

Pôle national de données de biodiversité, une infrastructure virtuelle au service de la recherche



anr®

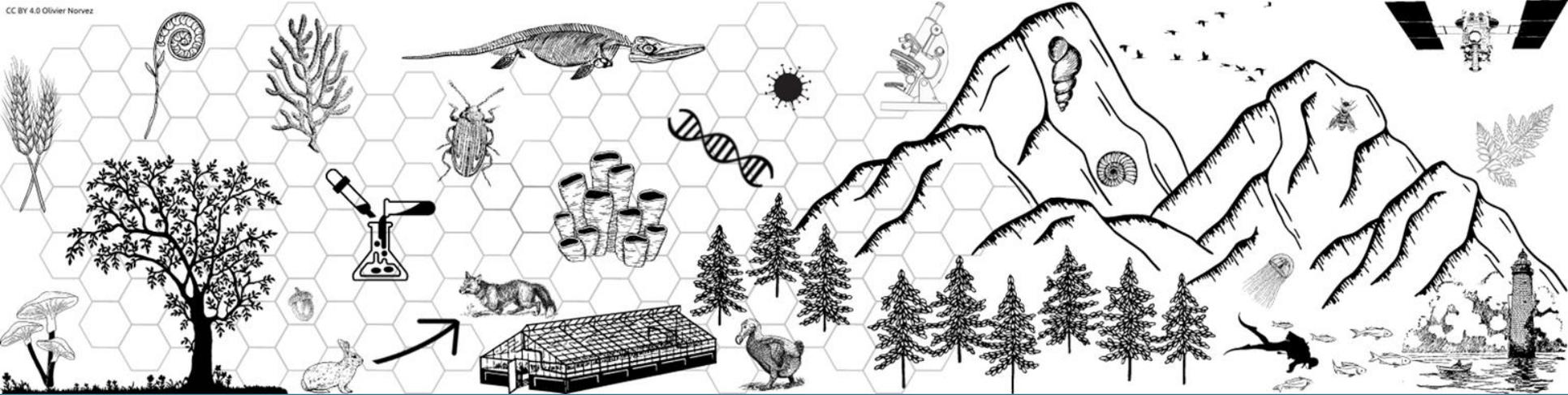


DATA TERRA



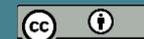
Pôle National de Données de Biodiversité

CC BY 4.0 Olivier Norvez



Yvan Le Bras (MNHN) *et al.*

@PNDBiodiv & www.pndb.fr



Le PNDB par les projets



#PNDB @PNDBiodiv



Equipex+ : e-infrastructure nationale intégrée biodiv / système terre / climat

Réseau implémentation principes FAIR en biodiv



Openmetapaper : dev application R Shiny MetaShARK & data paper



Outils et Formations e-learning via Galaxy / Carpentries



Partage données et algo IA Biodiv indicateurs de biodiversité

Implementation comittee French BON + EBV operationalizatyion pilot



Partenaires & CoPil PNDB

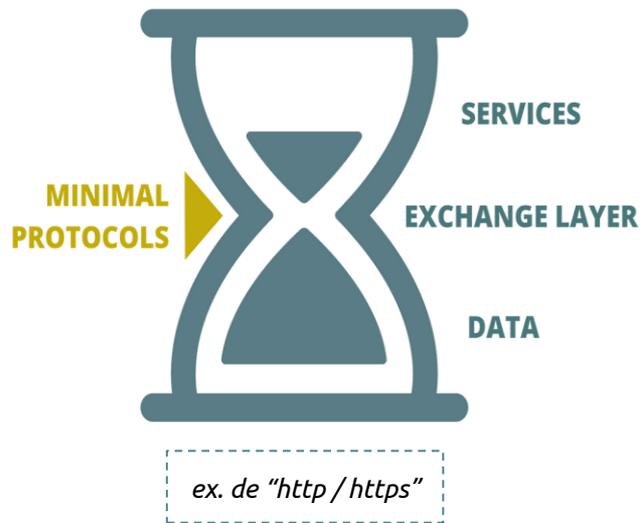


Introduction

Pourquoi Internet est une réussite ?

Introduction

Pourquoi Internet est une réussite ?



ex. de "http/https"

Protocole : ces règles vestimentaires que doit suivre la famille royale d'Angleterre

- Les tenues militaires sont obligatoires en cas d'obsèques. ...
- Lors d'un déplacement, la famille royale doit prévoir une tenue de rechange noire. ...
- Les petits garçons royaux doivent porter des shorts.



Si **protocole** pas respecté, pas de **communication** !

Communication entre machines / **logiciels**

Texte décrivant ce que doivent faire les 2 parties : **les spécifications**

Spécifications adoptées par un organisme qui les développe et les approuve

- ⇒ Ces spécifications deviennent des **normes / standards**
- ⇒ Consensus d'une partie des acteurs
- ⇒ Normes ouvertes préférables pour l'interopérabilité



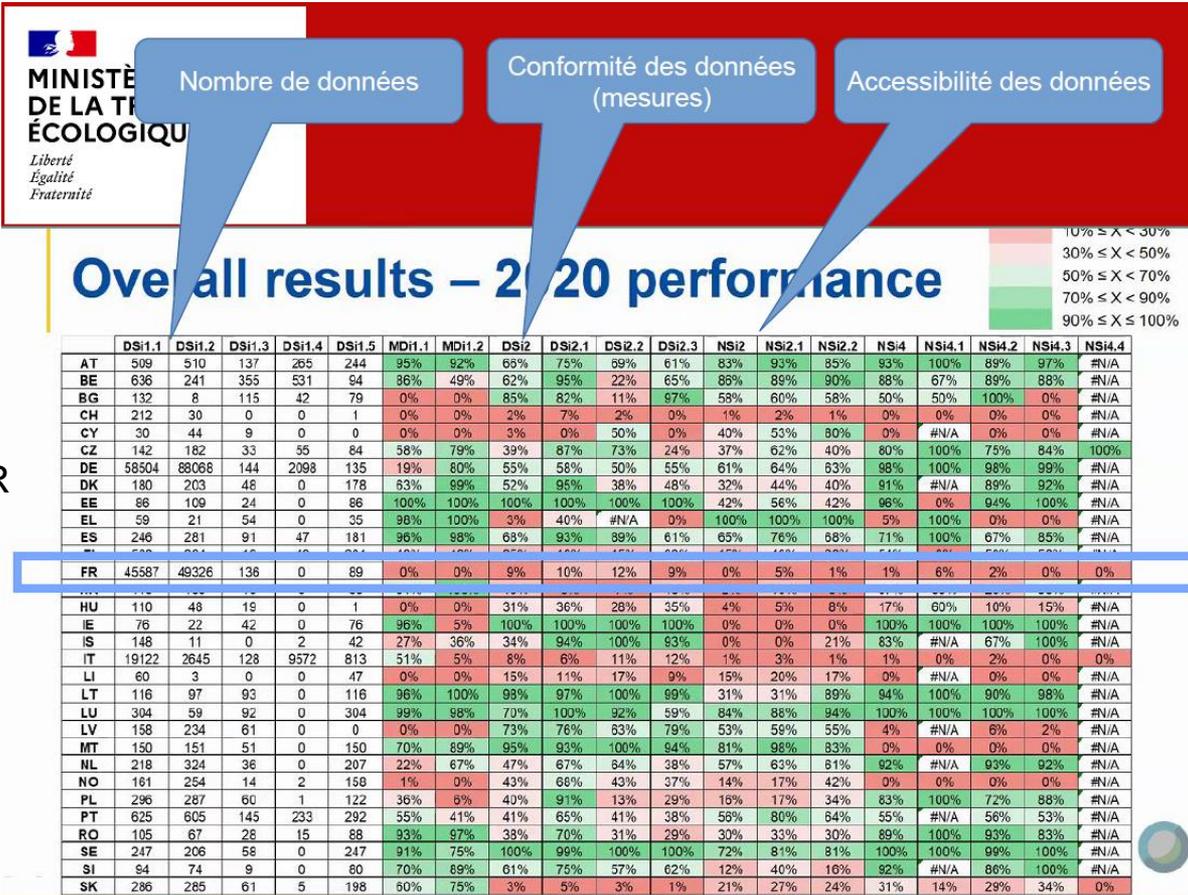
BiodiFAIRse

Pôle National
de Données de Biodiversité

D'INSPIRE à DCAT / GeoDCAT

Retours dans le cadre du groupe de travail sur les métadonnées du conseil national de l'information géolocalisée (CNIG)

