

Partager des données interoperables en recherche

Focus sur les standards et les entrepôts

Louise Percevault

louise.percevault@umontpellier.fr



UNIVERSITÉ DE
MONTPELLIER



CC BY 4.0



1. Introduction

2. Les entrepôts de données de recherche

3. Les standards de partage de (méta)données

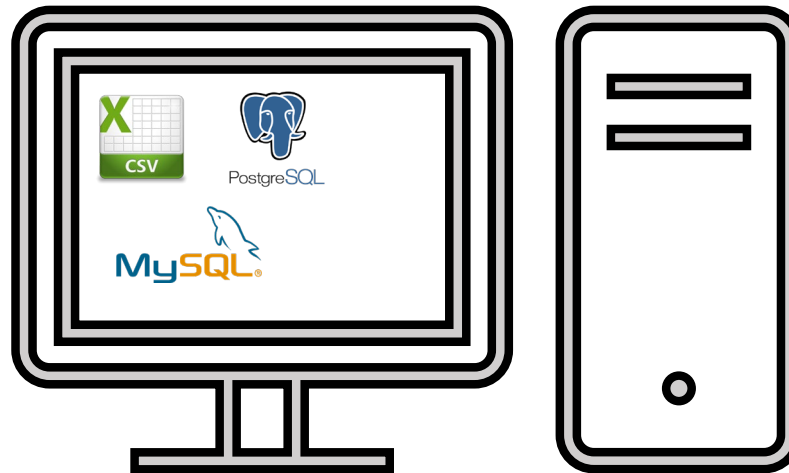
4. Pour conclure



1.Introduction

”Les « données de la recherche » sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche.” ([OCDE 2007](#))

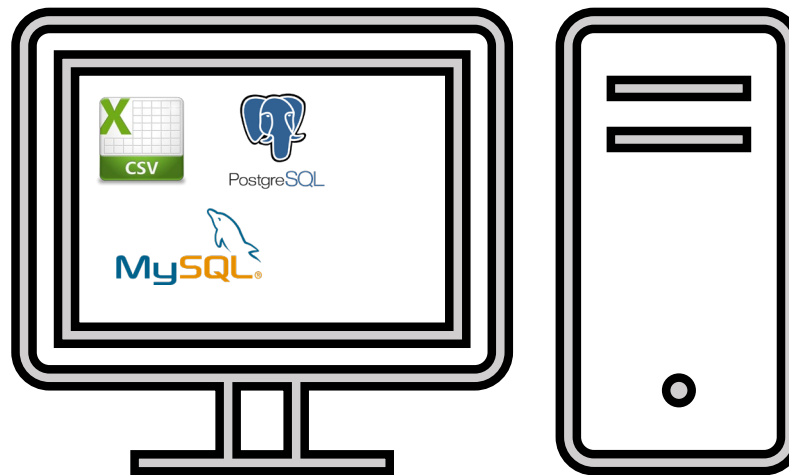
Les données de la recherche



1.Introduction

”Les « données de la recherche » sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche.” ([OCDE 2007](#))

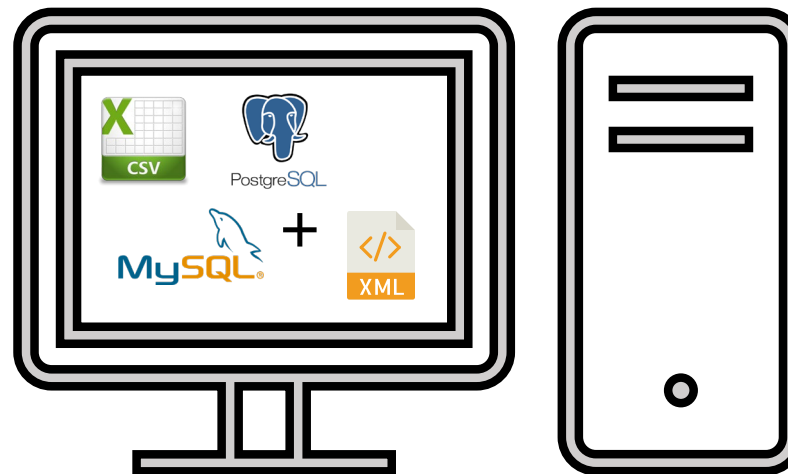
Les données de la recherche



Ce terme ne s'applique pas aux : carnets de laboratoire, analyses préliminaires [...], programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris).

1.Introduction

Les données de la recherche



et les métadonnées associées

Les **métadonnées** (la documentation des données) peuvent être définies comme représentant les informations ou instructions de niveau supérieur qui décrivent le contenu, le contexte, la qualité, la structure et l'accessibilité **d'un ensemble de données spécifique** (Michener et al., 1997)

Doivent contenir **toutes les informations nécessaires** et suffisantes pour permettre l'**utilisation secondaire (réutilisation) à long terme de l'ensemble de données** par le(s) chercheur(s) initial(aux), ainsi que l'utilisation par d'autres scientifiques qui n'ont pas été directement impliqués dans les efforts de recherche initiaux (Michener et al., 1997).

1.Introduction

Le renseignement incomplet des métadonnées engendre inévitablement une diminution, une perte d'information sur ou contenue dans les données au cours du temps, jusqu'à la rendre inutilisable (Delalandre 2024 ; Garnier et al., 2016 ; Michener et al., 1997).

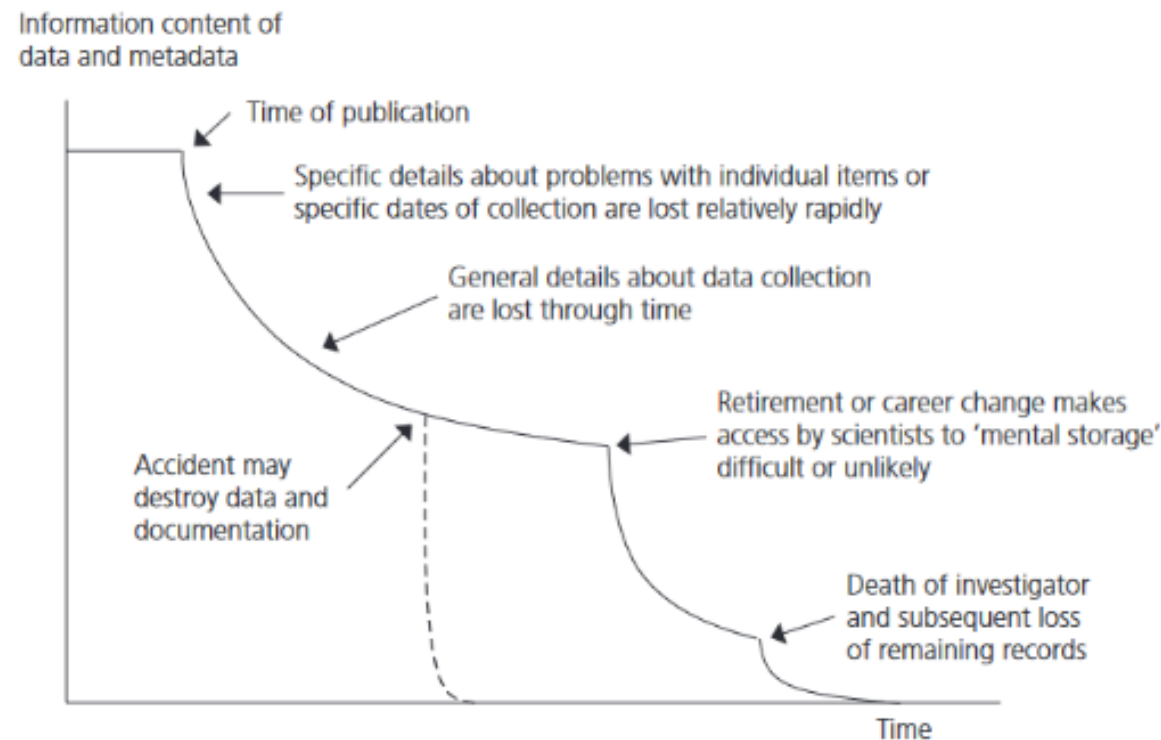


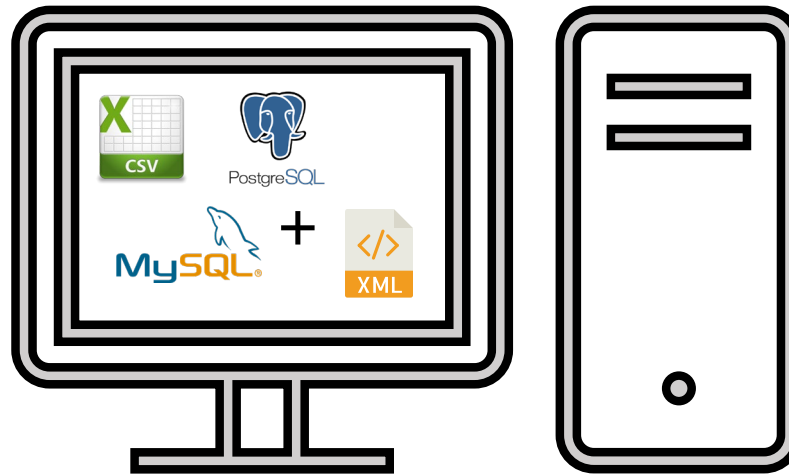
Figure 29 : Perte de données au fil du temps suite à leur publication. Figure issue de Garnier et al. (2016), adaptée de Michener et al. (1997).

