et d'un paquet d'instrument (core, instruments scientifiques par PI)
  - Tous les vols IAGOS (en fait découpé en 2 fiches : croisière et profils verticaux lors des phases de montée/descente)
  - Vol ballon en général, mais parfois tous les vols d'une année
- Les fiches renvoient vers des portails dédiés qui permettent d'autres accès
  - SAFIRE+ : recherche et téléchargement par avion, campagne, vol, dataset (package instrumental pendant 1 campagne)
  - IAGOS : accès fin (vol, variable, etc.)



### Données spatiales

- Diversité : données issues de multiples satellites d'observation de l'atmosphère
  - Multiple paramètres : radiances/reflectance à diverses longueurs d'ondes, rétrodiffusion LIDAR, réflectivité RADAR, dépolariation...
  - Diverses géométries : données "griddées" (rares), géostationnaire, visée instrument, images, profils atmosphériques...
  - Granularité fichier dépendant de la mission : 1/2 orbites jour/nuits, orbite, sections de 5 minutes, horaire, journalier, mensuel, annuel...
  - Diversité de formats (hdf4/4, NetCDF, BUFR, GRIB, binaire propriétaire...)
- Vocabulaire
  - Un **produit** est un type de fichier généré par un algorithme de traitement (aka une chaîne de traitement)
  - Un produit est composé de **dataset** (variables) et de métadonnées.
  - Une **version** de produit est unique et correspond à la combinaison collection des fichiers d'entrée + version de l'algorithme
  - Une **collection** est une agrégation de versions dans le temps (avec périodes de recouvrement possible, mais pas toujours, exclusion des périodes avec des versions "non standard" de la collection)
- Structuration des données
  - Données rangées dans une arborescence UNIX  
 <capteur>/<produit>/<version>/<année>/<année-mois-jour>  
 Product > dataset  
 Product versions <=> Collections
  - Choix fait car permettant une navigation "logique" dans les répertoires depuis le cluster de calcul, compatible FTP et interface web FTP-like, opendap... => communément utilisé sur les sites de distribution de données satellite





- Land Surface, In situ component



## Structuration des données: modèle de données Theia/OZCAR

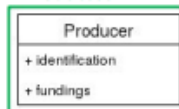


DJOUGOU - Metro station

**Observation: basé sur O&M**

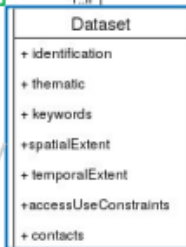
1 observation = 1 série temporelle d'1 variable à 1 station

### Producteur

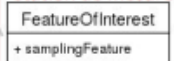


lieu de données : basé sur ISO 19115 /  
nspire / dataCite

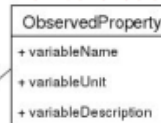
1 jeu de données = 1 ensemble  
d'observations, ie 1 ou plusieurs séries  
temporelles de différentes variables à  
différentes stations



### Station de mesure

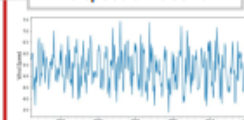


### Nom de variable

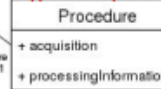


Wind speed

Wind speed time serie



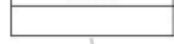
### Type de capteur



Wind monitor



### Fichier de données



Meteorological dataset (including radiative budget), within the  
Donga watershed (600 km2), Benin

2021-05-07T09:00:00.000Z;2.165  
2021-05-07T09:30:00.000Z;2.359  
....