<http://cnig.gouv.fr/metadonnees-r21413.html>



0% de compatibilité des métadonnées FR avec ISPIRE



D'INSPIRE à DCAT / GeoDCAT

About | Contact | Privacy policy | Legal notice | Cookies



INSPIRE GEOPORTAL

Enhancing access to European spatial data

European Commission > INSPIRE > Geoportal

Home | Priority Data Sets Viewer | Thematic Viewer | Harvesting status | Find out more about

Data sets by

Legislation: Priority Country: France

148 | 51 | 2

Filter results by:

Properties:

- Downloadable
- Viewable

Spatial scope coverage:

- National
- Regional
- Other

Reset

Title	Show dataset titles in:	Properties
Répartition des espèces rapportées sous la Directive Habitats Faune Flore en 2013 (DHFF 2013)	original MD language	
Répartition des espèces exotiques envahissantes dans le cadre du rapportage EEE en 2019 (EEE 2019)		
Masses d'eau cours d'eau - Réunion - Version Rapportage 2010		
Masses d'eau côtières - France entière - Version Rapportage 2010		
Masses d'eau souterraines - France entière - Version Rapportage 2010		

Mais pas 0 données !

D'INSPIRE à DCAT / GeoDCAT

« Complexité INSPIRE et ISO19115-ISO19139, mieux vaut qqchse de simple ! »

« Modèle INSPIRE rebute plus qu'il ne plait ! »

« Si on part en arrière sur du dublin core et titre / auteurs / résumés, on met tout un travail à la poubelle »



« Ce qu'il faut, c'est de la flexibilité ! »



« Il faut garder la place de pouvoir renseigner les métadonnées finement ! »

D'INSPIRE à DCAT / GeoDCAT

« Complexité INSPIRE et ISO19115-ISO19139, mieux vaut qqchse de simple ! »

« Modèle INSPIRE rebute plus qu'il ne plait ! »

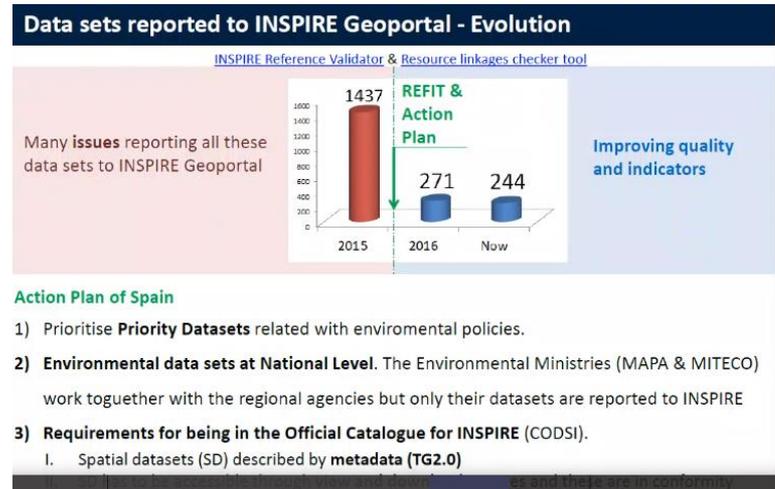
« Si on part en arrière sur du dublin core et titre / auteurs / résumés, on met tout un travail à la poubelle »

« Ce qu'il faut, c'est de la flexibilité ! »

« Il faut garder la place de pouvoir renseigner les métadonnées finement ! »

Vous ne connaissez pas l'EML ???? ;)

« Pour passer les tests, on diminue le nombre de champs de métadonnées »



On perd en qualité / degré de FAIRitude !?

D'INSPIRE à DCAT / GeoDCAT

« données spatiales sont pas spéciales »

-> hmmm alors que l'Europe a structurer par exemple les domaines comme l'environnement ont été mis dans la case données spatiales

« publier les données en tant que données ouvertes LOD / web sémantique »

->LOD ok mais pas direct, il vaut mieux un passage par un standard de MTD

D'INSPIRE à DCAT / GeoDCAT

« données spatiales sont pas spéciales »

-> hmmm alors que l'Europe a structurer par exemple les domaines comme l'environnement ont été mis dans la case données spatiales (ce qui est un problème je trouve)

« publier les données en tant que données ouvertes LOD / web sémantique »

->LOD ok mais pas direct, il vaut mieux un passage par un standard de MTD de domaine si existant / pertinent

D'INSPIRE à DCAT / GeoDCAT

=> intégrer des solutions logicielles qui utilisent les différentes normes

=> **outils et services FAIR**

=> définir une nouvelle norme

=> **NON !**

Cahier de l'Observatoire Data Publica : la standardisation des données ouvertes

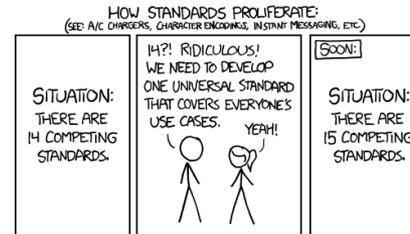
=> <https://observatoire.data-publica.eu/nos-publications>

=> publier les données en tant que données ouvertes LOD / web sémantique

=> **pas sans passer par de la métadonnées**

=> mapper les métadonnées d'un standard à un autre

=> **oui (et potentiellement en lien avec point intégration solutions logicielles)**



Vous ne connaissez pas l'EML ???? ;)

Interopérabilité : Générer et partager les [méta]données

Génération de données « en silos », partage, usage et intégration
=> nécessite un travail particuliers d'interopérabilité

Interopérable
Standards internationaux, ontologies, thésaurus, outils open source

Study A

Metadata (from EML)
Study A: White Mountains
Area column units: sq. meter
PIRU = *Picea rubens*
BEPA = *Betula papyrifera*

Date	Site	Species	Area	Count
10/1/1993	N654	PIRU	2	26
10/3/1994	N654	PIRU	2	29
10/1/1993	N654	BEPA	1	3

Study B

Metadata (from EML)
Study B: Green Mountains
Area sampled: 1 sq. meter
picrub = *Picea rubens*
betpap = *Betula papyrifera*

Date	Site	picrub	betpap
31 Oct 1993	1	13.5	1.6
14 Nov 1994	1	8.4	1.8

In order to combine the results from studies A and B, information contained in the metadata must be used to manipulate the data, since the two tables have different forms (schema). For example, the species columns from studies A and B use different structures and different annotations, but essentially describe the same organisms—a fact that would not be apparent without metadata.