1.Introduction

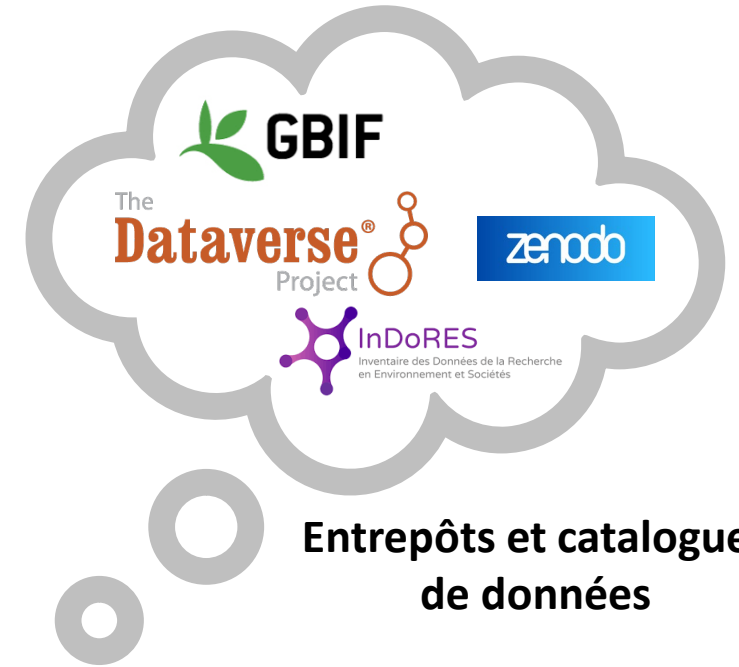
Standards d'échange de (méta)données



Les données de la recherche



et les métadonnées associées



Entrepôts et catalogues de données



1. Introduction

2. Les entrepôts de données de recherche

3. Les standards de partage de (méta)données

4. Pour conclure



2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).

2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).

Entrepôts

2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

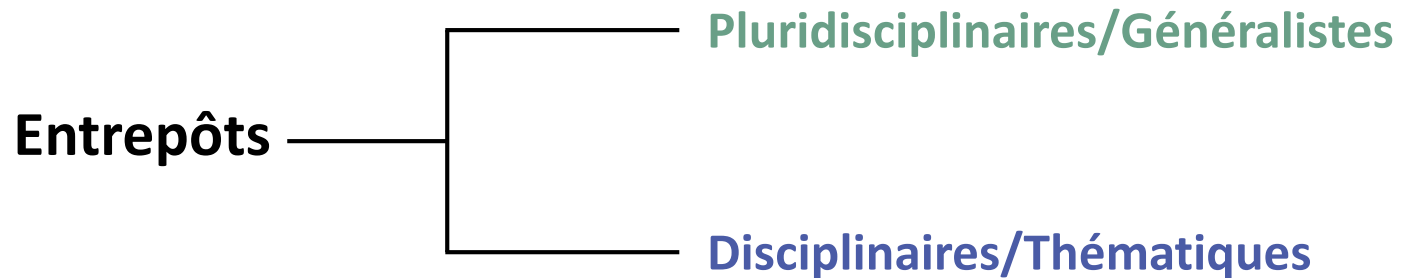
Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).

Entrepôts  **Pluridisciplinaires/Généralistes**

2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2023).



2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le catalogue d'entrepôts international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).



2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).

Thématiques

Généralistes - Institutionnels

Généralistes

2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).

Thématiques

- [SEANOE](#) (Sea scientific open data publication).
- [PANGAEA](#) (Data Publisher for Earth & Environmental Science) : Données sur le système Terre.
- [THEIA](#) : Données et services pour les surfaces continentales.
- [AnaEE-France](#) : Systèmes expérimentaux en situations contrôlées.
- [GBIF](#) (Global Biodiversity Information Facility) : Occurrence de taxon.
- [GenBank](#) : Séquence d'ADN.
- [TRY](#) (Plant Trait Database) : Traits fonctionnels des plantes.

Généralistes - Institutionnels

Généralistes

2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).

Thématiques

- [SEANOE](#) (Sea scientific open data publication).
- [PANGAEA](#) (Data Publisher for Earth & Environmental Science) : Données sur le système Terre.
- [THEIA](#) : Données et services pour les surfaces continentales.
- [AnaEE-France](#) : Systèmes expérimentaux en situations contrôlées.
- [GBIF](#) (Global Biodiversity Information Facility) : Occurrence de taxon.
- [GenBank](#) : Séquence d'ADN.
- [TRY](#) (Plant Trait Database) : Traits fonctionnels des plantes.

Généralistes - Institutionnels

- [DataSuds](#) (IRD)
- [Dataverse CIRAD](#)
- [Data INRAE](#)
- [Data Umontpellier](#) (UM)
- [InDoRES](#) (CNRS-MNHN)
- [Recherche Data Gouv](#)

Généralistes

2. Les entrepôts de données de recherche

Les **entrepôts de données** de recherche (Research Data Repository ou Data Repository) sont des emplacements centralisés destinés à **diffuser, rendre visibles et à faciliter l'accès** aux **données issues de la recherche** (Delalandre 2024).

Le moteur de recherche international Registry of Research data Repositories ([Re3data](#)), comptait en 2023 plus de 3000 entrepôts de données de recherche différents (Strecker et al., 2022).

Thématiques

- [SEANOE](#) (Sea scientific open data publication).
- [PANGAEA](#) (Data Publisher for Earth & Environmental Science) : Données sur le système Terre.
- [THEIA](#) : Données et services pour les surfaces continentales.
- [AnaEE-France](#) : Systèmes expérimentaux en situations contrôlées.
- [GBIF](#) (Global Biodiversity Information Facility) : Occurrence de taxon.
- [GenBank](#) : Séquence d'ADN.
- [TRY](#) (Plant Trait Database) : Traits fonctionnels des plantes.

Généralistes - Institutionnels

- [DataSuds](#) (IRD)
- [Dataverse CIRAD](#)
- [Data INRAE](#)
- [Data Umontpellier](#) (UM)
- [InDoRES](#) (CNRS-MNHN)
- [Recherche Data Gouv](#)

Généralistes

- [Zenodo](#) (Européen)
- [Dryad](#) (International)
- [Figshare](#) (International)

2. Les entrepôts de données de recherche

Quel entrepôt ? Pour quel type de données ?

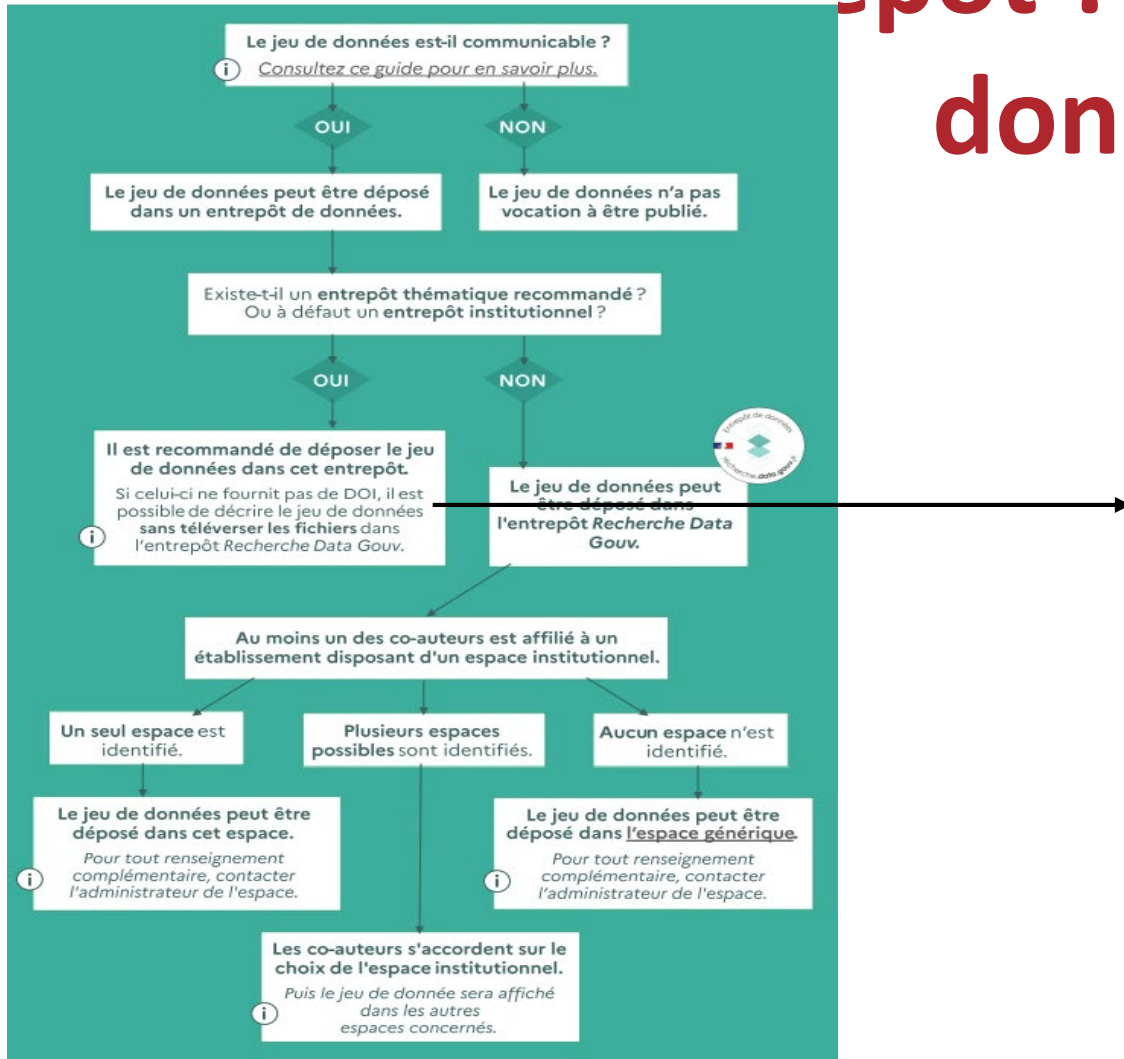
2. Les entrepôts de données de recherche