Different views on  
what is the Dataset

Super heterogenous  
granularity

Just 3 examples

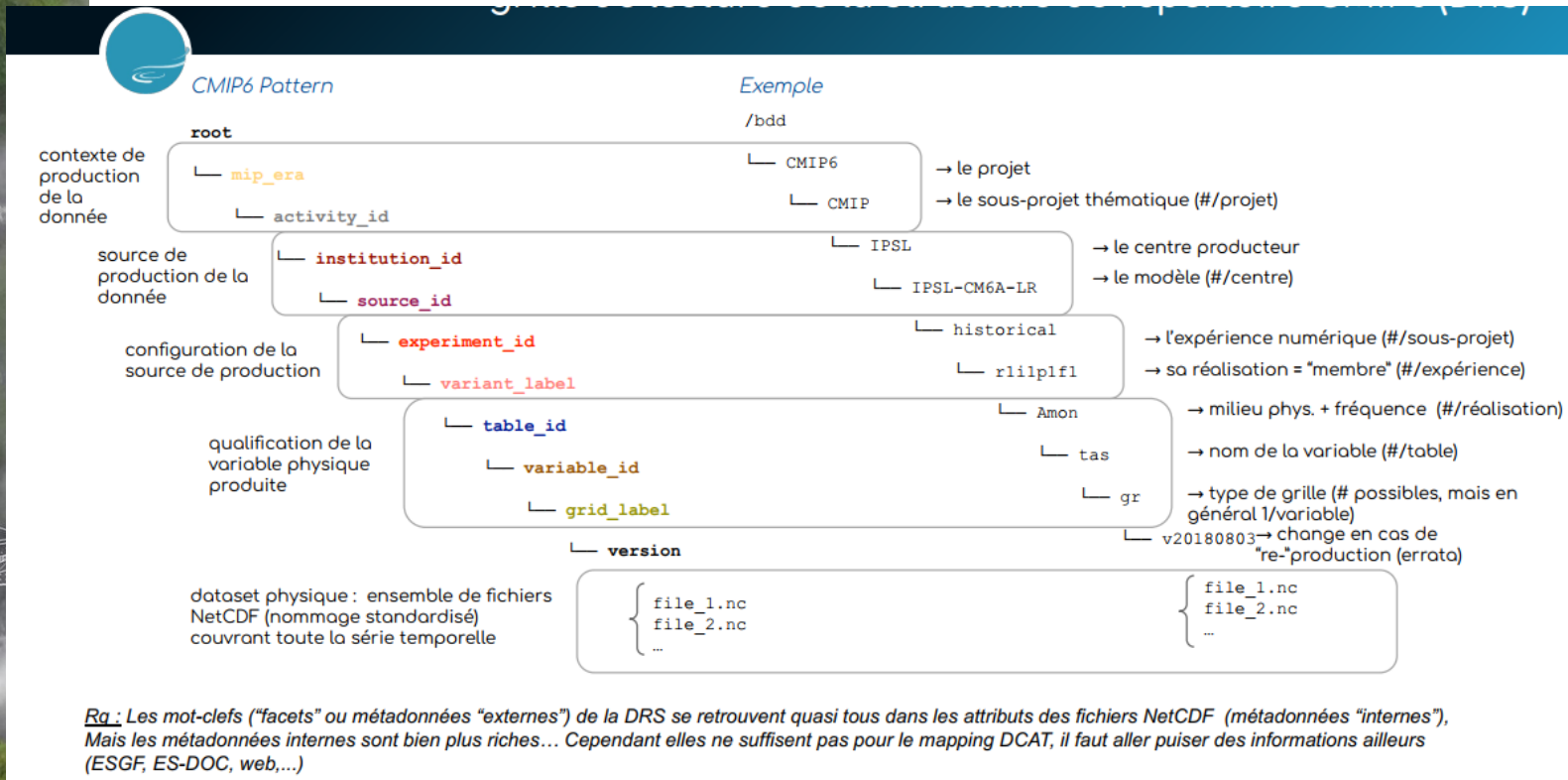


- CLIMERI : Climate modelling (GIEC CMIP)

Different views on  
what is the Dataset

Super heterogenous  
granularity

Just 3 examples







- Current status

- Maaaaany mappings done from reference centers <-> dataterra\_geodcatap profile
- Some already the subject of collaborative work : 19115-2003 / -2014 => DCAT or GeoDCAT