Integrated Data

Study	Date	Site	Species	Density
A	10/1/1993	N654	<i>Picea rubens</i>	13.0
A	10/3/1994	N654	<i>Picea rubens</i>	14.5
A	10/1/1993	N654	<i>Betula papyrifera</i>	3.0
B	10/31/1993	1	<i>Picea rubens</i>	13.5
B	10/31/1993	1	<i>Betula papyrifera</i>	1.6
B	11/14/1994	1	<i>Picea rubens</i>	8.4
B	11/14/1994	1	<i>Betula papyrifera</i>	1.8

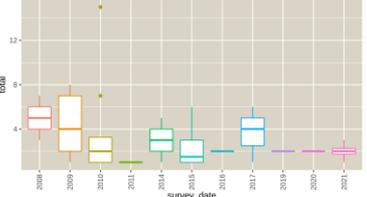
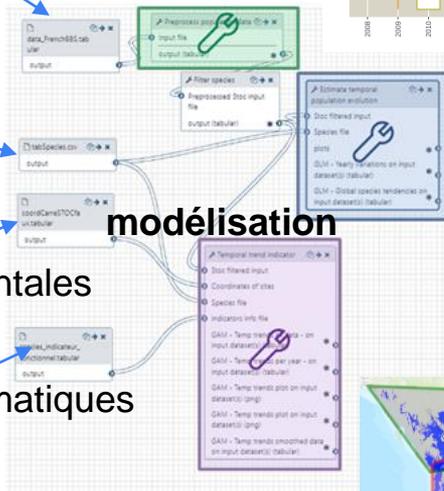
Metadata 'promoted' to become data
Format normalized using metadata
Species metadata from study B is now data (picrub/betpap column headings)
Density calculated using metadata

Données de biodiversité

Référentiel

Données environnementales

Scénarios climatiques



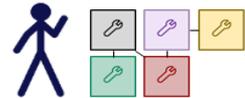
Interopérabilité : Générer et partager les codes source

Génération de codes sources, partage et intégration dans des workflows
 => nécessite un travail particuliers d'interopérabilité

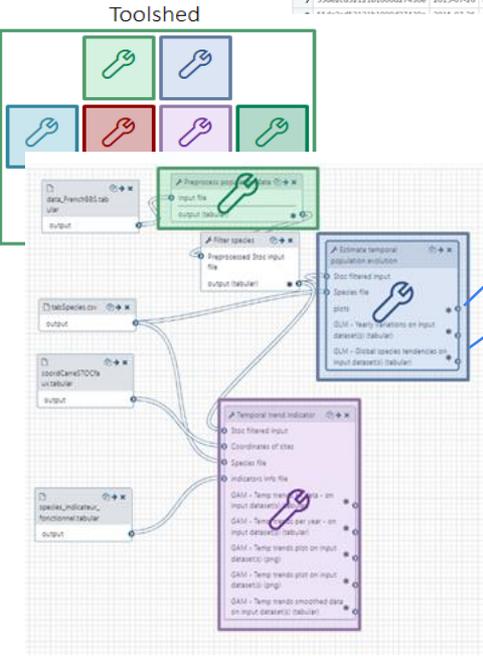
Interopérable
 Standards internationaux, conda, containers, Galaxy, cwl, RO-CRATE

carre	annee	espece	abond
1	2	2016	ACCGEN 0
2	2	2017	ACCGEN 0
3	2	2018	ACCGEN 0
4	2	2019	ACCGEN 0
5	183	2016	ACCGEN 0
6	183	2017	ACCGEN 0

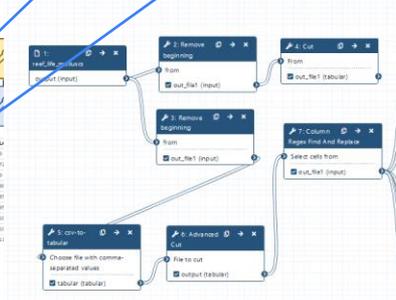
participation	huit	num_micro	groupe	espece	nb_contacts
1	556e2cf5212181000d27430e	2015-07-26	0	bat	Barbar 1
2	556e2cf5212181000d27430e	2015-07-26	0	boehcricket	Barfis 1
3	556e2cf5212181000d27430e	2015-07-26	0	noise	noise 5022
4	556e2cf5212181000d27430e	2015-07-26	0	boehcricket	Ducalls 5
5	556e2cf5212181000d27430e	2015-07-26	0	boehcricket	lylil 1B
6	556e2cf5212181000d27430e	2015-07-26	0	bat	Nyctel 1
7	556e2cf5212181000d27430e	2015-07-26	0	boehcricket	Rhanan 269



codeSp	rotation	codeSp	sexe	taille	classe_taille	poids	nb_jeu
1	AS140155	3	Hemiteles	-999	P	-999	1
2	AS140155	1	Nanosp.	-999	P	-999	3
3	AS140155	3	Gompvri	-999	P	-999	1
4	AS140160	3	Gompvri	-999	P	-999	1



Workflow constructed from history "TEST for RLS for biodiversity data exploration tutorial"



Workflow analytique de Galaxy pour les codes sources



Interopérabilité : Données et métadonnées, les bases !

"Les données de recherche sont définies comme des enregistrements factuels utilisés comme sources primaires pour la recherche scientifique, et qui sont généralement acceptés dans la communauté scientifique comme nécessaires pour valider les résultats de la recherche."

OCDE, 2007. <https://www.oecd.org/sti/inno/38500813.pdf>

OECD Principles and Guidelines for Access to Research Data from Public Funding



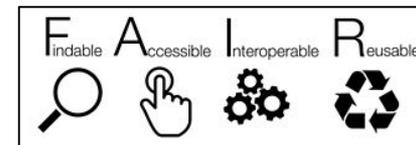
OECD

Qui ?
Quoi ?
Où ?
Quand ?
Comment ?

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip F. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, ... Barend Mons [✉](#) [+ Show authors](#)

Scientific Data 3, Article number: 160018 (2016) | [Cite this article](#)



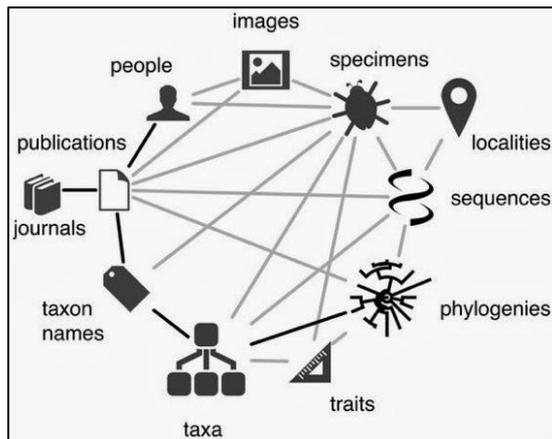
The FAIR Guiding Principles => Métadonnées est LA composante !

"Les métadonnées, que l'on peut définir simplement comme « des données sur les données », sont un moyen de nommer les choses et de représenter les données et leurs relations."

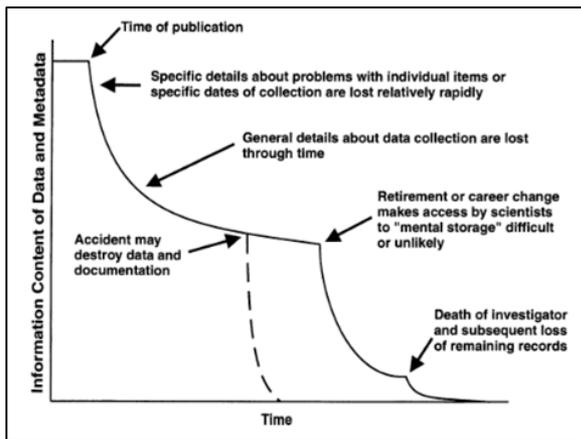
Christine L. Borgman, 2020. <https://books.openedition.org/oepp/14692>



Interopérabilité : Le cas des pôles de données, focus sur PNDB



R. Page, 2016 : doi: 10.3897/rio.2.e8767



W.K. Michener et al., 1997 doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]

Une diversité “d’objets” à relier entre eux
(Hétérogénéité dans les types de données, leur provenance et les standards utilisés)

Perte d’informations dans le temps

reproductibilité

Forum

Ecological Data Should Not Be So Hard to Find and Reuse

Timothée Poisot¹✉, Anne Bruneau^{1,2}, Andrew Gonzalez³✉, Dominique Gravel⁴, Pedro Peres-Neto⁵

Towards a more reproducible ecology

Michael Krabbe Borregaard and Edmund M. Hart

M. K. Borregaard (mkrborregaard@nmn.ku.dk), Unit of Copenhagen, Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, Copenhagen, Denmark. – E. M. Hart, Unit of Vermont, Dept of Biology, Marsh Life Science Building, Burlington, VT, USA.