Où publier vos données ?

Ce logigramme a pour objectif d'aider le déposant à identifier où il pourra publier son jeu de données.



Entrepôt ? Pour quel type de données ?



Entrepôts de confiance

Sur cette page, Recherche Data Gouv se fait le relais des entrepôts thématiques de confiance proposés par le Collège des Données de la recherche du Comité pour la science ouverte.

Le partage de données issues des activités de recherche implique de les mettre à disposition dans des entrepôts reconnus par les communautés scientifiques, thématiques, institutionnels ou généralistes.

Pour faciliter le travail des équipes de recherche, le collège des Données de la recherche propose une méthode d'identification des entrepôts thématiques de confiance, ainsi qu'une première liste d'entrepôts découlant de la grille de critères d'analyse retenus à découvrir ci-dessous.

Cette liste est évolutive et a vocation à être progressivement complétée par le collège des Données de la Recherche. Elle n'est par ailleurs pas exhaustive : ainsi, ne pas figurer dans la liste n'implique pas que l'entrepôt concerné n'est pas de confiance.

⚠ N'oubliez pas, la dernière version à jour de la liste de référence est celle publiée sur le site Ouvrir la Science.

📧 Pour toute question relative aux entrepôts de confiance, contactez le Collège des Données de la recherche du Comité pour la science ouverte.

Archéologie	+
Astronomie	+
Biologie	+
Chimie	+
Physique	+
Sciences de l'environnement	-
EaSy Data	Sciences de l'environnement Sciences de la Terre
Global Biodiversity Information Facility (GBIF)	Biodiversité Biologie animale Biologie végétale Ecologie
Inventaire des données de recherche en environnement et sociétés (data.InDoRES)	Bioarchéologie Ecologie Environnement
PANGAEA	Ecologie Environnement Sciences de la Terre Sciences du vivant
SEANOE	Océanographie
Sextant	Océanographie
World Data Center of Climate (WDC)	Sciences de l'environnement
Sciences humaines et sociales	+

2. Les entrepôts de données de recherche

Quel entrepôt ? Pour quel type de données ?

Avons-nous besoin d'un DOI ?

Un entrepôt unique ?

Notre discipline a-t-elle un ou plusieurs entrepôts dédiés ?

Existe-t-il un entrepôt spécifique pour notre type de données ?

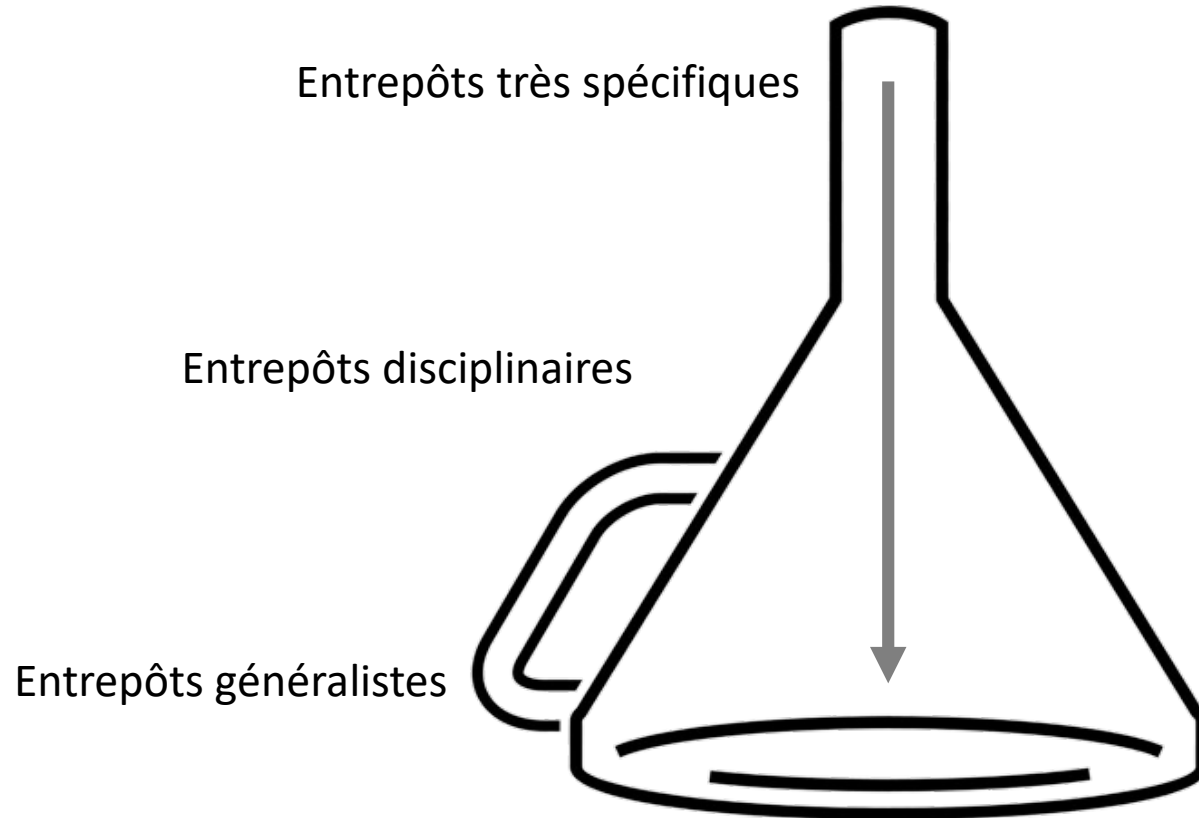
A-t-on le projet de faire un data paper ?

Quelle visibilité nous souhaitons apporter à nos données ?

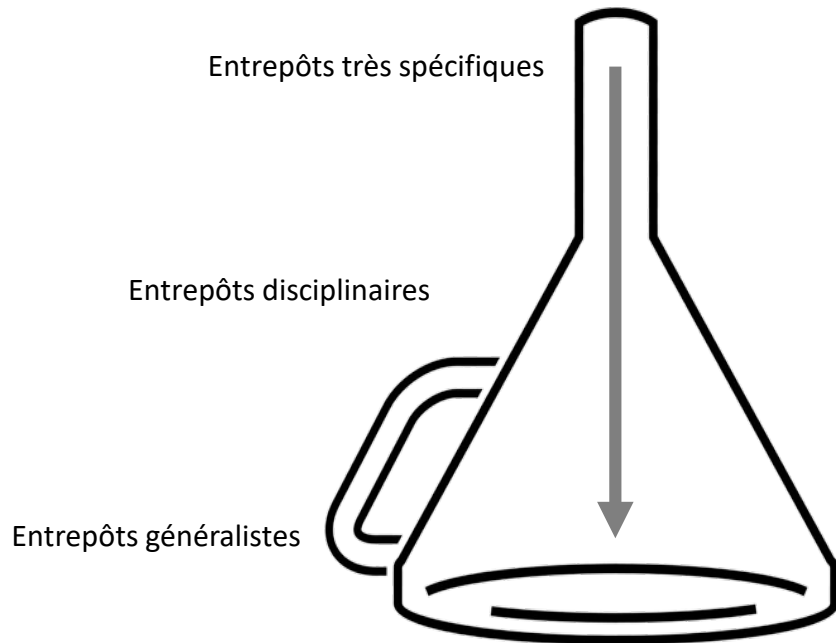
Quelles moyens humains et financiers peut-on allouer au processus de publication des données ?

2. Les entrepôts de données de recherche

Quel entrepôt ? Pour quel type de données ?