Note : we'll also need a "dataterra\_geodcatap" to 19115-2014 for CSW exposition

- Some not

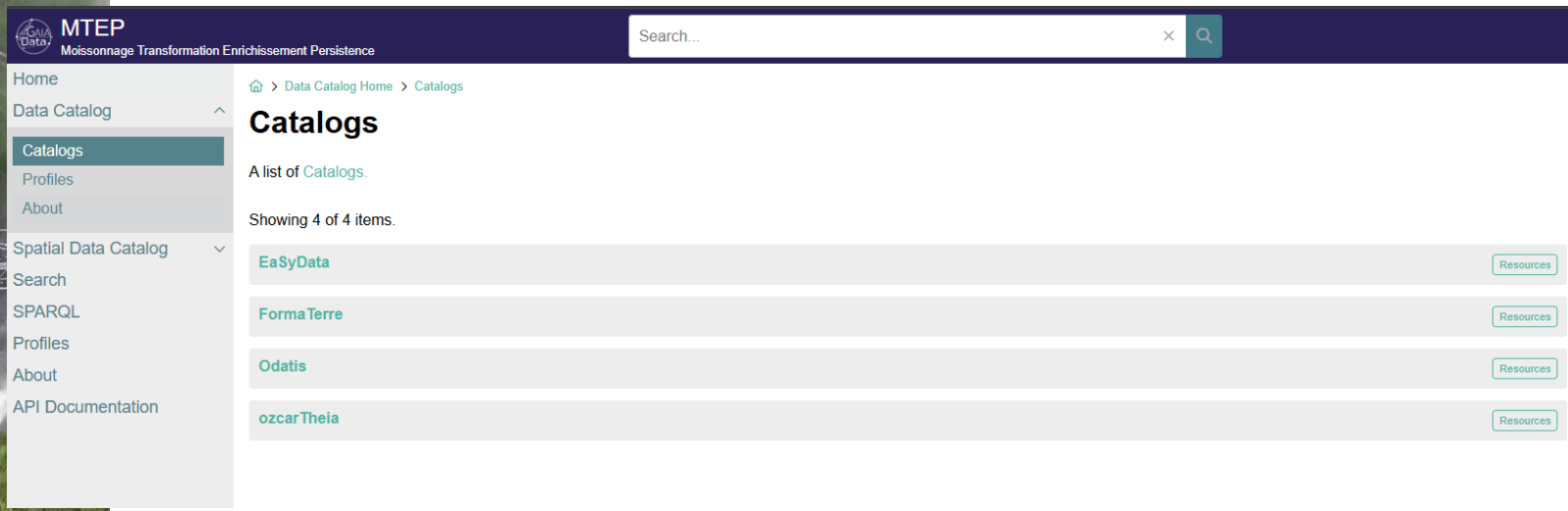
- STAC ⇔ DCAT : however several initiatives already identified (FAIR-Ease, FAIRiCUBE projects)
- EML (ecological metadata language) ⇔ DCAT