Metaresearch for Evaluating Reproducibility in Ecology and Evolution

Fiona Fidler, Yung En Chee, Bonnie C. Wintle, Mark A. Burgman, Michael A. McCarthy, Ascelin Gordon

BioScience, Volume 67, Issue 3, March 2017, Pages 282–289, <https://doi.org/10.1093/biosci/biw159>

Published: 13 January 2017

Interopérabilité : Le cas des pôles de données, focus sur PNDB

Standardiser par la métadonnée permet la description fine, l'inférence, l'identification et l'interopérabilité des données

Study A

Metadata (from EML)				
Study A: White Mountains Area column units: sq. meter PIRU = <i>Picea rubens</i> BEPA = <i>Betula papyrifera</i>				
Date	Site	Species	Area	Count
10/1/1993	N654	PIRU	2	26
10/3/1994	N654	PIRU	2	29
10/1/1993	N654	BEPA	1	3

Study B

Metadata (from EML)			
Study B: Green Mountains Area sampled: 1 sq. meter picrub = <i>Picea rubens</i> betpap = <i>Betula papyrifera</i>			
Date	Site	picrub	betpap
31 Oct 1993	1	13.5	1.6
14 Nov 1994	1	8.4	1.8

In order to combine the results from studies A and B, information contained in the metadata must be used to manipulate the data, since the two tables have different forms (schema). For example, the species columns from studies A and B use different structures and different annotations, but essentially describe the same organisms—a fact that would not be apparent without metadata.

Integrated Data

Study	Date	Site	Species	Density
A	10/1/1993	N654	<i>Picea rubens</i>	13.0
A	10/3/1994	N654	<i>Picea rubens</i>	14.5
A	10/1/1993	N654	<i>Betula papyrifera</i>	3.0
B	10/31/1993	1	<i>Picea rubens</i>	13.5
B	10/31/1993	1	<i>Betula papyrifera</i>	1.6
B	11/14/1994	1	<i>Picea rubens</i>	8.4
B	11/14/1994	1	<i>Betula papyrifera</i>	1.8

Metadata 'promoted' to become data

Format normalized using metadata

Species metadata from study B is now data (picrub/betpap column headings)

Density calculated using metadata

Interopérable

Standards internationaux, ontologies, thésaurus, outils open source



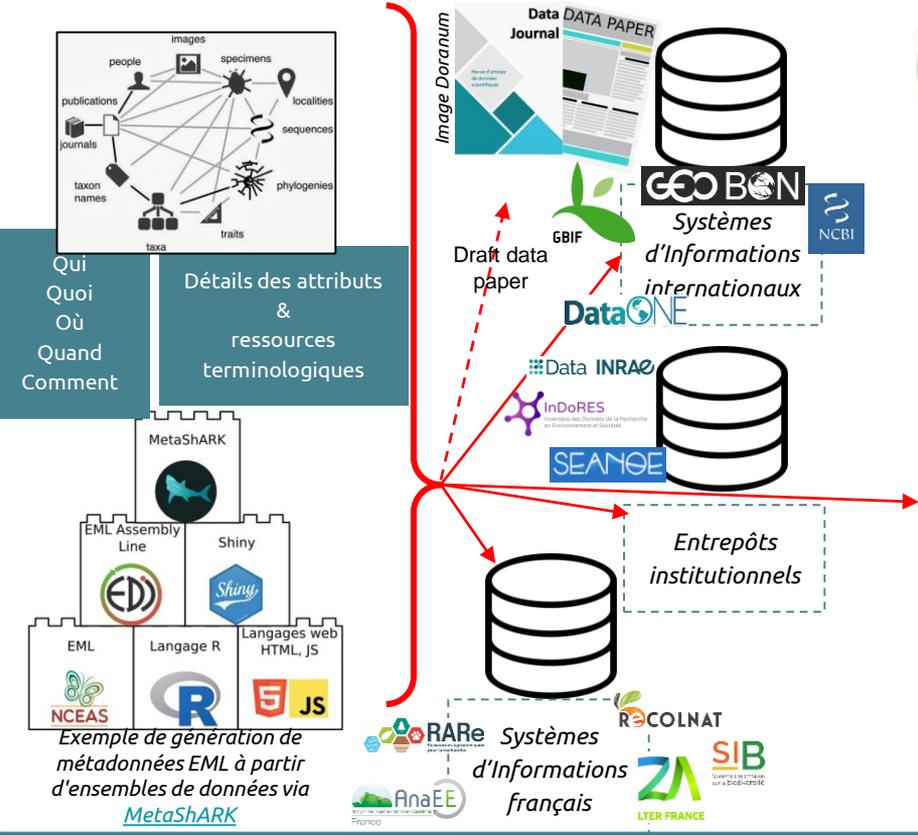
ex. de "http/https"

L'Ecological Metadata Language - EML- est un standard pivot mondialement reconnu et qui a plus de 25 ans de retours d'expériences par les écologues / écoinformaticiens