2. Les entrepôts de données de recherche



Les catalogues d'entrepôts :

- [Re3Data](#)
- [Cat OPIDoR – Entrepôts de données](#)
- [Datacite Repository Finder](#)

Autres ressources :

- Sélectionner un entrepôt thématique de confiance (Ouvrir la science) : <https://www.ouvrirlascience.fr/selectionner-un-entrepot-thematique-de-confiance-pour-la-diffusion-des-donnees-de-recherche-note-methodologique/>
- Les principaux entrepôts et catalogues de données en sciences de l'Environnement (DORANum): https://doranum.fr/environnement/principaux-entrepots-et-catalogues-de-donnees-en-sciences-de-lenvironnement_10_13143_bevh-862/
- *Sylvie Cocaud, Pascal Aventurier(2017). Les entrepôts de données de recherche. ANF réseaux Renatis et Médiçi (CNRS). pp.63 slides.* <https://hal.science/hal-01595599/document>
- Disponibilité à long terme des données de recherche – disparition d'entrepôt : Strecker et al. 2023 [\[2310.06712\]](#) [Disappearing repositories -- taking an infrastructure perspective on the long-term availability of research data \(arxiv.org\)](#)



1. Introduction

2. Les entrepôts de données de recherche

3. Les standards de partage de (méta)données

4. Pour conclure



3. Les standards de partage de (méta)données

Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

Interopérabilité : capacité d'un produit ou d'un système à fonctionner avec d'autres produits ou systèmes existants ou futurs, sans restriction ni modification supplémentaire.

3. Les standards de partage de (méta)données

Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

Interopérabilité : capacité d'un produit ou d'un système à fonctionner avec d'autres produits ou systèmes existants ou futurs, sans restriction ni modification supplémentaire.

- - > Le rôle des standards est de permettre, d'augmenter **l'interopérabilité** d'un produit ou d'un système.

3. Les standards de partage de (méta)données

Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

Interopérabilité : capacité d'un produit ou d'un système à fonctionner avec d'autres produits ou systèmes existants ou futurs, sans restriction ni modification supplémentaire.

--> Le rôle des standards est de permettre, d'augmenter **l'interopérabilité** d'un produit ou d'un système.

Standards de données

- Structuration des données
- Interopérabilité forte des données
- Facilite l'agrégation de données

Standards de métadonnées

- Structuration des métadonnées
- Niveaux de précision très variables
- Facilite la bonne compréhension des données



Complémentaires

3. Les standards de partage de (méta)données

Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

eventDate	decimalLatitude	decimalLongitude	footprintWKT	scientificName	originalNameUsage
10/03/2016	43.38678	6.45929	POINT(6.45929 43.38678)	Poa bulbosa L.	Poa bulbosa L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.39907)	Gratiola officinalis L.	Gratiola officinalis L., 1753
06/05/2015	43.37602	6.47122	POINT(6.47122 43.37602)	Phillyrea angustifolia L.	Phillyrea angustifolia L., 1753

Term Name	dwc:eventDate
Term IRI	http://rs.tdwg.org/dwc/terms/eventDate
Modified	2023-06-28
Term version IRI	http://rs.tdwg.org/dwc/terms/version/eventDate-2023-06-28
Label	Event Date
Definition	The date-time or interval during which a dwc:Event occurred. For occurrences, this is the date-time when the dwc:Event was recorded. Not suitable for a time in a geological context.
Notes	Recommended best practice is to use a date that conforms to ISO 8601-1:2019.
Examples	<p>1963-03-08T14:07-0600 (8 Mar 1963 at 2:07pm in the time zone six hours earlier than UTC)</p> <p>2009-02-20T08:40Z (20 February 2009 8:40am UTC)</p> <p>2018-08-29T15:19 (3:19pm local time on 29 August 2018)</p> <p>1809-02-12 (some time during 12 February 1809)</p>

eventDate	decimalLatitude	decimalLongitude	footprintWKT	scientificName	originalNameUsage
16/09/2014	43.38785	6.47373	POINT(6.47373 43.38785)	Agrostis castellana Bois: Agrostis capillaris subsp. castellana (Boiss. & Reut.) O.B	

3. Les standards de partage de (méta)données

Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

eventDate	decimalLatitude	decimalLongitude	footprintWKT	scientificName	originalNameUsage
10/03/2016	43.38678	6.45929	POINT(6.45929 43.38678)	Poa bulbosa L.	Poa bulbosa L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.39907)	Gratiola officinalis L.	Gratiola officinalis L., 1753
06/05/2015	43.37602	6.47122	POINT(6.47122 43.37602)	Phillyrea angustifolia L.	Phillyrea angustifolia L., 1753
23/04/2014	43.38765	6.47164	POINT(6.47164 43.38765)	Mentha pulegium L.	Mentha pulegium L., 1753
23/04/2014	43.38689	6.45972	POINT(6.45972 43.38689)	Prunella vulgaris L.	Prunella vulgaris L., 1753
17/05/2018				Term Name dwc:scientificName	(Fenzl) Walters
23/04/2014					
10/03/2016				Term IRI	http://rs.tdwg.org/dwc/terms/scientificName
10/03/2016					abbayes
06/05/2015				Modified	2023-06-28
13/05/2015				Term	http://rs.tdwg.org/dwc/terms/version/scientificName-2023-06-28
13/05/2015				version IRI	
13/05/2015					
07/05/2015				Label	Scientific Name
24/06/2015					uv., 1812
10/03/2016				Definition	The full scientific name, with authorship and date information if known. When forming part of a dwc:identi
24/06/2015					this should be the name in lowest level taxonomic rank that can be determined. This term should not cont
20/11/2015					identification qualifications, which should instead be supplied in the dwc:identificationQualifier term.
16/04/2016					
06/05/2015				Notes	This term should not contain identification qualifications, which should instead be supplied in the
23/04/2014	43.38585	6.45569	POINT(6.45569 43.38585)	Moenchia erecta (L.) G.	Moenchia erecta (L.) P. Gaertn., B. Mey. & Scherb., 1795
13/05/2015	43.38758	6.45398	POINT(6.45398 43.38758)	Linum trigynum L.	Linum trigynum L., 1753
23/04/2014	43.38787	6.47391	POINT(6.47391 43.38787)	Isoetes durieui Bory	Isoetes durieui Bory, 1844
06/05/2015	43.37498	6.47076	POINT(6.47076 43.37498)	Crucianella angustifolia	Crucianella angustifolia L., 1753
13/05/2015	43.37351	6.45818	POINT(6.45818 43.37351)	Vincetoxicum hirsutum	Vincetoxicum hirsutum Medik., 1790
16/09/2014	43.38785	6.47373	POINT(6.47373 43.38785)	Agrostis castellana Boiss.	Agrostis capillaris subsp. castellana (Boiss. & Reut.) O.B

3. Les standards de partage de (méta)données

Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

eventDate	decimalLatitude	decimalLongitude	footprintWKT	scientificName	originalNameUsage
10/03/2016	43.38678	6.45929	POINT(6.45929 43.3	Poa bulbosa L.	Poa bulbosa L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.3	Gratiola officinalis L.	Gratiola officinalis L., 1753
06/05/2015	43.37602	6.47122	POINT(6.471220000	Phillyrea angustifolia L.	Phillyrea angustifolia L., 1753
23/04/2014	43.38765	6.47164	POINT(6.471640000	Mentha pulegium L.	Mentha pulegium L., 1753
23/04/2014	43.38689	6.45972	POINT(6.459720000	Prunella vulgaris L.	Prunella vulgaris L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.3	Montia arvensis Wallr.	Montia fontana subsp. chondrosperma (Fenzl) Walters
23/04/2014	43.39907	6.46744	POINT(6.467440000	Serapias lingua L.	Serapias lingua L., 1753

Column label	Term IRI
eventDate	http://rs.tdwg.org/dwc/terms/eventDate
decimalLatitude	http://rs.tdwg.org/dwc/terms/decimalLatitude
decimalLongitude	http://rs.tdwg.org/dwc/terms/decimalLongitude
footprintWKT	http://rs.tdwg.org/dwc/terms/footprintWKT
scientificName	http://rs.tdwg.org/dwc/terms/scientificName
originalNameUsage	http://rs.tdwg.org/dwc/terms/originalNameUsage