#### Class [dcatalog:CatalogRecord \(M\)](#)

|                           |  | Source  |  |   | Target  |   |
|---------------------------|--|---|--|---|---|---|
| Parent_source_element_url | Xpath  | Property_source_element_url (ENLEVER Colonne) | Range (set of valid property's values) | Example of value  | Property_dcat_url (V2.0)                              | Notes   |
| mdb_MD_Metadata           | /  | mdb:resourceScope                             | gmd:MD_ScopeCode                       | <codeList="http://standards.iso.org/iso/19115/resources/CodeLists/codelist.xml#MD_ScopeCode" codeListValue="dataset">   | foaf:primaryTopic (M)                                 | conditionnel pointer vers le dcat Datas   |
|                           | /mdb:MD_Metadata/mdb:dateInfo[1]/cit:CI_Date/cit:date/gco:Date                             | gmd:dateStamp                                 | gco:Date                               |   | revision -> dct.modified<br>publication -> dct.issued | dct:created feasible for catalogRecord  |
|                           | /mdb:MD_Metadata/mdb:metadataStandard/cit:CI_Citation/cit:title/gco:CharacterString        |   |  |   | dct:conformsTo (R)                                    | mettre en dur<br>https://www.w3.org/TR/vocab-dcat-2/  |
|                           | /mdb:MD_Metadata/mdb:metadataIdentifier/mcc:MD_Identifier/mcc:code/gco:CharacterString     | mcc:code                                      | gco:CharacterString                    | <mcc:code><br><gco:CharacterString>1b7ad219-023b-4544-b4eb-6e12340504df</gco:CharacterString> <!-- mapping DCAT : dcat:CatalogRecord/dct:identifier --><br></mcc:code><br><mcc:codeSpace><br><gco:CharacterString>urn:uuid:</gco:CharacterString><br></mcc:codeSpace> | dct:identifier (R)<br>dct:created (O)                 |   |
|                           | /mdb:MD_Metadata/mdb:defaultLocale/lan:PT_Locale/lan:language/lan:LanguageCode             | gmd:language                                  | gmd:LanguageCode                       | codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="fre">  | dct:language (O)                                      | dans dcat:CatalogRecord ?<br>n'existe pas dans dcat V2 => ajouter d<br>Voir si pas déjà dans le GaiaData_DC |
|                           | /mdb:MD_Metadata/mdb:metadataLinkage/cit:CI_OnlineResource/cit:linkage/gco:CharacterString |   | gco:CharacterString                    | <gco:CharacterString<br>xmlns:gco="http://standards.iso.org/iso/19115-3/gco/1.0">https://dev.easydata.earth/metadataRecord/46e3be28-0d46-448c-9baf-4626a61d7b44</gco:CharacterString>   | dcat:CatalogueRecord                                  | l'URI du catalogue Record   |
|                           |  |   |  |   |   |   |



- Current status
  - Core system running : <https://terra-vocabulary.org/portal/ld/>
  - tip of the iceberg based on PREZ (<https://github.com/RDFLib/prez>) as including this profile was not feasible in Geonetwork + also willing to go to various representations of the catalogue resources (mime types & profiles)
  - Minimum Viable Product mode : development made on identified datasets supporting scientific use cases (connexion to the entire catalogues will for sure trigger many logs when doing harvest, transform, enrich)



The screenshot shows the MTEP Data Catalog Home page. The header is dark blue with the MTEP logo and the text "Moissonnage Transformation Enrichissement Persistence". A search bar is on the right. The left sidebar contains a navigation menu with options: Home, Data Catalog, Catalogs (selected), Profiles, About, Spatial Data Catalog, Search, SPARQL, Profiles, About, and API Documentation. The main content area is titled "Catalogs" and shows a list of four catalog items: EaSyData, Forma Terre, Odatis, and ozcarTheia, each with a "Resources" button.

MTEP  
Moissonnage Transformation Enrichissement Persistence

Search...

Home  
Data Catalog  
Catalogs  
Profiles  
About  
Spatial Data Catalog  
Search  
SPARQL  
Profiles  
About  
API Documentation

> Data Catalog Home > Catalogs

## Catalogs


A list of Catalogs.

Showing 4 of 4 items.

|             |           |
|-------------|-----------|
| EaSyData    | Resources |
| Forma Terre | Resources |
| Odatis      | Resources |
| ozcarTheia  | Resources |



- Current status
  - Core system running, tip of the iceberg being based on PREZ

**MTEP**  
Moissonnage Transformation Enrichissement Persistence

Search...

File Home Com...

Start

[Home](#)  
[Data Catalog](#)  
[Spatial Data Catalog](#)  
**[Datasets](#)**  
[Conformance](#)  
[Profiles](#)  
[About](#)  
[Search](#)  
[SPARQL](#)  
[Profiles](#)  
[About](#)  
[API Documentation](#)

[Spatial Data Catalog Home](#) > [Datasets](#) > [Argo float data and metadata from Global Data Asse...](#)

## Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)

IRI <https://terra-vocabulary.org/catalogRecord/c15152c7-b331-409f-b702-0be977d36d4d>  
Type [dcat:CatalogRecord](#), [dcat:Dataset](#)

Argo is a global array of 3,000 free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean. This allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection. The array provides 100,000 temperature/salinity profiles and velocity measurements per year distributed over the global oceans at an average of 3-degree spacing. Some floats provide additional bio-geo parameters such as oxygen or chlorophyll. All data collected by Argo floats are publicly available in near real-time via the Global Data Assembly Centers (GDACs) in Brest (France) and Monterey (California) after an automated quality control (QC), and in scientifically quality controlled form, delayed mode data, via the GDACs within six months of collection.