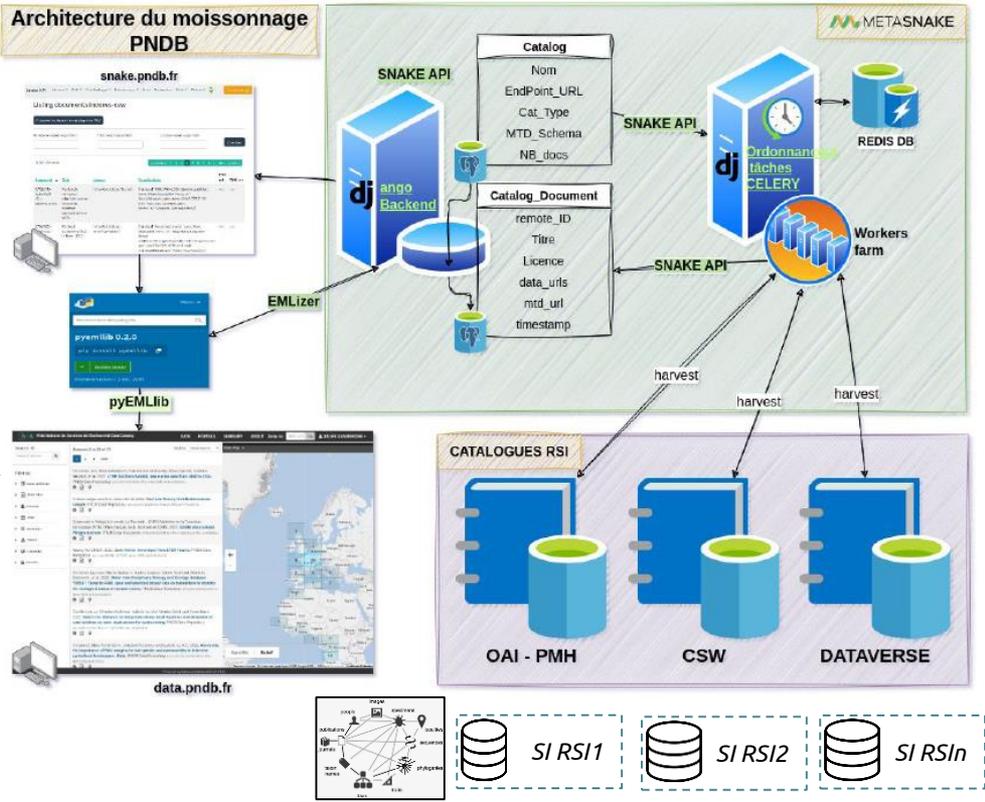


Interopérabilité : Le cas des pôles de données, focus sur PNDB

Approche jeu de données par jeu de données

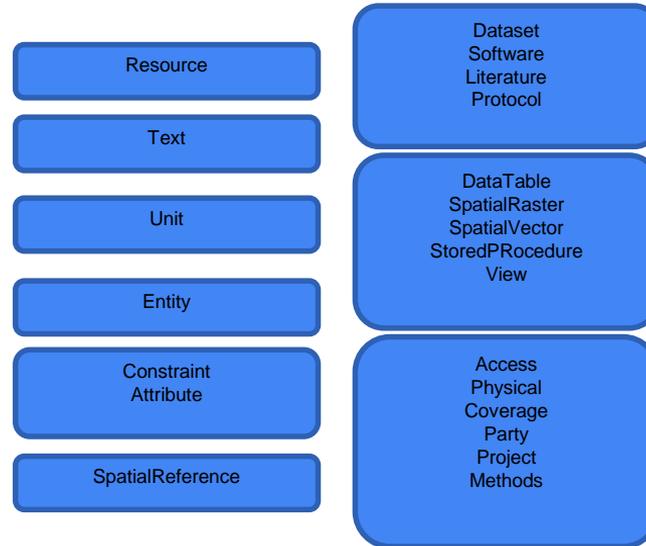


Approche lots de données (par SI)



EML in one slide!

EML -> Ecological Metadata Language !



PNDB « data & metadata » proposal

- **Focus on attributes / variables**

From creating EML: <https://ropensci.github.io/EML/articles/creating-EML.html>

We will build this EML file from the bottom up, starting with the two main components of a `dataTable` indicated above: the `attributeList` and the `physical` file type. We will then slip these two pieces into place inside a `dataTable` element, and slip that into our `eml` element along with the rest of the generic metadata, much like building a puzzle or nesting a set of Russian dolls.

The original metadata file was created in association with the publication in PNAS based on a Microsoft Word document template that Harvard Forest provides to the academic researchers. Metadata from this template is then read off by hand and an EML file is generated using a combination of a commercial XML editing platform (Oxygen) for commonly used higher-level elements, and the Java platform `Morpho` provided by the EML development team for lower level attribute metadata.

Attribute Metadata

A fundamental part of EML metadata is a description of the attributes (usually columns) of a text file (usually a csv file) containing the data being described. This is the heart of many EML files.

```
attributes <-
tibble::tribble(
  ~attributeName, ~attributeDefinition, ~formatString, ~definition,
  "run.num", "which run number (=block). Range: 1 - 6. (integer)", NA, "which run numbe",
  "year", "year, 2012", "YYYY", NA,
  "day", "Julian day. Range: 170 - 209.", "DDD", NA,
  "hour.min", "hour and minute of observation. Range 1 - 2400 (integer)", "hhmm", NA,
  "i.flag", "is variable Real, Interpolated or Bad (character/factor)", NA, NA,
  "variable", "what variable being measured in what treatment (character/factor).", NA, NA,
  "value.i", "value of measured variable for run.num on year/day/hour.min.", NA, NA,
  "length", "length of the species in meters (dummy example of numeric data)", NA, NA,
```

PNDB « data & metadata » proposal

- Focus on attributes / variables

ISO19115/ISO19139
Métadonnées données
géographiques



ISO19110
Métadonnées catalogue
d'attributs



Extension Information:
Name: Data_Field_Name
Short_Name: dfieldnm
Type: text
Domain: free text
Parent: Data_Field
Optionality: mandatory
Definition:



FGDC Biological Data
Profile (BDP)
Notion d'attributs

Name of the data field. This name should be the same as an Attribute Label documented in Section 5.1.2 Attribute (within Detailed Description which is itself within Entity and Attribute Information). The definition, domain, units

Why would I want to use EML when FGDC now supports biological data through the CSDGM?

EML is modular and extensible.

The Content Standard for Digital Geospatial Metadata (CSDGM) developed by the Federal Geographic Data Committee (FGDC) is a monolithic standard, and so it is difficult to mix and match parts of it with other standards – mainly because of all of the spatial requirements. So, we built EML as a series of modules that can be linked together and can be linked to other metadata standards. This gives us the most flexibility, and given that we can easily translate into CSDGM compliant documents, there is little cost.

EML
Taxonomic/Temporal/Geographic
Entities>attributs

Attribute Metadata

A fundamental part of EML metadata is a description of the attributes (usually columns) of a text file (usually a csv file) containing the data being described **This is the heart of many EML files.**

```
attributes <-  
tibble::tribble(  
  ~attributeName, ~attributeDefinition, ~formatString, ~definition,  
  "run.num", "which run number (=block). Range: 1 - 6. (integer)", NA, "which run numbe  
  "year", "year, 2012", "YYYY", NA,  
  "day", "Julian day. Range: 170 - 209.", "DDD", NA,  
  "hour.min", "hour and minute of observation. Range 1 - 2400 (integer)", "hhmm", NA,  
  "i.flag", "is variable Real, Interpolated or Bad (character/factor)", NA, NA,  
  "variable", "what variable being measured in what treatment (character/factor).", NA, NA,  
  "value.i", "value of measured variable for run.num on year/day/hour.min.", NA, NA,  
  "length", "length of the species in meters (dummy example of numeric data)", NA, NA,
```

Pour permettre la reproductibilité

NONGEOSPATIAL METADATA FOR THE ECOLOGICAL SCIENCES, Michener *et al.* 1997

330

4 sources

4 outputs

Provenance tracking!

Clic on Download to reproduce the treatment !

Data Table

Other Entity

Entity Name: DataDownload.R

Download

Description: R script used to download PAH.csv, Alkane.csv, Sample.csv, and CollectionMethods.csv from the original MS Access database.

Data Object Type: Other

Physical Structure Description:

Object Name: DataDownload.R

Size: 1846 byte

Externally Defined Format: Format Name: application/R

Online Distribution Info: <https://cn.dataone.org/cn/v2/resolve/urn:uuid:c2f47e88-cc2b-47db-a105-101b80e80334>

Source Data

PAH.csv

Citation: Mark Carls. 2017. Analysis of hydrocarbons following the Exxon Valdez oil spill, Gulf of Alaska, 1989 - 2014. Gulf of Alaska Data Portal. urn:uuid:32494e9c-afa3-4d55-875e-0f7928a4c171

View >

This data was generated by <> DataDownload.R

This data prov_hasDerivations the data you are currently viewing. Total_Aromatic_Alkanes_PWS.csv and the data you are currently viewing. Total_Aromatic_Alkanes_PWS.csv

This data was used by <> Total_PAH_and_Alkanes_GoA_Hydrocarbons

This data was used as an input to create the data you are currently viewing. Total_Aromatic_Alkanes_PWS.csv and the data you are currently viewing. Total_Aromatic_Alkanes_PWS.csv

Records: 12142

Un exemple sur portail
PNDB

PNDB Pôle National de Données de Biodiversité Data Catalog DATA PORTALS SUMMARY ABOUT Jump to: DOIx or IIF Go SIGN IN