6.461650000	Allosorus tinaei (Tod.) C Cheilanthes tinaei Tod., 1886
6.463660000	Cladonia mediterranea Cladonia mediterranea P. A. Duvign. & Abbayes
6.473400000	Tuberaria guttata (L.) Fc Tuberaria guttata (L.) Fourr., 1868
6.45443 43.3	Juncus effusus L. Juncus effusus L., 1753
6.463610000	Carex flacca Schreb. Carex flacca Schreb., 1771
6.46579 43.3	Carex oedipostyla Duva Carex oedipostyla Duval-Jouve, 1870
6.46166 43.3	Brachypodium sylvaticu Brachypodium sylvaticum (Huds.) P. Beauv., 1812
6.474380000	Lythrum salicaria L. Lythrum salicaria L., 1753
6.46356 43.3	Juniperus oxycedrus sul Juniperus oxycedrus L. subsp. oxycedrus
6.45717 43.3	Airopsis tenella (Cav.) A Airopsis tenella (Cav.) Asch. & Graebn., 1899
6.472880000	Cyperus eragrostis Lam Cyperus eragrostis Lam., 1791
6.468490000	Agrostis castellana Bois: Agrostis castellana Boiss. & Reut., 1842
6.47259 43.3	Trifolium campestre Scf Trifolium campestre Schreb., 1804
6.455690000	Moenchia erecta (L.) G. Moenchia erecta (L.) P. Gaertn., B. Mey. & Scherb., 1795
6.45398 43.3	Linum trigynum L. Linum trigynum L., 1753
6.473910000	Isoetes durieui Bory Isoetes durieui Bory, 1844
6.47076 43.3	Crucianella angustifolia Crucianella angustifolia L., 1753
6.45818 43.3	Vincetoxicum hirundina Vincetoxicum hirundinaria Medik., 1790
6.473730000	Agrostis castellana Bois: Agrostis capillaris subsp. castellana (Boiss. & Reut.) O.B

3. Les standards de partage de (méta)données

Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

eventDate	decimalLatitude	decimalLongitude	footprintWKT	scientificName	originalNameUsage
10/03/2016	43.38678	6.45929	POINT(6.45929 43.3	Poa bulbosa L.	Poa bulbosa L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.3	Gratiola officinalis L.	Gratiola officinalis L., 1753
06/05/2015	43.37602	6.47122	POINT(6.471220000	Phillyrea angustifolia L.	Phillyrea angustifolia L., 1753
23/04/2014	43.38765	6.47164	POINT(6.471640000	Mentha pulegium L.	Mentha pulegium L., 1753
23/04/2014	43.38689	6.45972	POINT(6.459720000	Prunella vulgaris L.	Prunella vulgaris L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.3	Montia arvensis Wallr.	Montia fontana subsp. chondrosperma (Fenzl) Walters
23/04/2014	43.39907	6.46744	POINT(6.467440000	Serapias lingua L.	Serapias lingua L., 1753

Données

Métadonnées

Column label	Term IRI
eventDate	http://rs.tdwg.org/dwc/terms/eventDate
decimalLatitude	http://rs.tdwg.org/dwc/terms/decimalLatitude
decimalLongitude	http://rs.tdwg.org/dwc/terms/decimalLongitude
footprintWKT	http://rs.tdwg.org/dwc/terms/footprintWKT
scientificName	http://rs.tdwg.org/dwc/terms/scientificName
originalNameUsage	http://rs.tdwg.org/dwc/terms/originalNameUsage

6.461650000	Allosorus tinaei (Tod.) C Cheilanthes tinaei Tod., 1886
6.463660000	Cladonia mediterranea Cladonia mediterranea P. A. Duvign. & Abbayes
6.473400000	Tuberaria guttata (L.) Fc Tuberaria guttata (L.) Fourr., 1868
6.45443 43.3	Juncus effusus L. Juncus effusus L., 1753
6.463610000	Carex flacca Schreb. Carex flacca Schreb., 1771
6.46579 43.3	Carex oedipostyla Duva Carex oedipostyla Duval-Jouve, 1870
6.46166 43.3	Brachypodium sylvaticu Brachypodium sylvaticum (Huds.) P. Beauv., 1812
6.474380000	Lythrum salicaria L. Lythrum salicaria L., 1753
6.46356 43.3	Juniperus oxycedrus sul Juniperus oxycedrus L. subsp. oxycedrus
6.45717 43.3	Airopsis tenella (Cav.) A Airopsis tenella (Cav.) Asch. & Graebn., 1899
6.472880000	Cyperus eragrostis Lam Cyperus eragrostis Lam., 1791
6.468490000	Agrostis castellana Bois: Agrostis castellana Boiss. & Reut., 1842
6.47259 43.3	Trifolium campestre Scf Trifolium campestre Schreb., 1804
6.455690000	Moenchia erecta (L.) G. Moenchia erecta (L.) P. Gaertn., B. Mey. & Scherb., 1795
6.45398 43.3	Linum trigynum L. Linum trigynum L., 1753
6.473910000	Isoetes durieui Bory Isoetes durieui Bory, 1844
6.47076 43.3	Crucianella angustifolia Crucianella angustifolia L., 1753
6.45818 43.3	Vincetoxicum hirundina Vincetoxicum hirundinaria Medik., 1790
6.473730000	Agrostis castellana Bois: Agrostis capillaris subsp. castellana (Boiss. & Reut.) O.B

3. Les standards de partage de (méta)données

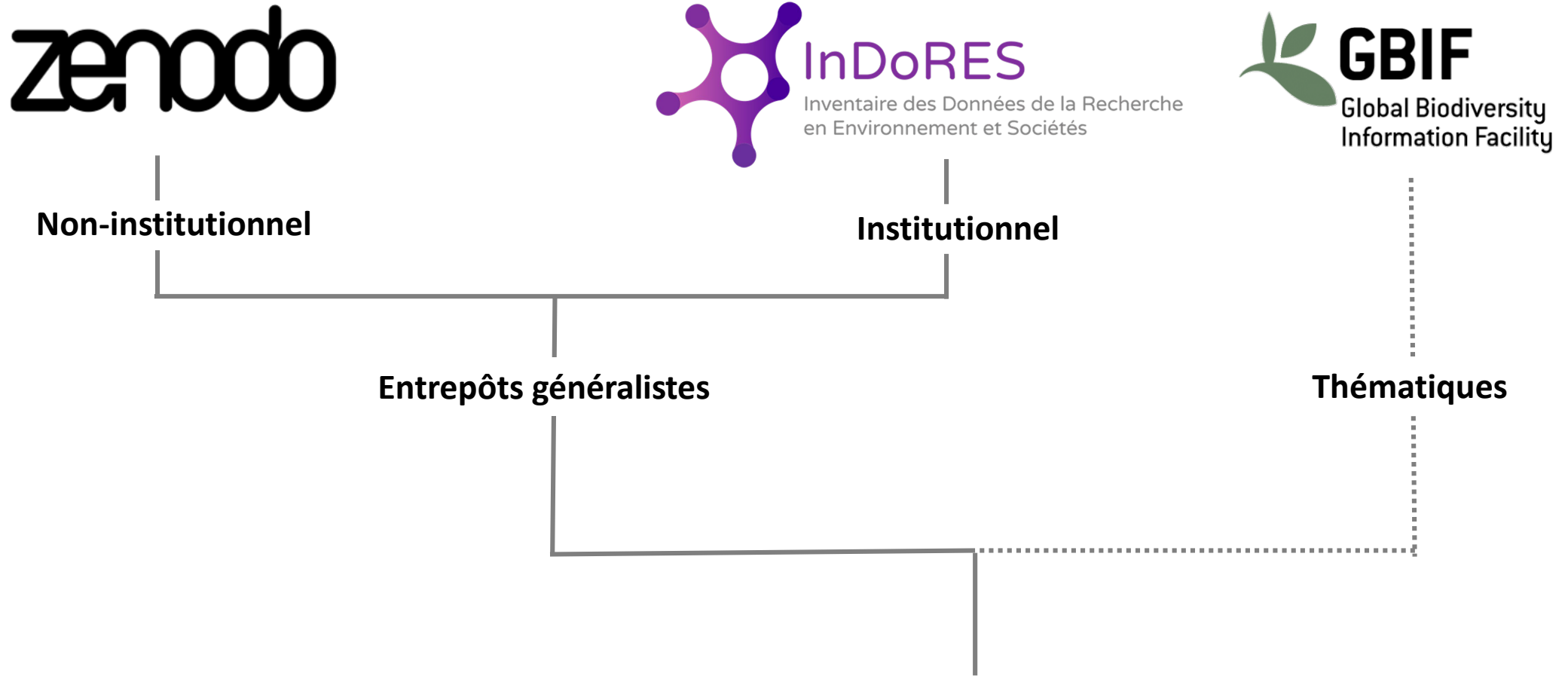
Un **standard** est une **référence** publiée dont la diffusion et l'utilisation sont **répandues** et **reconnues** par une grande partie des personnes travaillant dans un domaine (Garnier et al., 2016). Il permet d'établir un **langage commun** et ainsi **d'augmenter l'intégrité, la précision, la cohérence des données** en clarifiant des significations ambiguës.