|                                 |  |                                 |         |                            |                          |                  |  |                              |                         |  |  |
|---------------------------------|--|---------------------------------|---------|----------------------------|--------------------------|------------------|--|------------------------------|-------------------------|--|--|
| <b>Conforms To</b>              | ISO 19115:2003/19139 - SEXTANT   | <a href="#">xsd:string</a>      |         |                            |                          |                  |  |                              |                         |  |  |
| <b>dcat:contactPoint</b>        | <table><tr><td><a href="#">preferred label</a></td><td>Ifremer</td><td><a href="#">xsd:string</a></td></tr><tr><td><a href="#">rdf:type</a></td><td>ns2:Organization</td><td></td></tr><tr><td><a href="#">ns2:hasEmail</a></td><td>mailto:codac@ifremer.fr</td><td></td></tr></table> | <a href="#">preferred label</a> | Ifremer | <a href="#">xsd:string</a> | <a href="#">rdf:type</a> | ns2:Organization |  | <a href="#">ns2:hasEmail</a> | mailto:codac@ifremer.fr |  |  |
| <a href="#">preferred label</a> | Ifremer  | <a href="#">xsd:string</a>      |         |                            |                          |                  |  |                              |                         |  |  |
| <a href="#">rdf:type</a>        | ns2:Organization   |                                 |         |                            |                          |                  |  |                              |                         |  |  |
| <a href="#">ns2:hasEmail</a>    | mailto:codac@ifremer.fr  |                                 |         |                            |                          |                  |  |                              |                         |  |  |
| <b>foaf:primaryTopic</b>        | <a href="#">Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)</a>  |                                 |         |                            |                          |                  |  |                              |                         |  |  |

**Members**  
[Collections](#)

Search...

[Alternate Profiles](#)  
View alternate views & formats  
[DCAT](#) [current](#)  
[RDF/XML](#) [JSON-LD](#) [Turtle](#)  
[DCAT](#) [text/annot+turtle](#) [Turtle](#) [RDF/XML](#) [JSON-LD](#)  
[OGC API Features](#)  
[text/annot+turtle](#) [GeoJSON](#)  
[Alternates Profile](#)  
[text/annot+turtle](#) [Turtle](#) [JSON-LD](#) [RDF/XML](#)





- Current status

- Core system running, tip of the iceberg being based on PREZ

[Spatial Data Catalog Home](#) > [Datasets](#) > [Argo float data and metadata from Global Data Asse...](#)

## Argo float data and metadata from Global Data Assembly Centre (Argo GDAC)

[IRI https://terra-vocabulary.org/id/DataSet/af1420db-cee0-4152-bd63-47812cce80b](https://terra-vocabulary.org/id/DataSet/af1420db-cee0-4152-bd63-47812cce80b)

Type dcat:Dataset

Argo is a global array of 3,000 free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean. This allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, made publicly available within hours after collection. The array provides 100,000 temperature/salinity profiles and velocity measurements per year distributed over the global oceans at an average of 3-degree spacing. Some floats provide additional data for oxygen or chlorophyll. All data collected by Argo floats are publicly available in near real-time via the Global Data Assembly Centers (GDACs) in Brest (France) and Monterey (California) after an automated quality control (QC), and in scientifically validated mode data, via the GDACs within six months of collection.