Home / Search / Metadata

Lorraine Coché, Elie Arnaud, Bouveret Laurent, Romain David, Eric Foulquier, et al. 2021. **Kakila database of marine mammal observation data around the French archipelago of Guadeloupe in the AGOA sanctuary - French Antilles**. urn:node:PNDB. doi:10.48502/8bb5-pk85

Downloads 0 Citations 0 Views 0 Copy Citation Analyze Assessment report

Files in this dataset Package: urn:uuid:be50a625-e8a7-4cd5-b3ba-9ab591d41aa2

Name	File type		
Metadata: Kakila database of marine mammal observation data in the AGOA sanctuary - French Antilles.xml	EML v2.2.0		
BDD_Kakila_v2_20210420_observateur.tsv	text/csv	11 KB	Download
BDD_Kakila_v2_20210420_observation.tsv	text/csv	463 KB	Download
BDD_Kakila_v2_20210420_sortie.tsv	text/csv	217 KB	Download

[Show 5 more items in this data set](#)

General

Annotations

- is about biodiversity
- is about marine mammals
- is about geographical distribution
- is about Species distribution
- is about Sea regions
- is about Ecological stocktaking
- is about Natural area
- is about Landscape
- is about Ecosystem
- is about Animal ecology

Identifier doi:10.48502/8bb5-pk85

Un exemple sur portail
PNDB

PNDB Pôle National de Données de Biodiversité Data Catalog DATA PORTALS SUMMARY ABOUT Jump to: DOIx or IIT Go SIGN IN

Home / Search / Metadata

Lorraine Coché, Elie Arnaud, Bouveret Laurent, Romain David, Eric Foulquier, et al. 2021. **Kakila database of marine mammal observation data around the French archipelago of Guadeloupe in the AGOA sanctuary - French Antilles.** urn:node:PNDB. doi:10.48502/8bb5-pk85

Downloads 0 Citations 0 Views 0 Copy Citation Analyze Assessment report

Files in this dataset Package: urn:uuid:be50a625-e8a7-4cd5-b3ba-9ab591d41aa2

Name	File type	
Metadata: Kakila database of marine mammal observation data in the AGOA sanctuary - French Antilles.xml	EML v2.2.0	
BDD_Kakila_v2_20210420_observateur.tsv	text/csv	11 KB Download
BDD_Kakila_v2_20210420_observation.tsv	text/csv	463 KB Download
BDD_Kakila_v2_20210420_sortie.tsv	text/csv	217 KB Download

Show 5 more items in this data set

PNDB Pôle National de Données de Biodiversité Data Catalog DATA PORTALS SUMMARY ABOUT Jump to: DOIx or IIT Go SIGN IN

Number Of Records 4703

Attribute Information

Variables

- code_observation
- code_sortie
- code_observateur
- code_secteur_geog
- latitude
- longitude
- profondeur
- heure_observation
- code_taxon
- nombre_minimum
- nombre_maximum

Name: code_observation

Annotations: is a occurrenceID

Label:

Definition: Observation code combining the code_sortie and an observation number

Storage Type: string

Measurement Type: nominal

Measurement Domain: Definition Observation code combining the code_sortie and an observation number

Un exemple sur portail PNDB

Level III:
publishable and
auditable

X
X
X
X
X

Provenance



 Pôle National de Données de Biodiversité TEST Data Catalog

DATA PORTALS SUMMARY ABOUT Jump to: DOIx or IIF Go  YVAN LE BRAS ▾

Entity Name

[Download](#)  Data Object Type: image/png

4 sources   

Other Entity

Entity Name

[Download](#)  Data Object Type: application/octet-stream

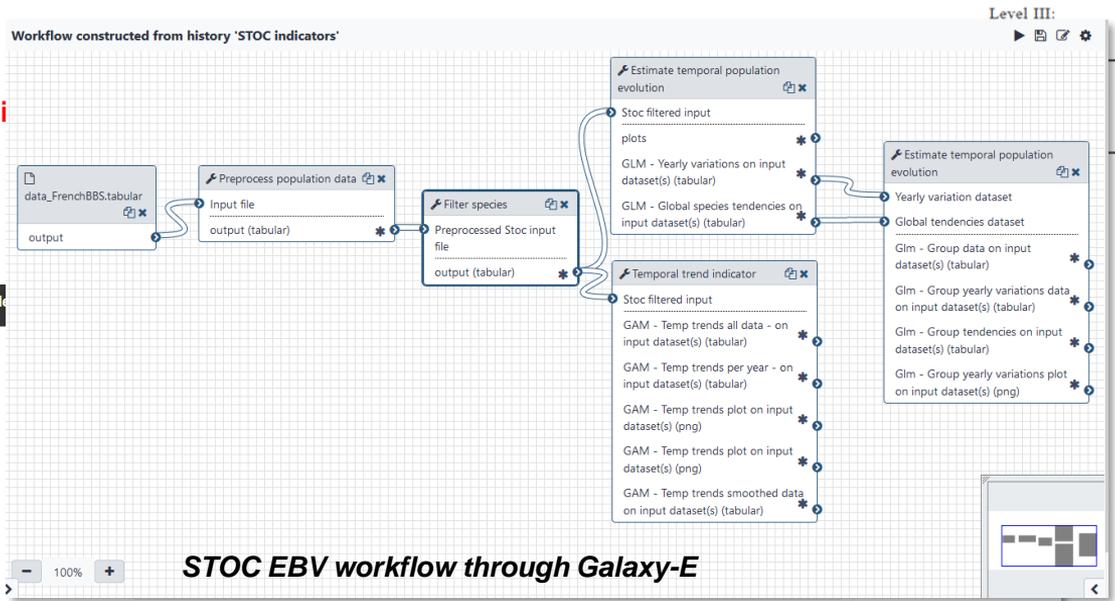
[view more](#) 

6 derivations   

[view more](#) 

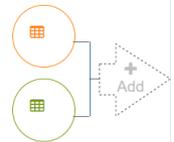
Un exemple sur portail PND

 Pôle National de Données de



Download  Data Object Type: image/png

4 sources



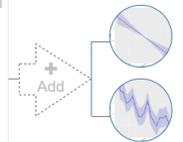
[view more](#)

Other Entity

Entity Name

Download  Data Object Type: application/octet-stream

6 derivations



[view more](#)

Amazing EML

Language + Modules -> flexibility!

Provenance tracking!

Data Table

Data: <https://pasta.lternet.edu/package/table/m1/knb-lter-and/4541/3/43232065a859b5ec150596b4af4a0158>

Name: HF02001

Identifier: 1

Description: Multi-scale tracer tests: Dye tracer test concentration breakthrough curves

Number of Records: 0

Number of Columns: 14

Time Period

Begin: 2003-06-14

End: 2003-06-21

Constraint:

Name:	PRIMARY
Primary Key:	HF02001.DATANUM
Key:	HF02001.SITENAME
Name:	NOTNULL
Key:	HF02001.BASIN_REG
Key:	HF02001.CONCEN
Key:	HF02001.DATANUM

Constraints definition!