Structuration des données pour analyse (\neq) Structuration des données pour diffusion

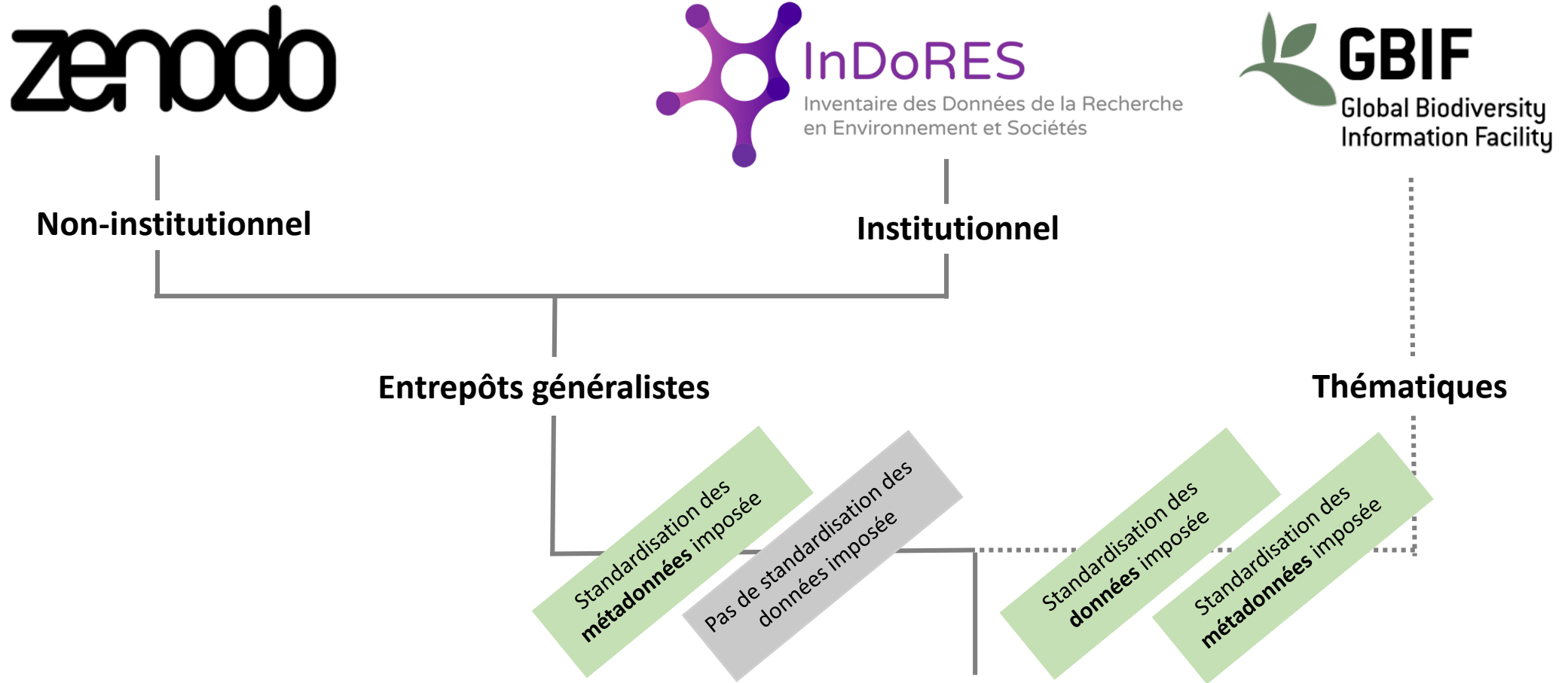


(Delalandre 2024 ; Poissot et al., 2019)

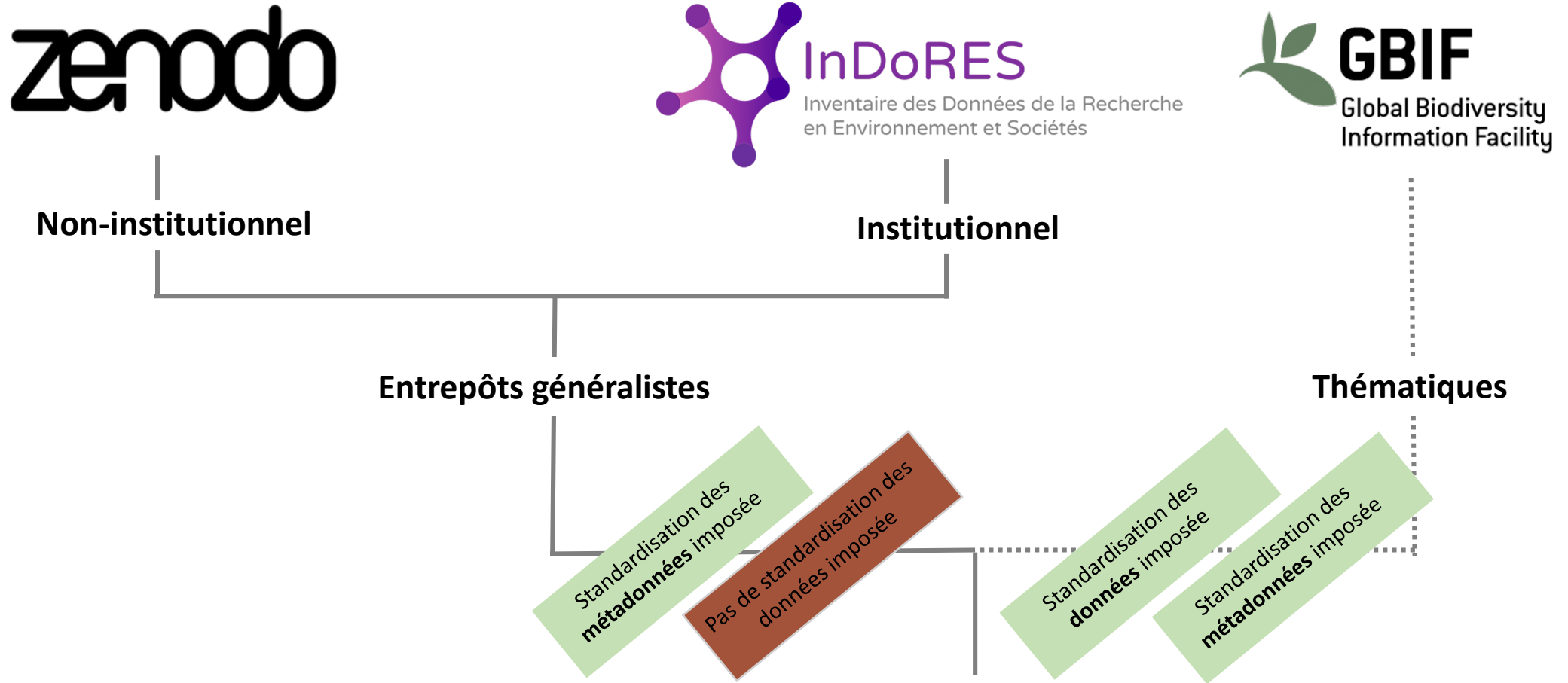
3. Les standards de partage de (méta)données



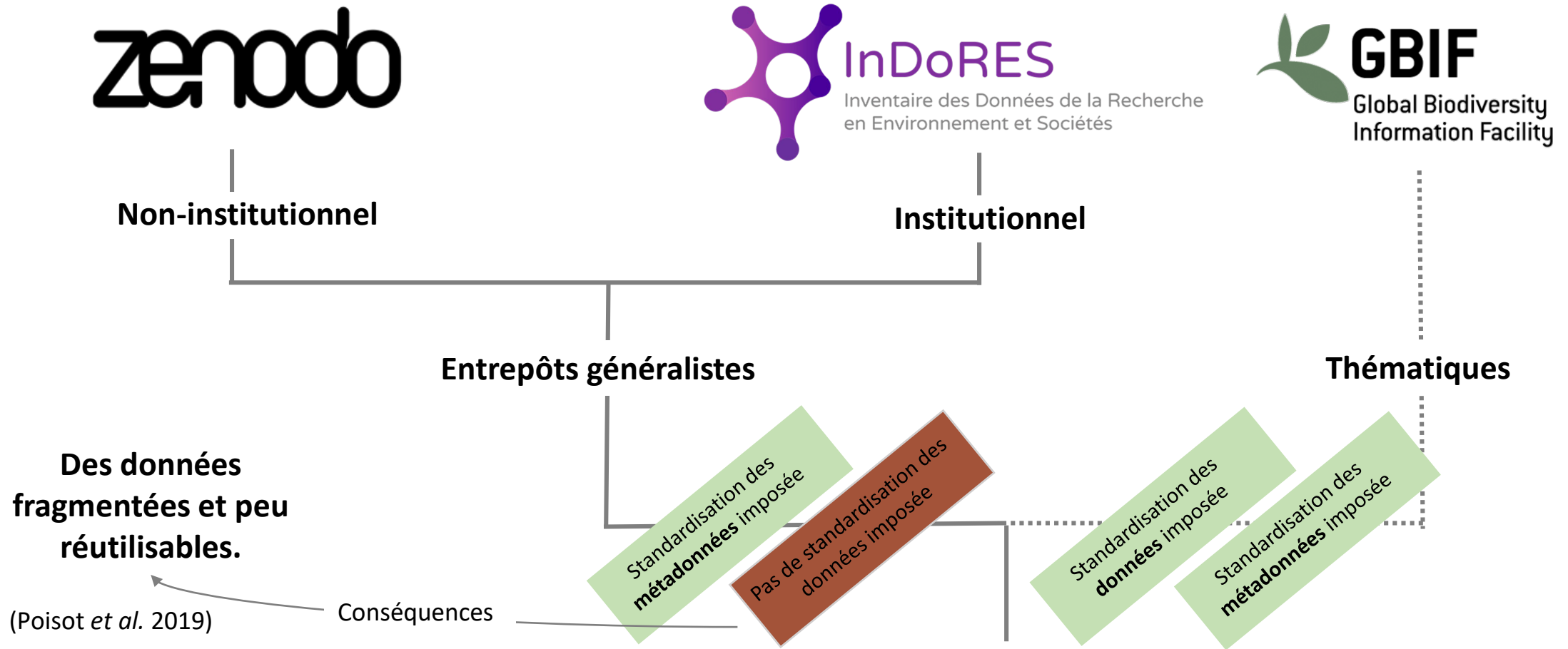
3. Les standards de partage de (méta)données



3. Les standards de partage de (méta)données



3. Les standards de partage de (méta)données



3. Les standards de partage de (méta)données

Comment enrichir un dépôt de données dans un entrepôt généraliste pour en garantir son interopérabilité ?