|                    |                                    |   |   |
|--------------------|------------------------------------|---|---|
| Date Issued        | 2000-09-12                         |   |   |
| Spatial Coverage   | Spatial Coverage                   | rdf:type  | Location  |
|                    | dcat:bbox                          | <gml:Envelope srsName="http://www.opengis.net/def/crs/OGC/1.3/CRS84"><gml:lowerCorner>-90.00 -180.00</gml:lowerCorner><gml:upperCorner>90.00 180.00</gml:upperCorner></gml:Envelope> POLYGON((90.00 -180.00,90.00 180.00,-90.00 180.00,-90.00 -180.00,90.00 -180.00)) |   |
|                    | rdf:type                           | Concept   |   |
| Subject            | oceans                             |   |   |
| Temporal Coverage  | rdf:type                           | Period of Time  |   |
|                    | dcat:startDate                     | 1995-01-01  |   |
| dcat:distribution  | Description                        | Argo_ARGOLocations:Latest_Locations   |   |
|                    | rdf:type                           | dcat:Distribution   |   |
|                    | dcat:accessService                 | rdf:type  | dcat:DataService  |
|                    |                                    | dcat:endpointDescription  | Argo_ARGOLocations:Latest_Locations   |
|                    |                                    | dcat:endpointURL  | http://gis.jcommops.org/arcgis/services/Argo/ARGOLocations/MapServer/WFSServer? |
|                    | dcat:accessURL                     | http://gis.jcommops.org/arcgis/services/Argo/ARGOLocations/MapServer/WFSServer?   |   |
|                    | Description                        | FTP ARGO  |   |
| rdf:type           | dcat:Distribution                  |   |   |
| dcat:accessService | rdf:type                           | dcat:DataService  |   |
|                    | dcat:endpointDescription           | FTP ARGO  |   |
|                    | dcat:endpointURL                   | ftp://ftp.ifremer.fr/ifremer/argo/  |   |
| dcat:accessURL     | ftp://ftp.ifremer.fr/ifremer/argo/ |   |   |



- Core questions

- Where do we stop shoehorning observation level information into DCAT ? Ex : all observed properties in keywords ? Really ?
  - ⇒ Where do we start adding other ontologies ex : SOSA?
  - ⇒ But some reference centers already handle this but how do we do for those only running a 19115 based system (ex : Geonetwork) ? How do we extract more observation related info for those (need to traverse to the API exposing them through 'onlineResource')

- Content to allow proper machine 2 machine discovery and connection do dataservices
  - ⇒ Patterns to describe the URL, need to embark W3C:hydra for proper machine actionable API description (done with EU EPOS Research Infrastructure) ?

Could be generic for standard OGC webservices/APIs

- Document the transformation
  - ⇒ W3C:Prov





- Core questions

- How not to loose all that geo / scientific / metadata when harvest by the nation endpoint for research data (<https://recherche.data.gouv.fr/fr>) ?

- Running under Dataverse => Dublin core...

=> Different vision on what level of metadata is expected : quite a split !



Dublin Core

DCAT  
ISO 19115

DCAT + SSN/SOSA + PROV + Hydra  
ISO 19115 + OMS + etc...



- Core questions
  - Dynamic to embark the communities into a community of practice
    - ⇒ Pretty loose harvester rules with lots of information in the logs
  - Ex : 'keyword' are recommended, they are missing, that would be nice to add some
  - Where do we draw a line to identify datasets ?



CMIP6 Dataset Name (CLIMER):

esgpull search --distrib true project:CMIP6 -0 -f  
Found 63\_935\_043 files.

esgpull search --distrib true project:CMIP6 -0  
**Found 13\_261\_440 datasets**

Proposition Granularité GAIA :

CMIP6 Dataset Name (GAIA):

$\%(mip\_era)s.\%(experiment\_id)s.\%(table\_id)s.\%(variable\_id)s.\%$

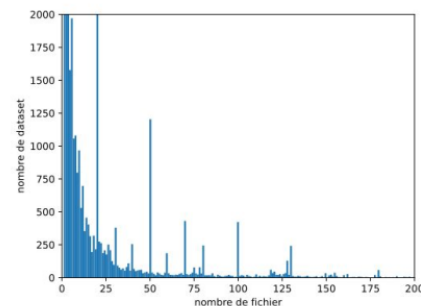
CMIP6.ssp585.Amon.tasmin

=> En cours d'analyse pour avoir le nombre exact (pour l'instant) une bonne moitié est quantifié

**environ 100\_000 Datasets**



Searching for the right grouping to avoid hiding the needles from the others in a big hay stack





# Thank you / merci

[s.grellet@brgm.fr](mailto:s.grellet@brgm.fr)