	A	B	C		D	E	F	G	H	I	J	K	L	M	N
1	ORCODE	ENTITY	SITE	SITENAME	TEST	DATANUM	TIME	CONCEN	MASSFLUX	DISCHARGE	BASIN	REG	TEMP	TF1	
2	HF0200	1	1-T-5	Notapparent	1	01/09/2003	16790.0100	2.88	202.3	MACK	M				
3	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	2	01/09/2003	20390.0100	20.88	202.2	MACK	M			
4	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	3	01/09/2003	23990.0400	90.41	202.1	MACK	M			
5	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	4	01/09/2003	27590.0400	170.92	202.0	MACK	M			
6	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	5	01/09/2003	31190.1240	249.64	201.9	MACK	M			
7	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	6	01/09/2003	34790.1380	318.12	201.8	MACK	M			
8	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	7	01/09/2003	38390.1980	398.62	201.7	MACK	M			
9	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	8	01/09/2003	41990.2350	473.07	201.6	MACK	M			
10	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	9	01/09/2003	45590.2470	527.28	201.5	MACK	M			
11	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	10	01/09/2003	49190.3010	605.33	201.4	MACK	M			
12	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	11	01/09/2003	52790.3300	663.34	201.2	MACK	M			
13	HF0200	1	1-T-5	CC_IOTT_ALC	IOTT	12	01/09/2003	56390.3360	714.84	201.0	MACK	M			

Semantic Module fonctionnality Example

Principle

Write annotations: kind of "sentence" composed with:

- a subject
- a property
- a property's value

Semantics!

An annotation is added inside an *Attribute* as its direct child (cf. example).

```
<attribute id="att.12">
  <attributeName>biomass</attributeName>
  ...
  <annotation>
    <propertyURI label="of characteristic">http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#ofCharacteristic</propertyURI>
    <valueURI label="Mass">http://ecoinformatics.org/oboe/oboe.1.2/oboe-characteristics.owl#Mass</valueURI>
  </annotation>
  <annotation>
    <propertyURI label="of entity">http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#ofEntity</propertyURI>
    <valueURI label="Plant Sample">http://example.com/example-vocab-1.owl#PlantSample</valueURI>
  </annotation>
</attribute>
```

Deploy Analytics environment on the cloud!



Porté par 3 Infrastructures de Recherche



OBJECTIF

Développer et mettre en œuvre **une infrastructure/plate-forme intégrée de données FAIR** et de services distribués pour l'**observation**, la **modélisation** et la **compréhension du Système Terre**, de la **Biodiversité** et de l'**Environnement**

- sur l'ensemble du cycle de la donnée, de son **acquisition** (spatiale, sols, in-situ) jusqu'à ses **multi-usages** (qualification/validation, stockage, accès, traitements/croisements de données multi-sources/extraction de connaissances, produits/services)
- pour la **communauté scientifique** contribuant à la connaissance du système Terre, de la biodiversité et de l'environnement ; **acteurs publics et privés**

SERVICES GAIA DATA



Services découverte, Accès et Gestion de données

Catalogue (métadonnées, vocabulaires, ontologies), systèmes évolués d'accès et de recherche

Consultation et accès aux données via **web services interopérables** (INSPIRE, Opensearch, STAC, ...)

Services avancés de visualisation

Accompagnement des communautés pour la **FAIRisation**



Services transversaux pour faciliter les travaux transdisciplinaires

Grille de données,

Cloud GAIA Data,

Portail connaissances,

Authentification unique

Support utilisateurs & formation – animation communautés

Support aux campagnes

Analysis Ready Data
Datacubes, ...



Earth Analytics Lab Exploration de la donnée, Bac à sable, Virtual Research Environment

Capacité à se connecter directement sur les centres

Traitements à la demande

Notebook/PANGEO/STAC

Low code / NoCode : Galaxy-E, FG/VIP



Services de production réguliers

Optimisation des traitements (outils orchestration) et formats de données (Zarr, CoG, ...)

Supporté sur un continuum d'infrastructures partagées

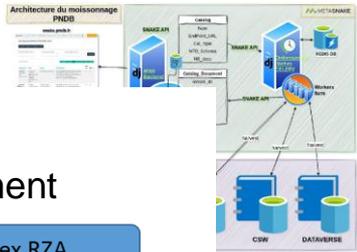
contact@gaia-data.org

www.gaia-data.org



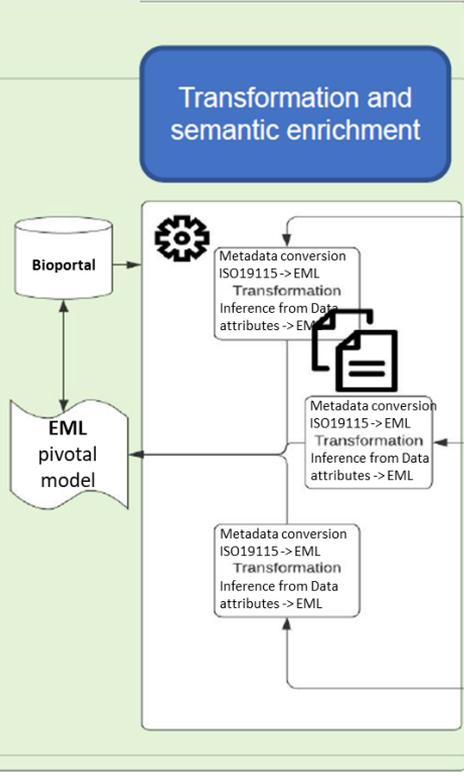
Intégration de données de biodiversité

Moissonnage, transformation, Enrichissement



```

<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<eml:eml xmlns:eml="https://eml.ecoinformatics.org/eml-2.2.0" xmlns:stxml="http://www.xml-cml.org/schema/stxml-1.2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" packageId="doi:10.48502/8005-pk85" system="unknown"
  xsi:schemaLocation="https://eml.ecoinformatics.org/eml-2.2.0 https://nls.lternet.edu/schemas/EML/eml-2.2.0/xsd/eml.xsd">
  <dataset id="dataset-01">
    <title>Kakilia database of marine mammal observation data around the French archipelago of Guadeloupe in the AGOA sanctuary - French Antilles</title>
    <creator>
      <individualName>
        <givenName>Lorraine</givenName>
        <surName>Coché</surName>
      </individualName>
      <organizationName>Université de Bretagne Occidentale</organizationName>
      <electronicMailAddress>lorraine.coché@gmail.com</electronicMailAddress>
      <userID directory="https://orcid.org">https://orcid.org/0000-0003-4909-4848</userID>
    </creator>
  </dataset>
  <attributes>
    <attribute id="att-codeobs">
      <attributeName>code_observation</attributeName>
      <attributeDefinition>Observation code combining the code_sortie and an observation number</attributeDefinition>
      <storageType>string</storageType>
      <measurementScale>
        <nominal>
          <nonNumericDomain>
            <textDomain>
              <definition>Observation code combining the code_sortie and an observation number</definition>
            </textDomain>
          </nonNumericDomain>
        </nominal>
      </measurementScale>
      <missingValueCode>
        <code>NA</code>
        <codeExplanation>No value</codeExplanation>
      </missingValueCode>
      <annotation>
        <propertyURI label="is a">http://www.w3.org/1999/02/22-rdf-syntax-ns#type</propertyURI>
        <valueURI label="occurrenceID">http://rs.tdwg.org/dwc/terms/occurrenceID</valueURI>
      </annotation>
    </attribute>
  </attributes>
</eml:eml>
  
```



SI Biodiv 1 IR ex RZA

Data & metadata Endpoints (OAI-PMH, CSW)

SI Biodiv 2 IR ex ANAEE

Data & metadata Endpoints (OAI-PMH)