- Utiliser un **ou plusieurs standards de (méta)données**.
- Fournir des métadonnées riches sous forme de **ReadMe.txt**.
- Ajouter un **dictionnaire de données/d'attributs** qui décrit l'ensemble des variables de votre jeu de données.

Ressources

Les catalogues de standards :

- [Metadata Standard Directory \(RDA\)](#)
- [FAIR Sharing](#)
- [Metadata Standards Catalog](#)
- [Disciplinary Metadata](#)
- [FGDC GEOSPATIAL STANDARDS](#)
- [OGC Standards](#)
- [Darwin Core Extensions](#)
- [TDWG Catalogue](#)

Les catalogues d'entrepôts :

- [Re3Data](#)
- [Cat OPIDoR – Entrepôts de données](#)
- [Datacite Repository Finder](#)

Autres ressources :

- Sélectionner un entrepôt thématique de confiance (Ouvrir la science) : <https://www.ouvrirlascience.fr/selectionner-un-entrepot-thematique-de-confiance-pour-la-diffusion-des-donnees-de-recherche-note-methodologique/>
- Les principaux entrepôts et catalogues de données en sciences de l'Environnement (DORANum): https://doranum.fr/environnement/principaux-entrepots-et-catalogues-de-donnees-en-sciences-de-lenvironnement_10_13143_bevh-862/
- *Sylvie Cocaud, Pascal Aventurier(2017). Les entrepôts de données de recherche. ANF réseaux Renatis et Médiçi (CNRS). pp.63 slides.* <https://hal.science/hal-01595599/document>
- Disponibilité à long terme des données de recherche – disparition d'entrepôt : Strecker et al. 2023 [\[2310.06712\]](#) [Disappearing repositories -- taking an infrastructure perspective on the long-term availability of research data \(arxiv.org\)](#)

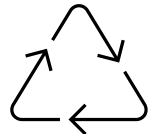
- 
- 
1. Introduction
 2. Les entrepôts de données de recherche
 3. Les standards de partage de (méta)données
 - 4. Pour conclure**

4. Pour conclure

Comment aboutir à des données disponibles et interopérables ?



Créer des nouveaux standards
spécifiques



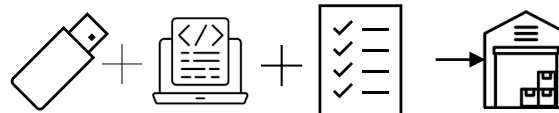
Valoriser l'existant et l'adapter
aux besoins



S'inspirer et apprendre des initiatives
concluantes



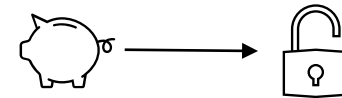
Encourager à l'utilisation
d'entrepôts spécialisés



Des entrepôts de données avec des
contraintes de standardisation



Valoriser le partage des données
autant que les résultats



Des financeurs et des éditeurs qui
encouragent à l'interopérabilité

Bibliographie :

Delalandre, L. (2024). Relations traits-environnement chez les végétaux : du cycle de vie des organismes au cycle de vie des données. (Thèse)

Garnier, E., Navas, M.-L. & Grigulis, K. (2016). *Plant functional diversity: organism traits, community structure, and ecosystem properties*. 1st ed. Oxford university press, Oxford (GB). (Livre)

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B. & Stafford, S.G. (1997). NONGEOSPATIAL METADATA FOR THE ECOLOGICAL SCIENCES. *Ecological Applications*, 7, 330–342. [https://doi.org/10.1890/1051-0761\(1997\)007\[0330:NMFTES\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2)

Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D. & Peres-Neto, P. (2019). Ecological Data Should Not Be So Hard to Find and Reuse. *Trends in Ecology & Evolution*, 34, 494–496. <https://doi.org/10.1016/j.tree.2019.04.005>

Strecker, D., Pampel, H., Schabinger, R. & Weisweiler, N.L. (2023). Disappearing repositories -- taking an infrastructure perspective on the long-term availability of research data. <https://doi.org/10.48550/arXiv.2310.06712>

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., *et al.* (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*, 7, e29715. <https://doi.org/10.1371/journal.pone.0029715>

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Glossaire :

Table 9.1 A concise (eco)informatics glossary

Term	Definition
Automated reasoning	Algorithms and software systems for automating the computation of logical inferences (typically refers to procedures for deductive reasoning)
Class	A class regroups an ensemble of objects having the same properties and behaviours (i.e. the evolution of properties over time)
Concept	An abstract, general, objective, and stable verbal representation of a real world object
Conceptualization	The action of elaborating a concept or an ensemble of communicable concepts
Controlled vocabulary	A list of terms and their definitions, established by a community of users. Controlled vocabularies can be considered as the language spoken by researchers within a scientific domain so as to clearly understand each other
Data integration	The ensemble of procedures aiming to facilitate the correspondence and combination of information derived from different sources, in a pertinent and useful manner
Data model	A data model organizes data elements and standardizes how the data elements relate to one another
Distributed (data) sources	Data sources located on different computer servers, or in different locations on the same server
Ecoinformatics	A field of research and development focused on the interface between ecology, computer science, and information technology
Entity	A real world, existing object, but representable only by a concept
Instance	A real world, existing object, with behaviour corresponding to the class to which it belongs: all of the instances of a class share the same characteristics
Interoperability	The capacity of a product or system to function with other existing or future products or systems, without restriction or requiring further modification
Mapping	Procedure allowing for the matching of corresponding data between two <i>data models</i> * (in particular between data located in different sources and between the concepts used to represent them in ontologies)
Metadata	Information about data, which is necessary to understand and interpret these data
Ontology	A formal representation of the concepts of a domain of interest and the relationships between these concepts
Open access	The making available on line of digital content, which can itself be open access (for example a Creative Commons licence), or subject to various intellectual property restrictions
Semantic integration	The process of inter-relating information from diverse sources
Semantic web initiative	A broadly scoped effort led by the world wide web consortium (W3C) to enable software systems to easily find, analyse, share, and integrate web content. The W3C has created many technology specifications for extending web content, including via ontologies using the web ontology language (OWL)
Semantics	The study of meaning. It focuses on the relation between signifiers, like words, phrases, signs, and symbols, and what they stand for, their denotation
Standard	A published reference whose diffusion and utilization are widespread and recognized by a large proportion of those working in the domain
Thesaurus	Allows for the organization and structuring of a controlled vocabulary based on semantic rules of hierarchy, association, or equivalence
Workflow	The coordinated conceptualization and automated control of the ensemble of tasks to be accomplished by the various actors implicated in a given operational process

Table tirée de Garnier *et al.* 2016

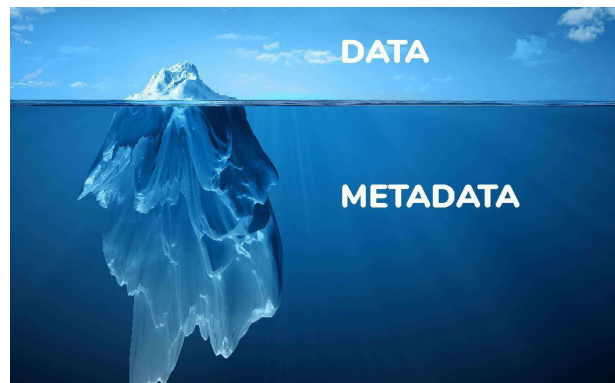
The image features a white background with decorative curved lines in shades of green and blue. One set of lines is in the top-left corner, and another is in the bottom-right corner. The text "Merci de votre attention" is centered in the middle of the page.

Merci de votre attention

Annexes

Les **métadonnées** représentent les informations ou instructions de niveau supérieur qui décrivent le contenu, le contexte, la qualité, la structure et l'accessibilité **d'un ensemble de données spécifique**. Elles doivent contenir **toutes les informations nécessaires** et suffisantes pour permettre **l'utilisation secondaire (réutilisation) à long terme de l'ensemble de données** par le(s) chercheur(s) initial(aux), ainsi que l'utilisation par d'autres scientifiques qui n'ont pas été directement impliqués dans les efforts de recherche initiaux (Michener et al., 1997).