SI Biodiv 3 ex Dataverse INRAE

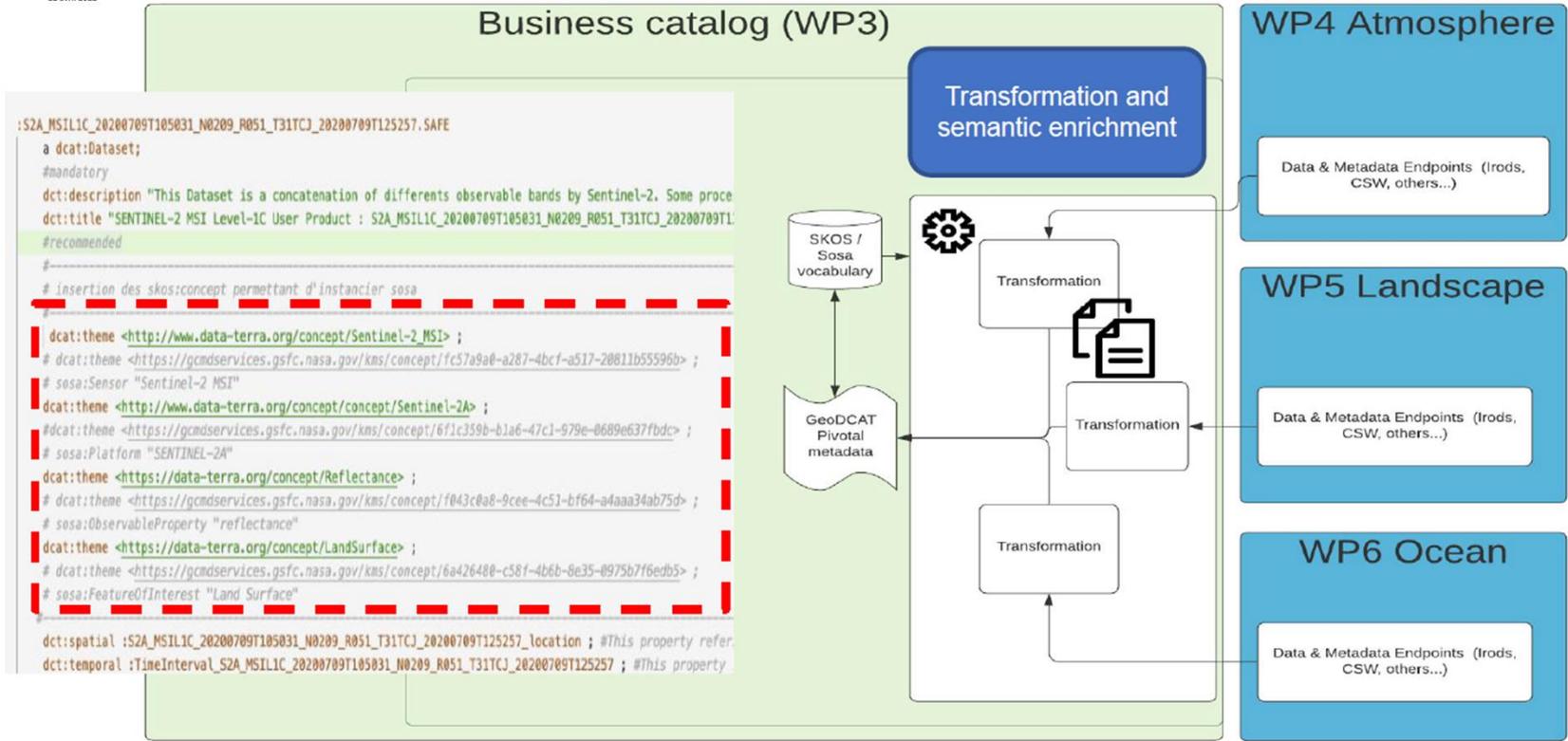
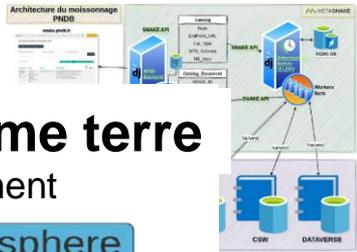
Data & metadata Endpoints (OAI-PMH)



Journée de lancement GAIA Data - EQUIPEX - 12 avril 2022

Intégration de données de biodiversité, climat, système terre

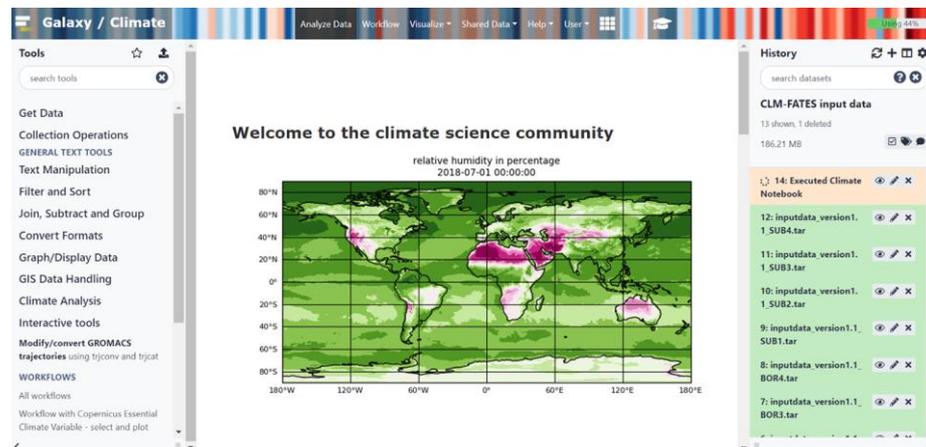
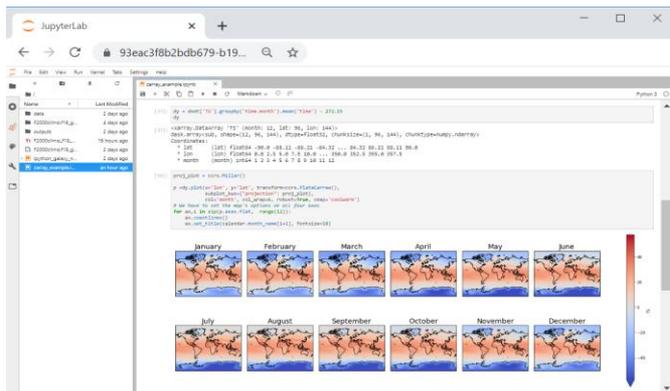
Moissonnage, transformation, Enrichissement



Interopérabilité : Le PNDB et GAIA DATA, focus sur l'analyse

Earth Analytics Lab

- VRE
- VAP
- Visualisation



Une infrastructure pour et par la Recherche : équipe opérationnelle du PNDB

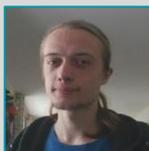


Du technique *et* de l'humain !

Elie ARNAUD

Développeur sur projet

MNHN - PatriNat



Yvan LE BRAS

Resp. de la coord. scien. & tech

MNHN - PatriNat



Coline ROYAUX

Doctorante en écologie
Data Scientist

MNHN - PatriNat & SU - Borea



Mélanie MADELIN

Développeuse sur projet

MNHN - PatriNat



Julien SANANIKONE

Responsable SI, technique

MNHN - PatriNat



Lynda BENNACI RABIEUX

Resp. de la gestion admin.

MNHN - PatriNat



Olivier NORVEZ

Resp. de l'animation

FRB & MNHN - PatriNat



Jean-Denis VIGNE

Coordinateur du PNDB

MNHN



Thomas BOUIX

Coord. SIB, GBIF, PNDB

MNHN - PatriNat



Sandrine PAVOINE

Conseillère scientifique

MNHN - CESCO



Recrutement en cours

Chef.fe de projet opérationnel

MNHN - PatriNat



Comprendre le

PN**D****B** Pôle National
de Données de Biodiversité

en 1 heure !

Merci !
Questions ?

Olivier Norvez (FRB & MNHN) *et al.*



@PNDBiodiv & www.pndb.fr



Séances pour une terre durable



AgroParisTech