« The most important reason to invest time and energy in developing metadata is that **human memory is short** »





Bien établir vos besoins et vos contraintes en termes de publication des données.



Choisir l'entrepôt de données qui convient à ces besoins et contraintes.



Tenter de privilégier les entrepôts disciplinaires et thématiques.



Pour les données complexes, ne pas se restreindre à un unique entrepôt.



Ajouter des métadonnées robustes et complètes.



Utiliser au maximum des thesaurus et des standards pour décrire vos données et métadonnées.



Bien relier les données aux articles associées en y faisant référence explicitement et inversement.

A minima :



Déposer les données dans un entrepôt généraliste.



Ajouter au dépôt un Readme et un dictionnaire de données.

Annexes

F

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

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

F4. (Meta)data are registered or indexed in a searchable resource

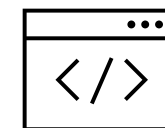
A

I

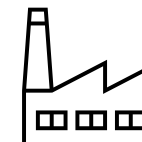
R



ID unique



Métadonnées



Entrepôts

Annexes

F

A

I

R

Accessible

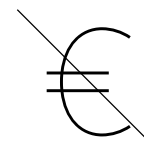
Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

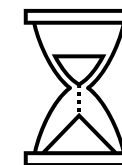
A2. Metadata are accessible, even when the data are no longer available



Gratuit



Accès libre



Durable

Annexes

F

A

I

R

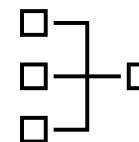
Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

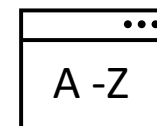
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

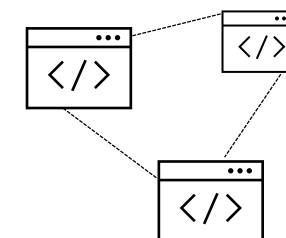
I3. (Meta)data include qualified references to other (meta)data



Standards



Vocabulaires contrôlés



Faire des liens

Annexes

F

A

I

R

Reusable

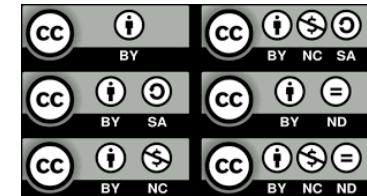
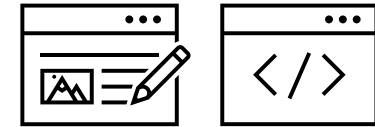
The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

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



Données et métadonnées au GBIF

eventDate	decimalLatitude	decimalLongitude	footprintWKT	scientificName	originalNameUsage
10/03/2016	43.38678	6.45929	POINT(6.45929 43.38678)	Poa bulbosa L.	Poa bulbosa L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.39907)	Gratiola officinalis L.	Gratiola officinalis L., 1753
06/05/2015	43.37602	6.47322	POINT(6.47322 43.37602)	Phillyrea angustifolia L.	Phillyrea angustifolia L., 1753
23/04/2014	43.38765	6.47164	POINT(6.47164 43.38765)	Mentha pulegium L.	Mentha pulegium L., 1753
23/04/2014	43.38689	6.45972	POINT(6.45972 43.38689)	Prunella vulgaris L.	Prunella vulgaris L., 1753
17/05/2018	43.39907	6.46759	POINT(6.46759 43.39907)	Montia arvensis Waltr.	Montia fontana subsp. chondrosperma (Fenzl) Walters
23/04/2014	43.39904	6.46744	POINT(6.46744 43.39904)	Serapias lingua L.	Serapias lingua L., 1753
10/03/2016	43.38689	6.46165	POINT(6.46165 43.38689)	Albosorus thaei (Tod.) C. Cheilanthes thaei Tod., 1886	
10/03/2016	43.38642	6.46366	POINT(6.46366 43.38642)	Cladonia mediterranea (Cladonia mediterranea P. A. Duvign. & Abbayes	
06/05/2015	43.37804	6.4734	POINT(6.4734 43.37804)	Tuberaria guttata (L.) Fc Tuberaria guttata (L.) Fourr., 1868	
13/05/2015	43.37452	6.45443	POINT(6.45443 43.37452)	Juncus effusus L.	Juncus effusus L., 1753
13/05/2015	43.37189	6.46361	POINT(6.46361 43.37189)	Carex flacca Schreb.	Carex flacca Schreb., 1771
13/05/2015	43.37512	6.46579	POINT(6.46579 43.37512)	Carex oedipostyla Duva	Carex oedipostyla Duval-Jouve, 1870
07/05/2015	43.37939	6.46166	POINT(6.46166 43.37939)	Brachypodium sylvaticum	Brachypodium sylvaticum (Huds.) P. Beauv., 1812
24/06/2016	43.38244	6.47438	POINT(6.47438 43.38244)	Lythrum salicaria L.	Lythrum salicaria L., 1753
10/03/2016	43.38659	6.46356	POINT(6.46356 43.38659)	Juniperus oxycedrus subsp. oxycedrus	Juniperus oxycedrus subsp. oxycedrus
24/06/2016	43.38267	6.45717	POINT(6.45717 43.38267)	Airoopsis tenella (Cav.) A. Airoopsis tenella (Cav.) Asch. & Graebn., 1899	
20/11/2019	43.3772	6.47288	POINT(6.47288 43.3772)	Cyperus eragrostis Lam	Cyperus eragrostis Lam., 1791
16/04/2019	43.39237	6.46849	POINT(6.46849 43.39237)	Agrostis castellana Boiss	Agrostis castellana Boiss. & Reut., 1842
06/05/2015	43.38684	6.47259	POINT(6.47259 43.38684)	Trifolium campestre Schreb.	Trifolium campestre Schreb., 1804
23/04/2014	43.38585	6.45569	POINT(6.45569 43.38585)	Moenchia erecta (L.) Gaertn.	Moenchia erecta (L.) Gaertn., B.Mey. & Scherb., 1796
13/05/2015	43.38758	6.45398	POINT(6.45398 43.38758)	Linum trigynum L.	Linum trigynum L., 1753
23/04/2014	43.38787	6.47391	POINT(6.47391 43.38787)	Isotetes duriei Bory	Isotetes duriei Bory, 1844
06/05/2015	43.37498	6.47076	POINT(6.47076 43.37498)	Crucianella angustifolia	Crucianella angustifolia L., 1753
13/05/2015	43.37351	6.45818	POINT(6.45818 43.37351)	Vincetoxicum hirundinaria	Vincetoxicum hirundinaria Medik., 1790
16/09/2014	43.38785	6.47373	POINT(6.47373 43.38785)	Agrostis castellana Boiss	Agrostis capillaris subsp. castellana (Boiss. & Reut.) O.E.



GBIF
Global Biodiversity
Information Facility

Métadonnées



Données



Darwin Core

GBIF Metadata Profile – How-to Guide

Version 2.1.

Term name	Description
alternateIdentifier	It is a Universally Unique Identifier (UUID) for the EML document and not for the dataset. This term is optional. A list of different identifiers can be supplied. E.g. 619a4b95-1a82-4006-be6a-7d0e3c9b33c5.
title	A description of the resource that is being documented that is long enough to differentiate it from other similar resources. Multiple titles may be provided, particularly when trying to express the title in more than one language (use the "xml:lang" attribute to indicate the language if not English/en). E.g. Vernal pool amphibian density data. Isla Vista, 1990-1996.
creator	The resource creator is the person or organization responsible for creating the resource itself. See section "People and Organizations" for more details.
metadataProvider	The metadataProvider is the person or organization responsible for providing documentation for the resource. See section "People and Organizations" for more details.
associatedParty	An associatedParty is another person or organization that is associated with the resource. These parties might play various roles in the creation or maintenance of the resource, and these roles should be indicated in the "role" element. See section "People and Organizations" for more details.
contact	The contact field contains contact information for this dataset. This is the person or institution to contact with questions about the use, interpretation of a data set. See section "People and Organizations" for more details.

Darwin Core List of Terms

Occurrence

dwcassociatedMedia | dwcassociatedOccurrences | dwcassociatedReferences | dwcassociatedTaxa | dwcbehavior | dwccasts | dwccatalogNumber | dwccatalogNumberNumeric | dwccatalogOfEstablishment | dwccatalogOfEstablishmentMeans | dwcgeoreferenceVerificationStatus | dwcindividualCount | dwcindividualID | dwcifeStage | dwcoccurrenceAttributes | dwcoccurrenceDetails | dwcoccurrenceID | dwcoccurrenceRemarks | dwcoccurrenceStatus | dwcorganismQuantity | dwcorganismQuantityType | dwcotherCatalogNumbers | dwcpathway | dwcrecordedBy | dwcrecordedByID | dwcrecordNumber | dwcreproductiveCondition | dwcsex | dwcsexuality

Organism

dwcassociatedOrganisms | dwcorganismID | dwcorganismName | dwcorganismRemarks | dwcorganismScope | dwcpreviousIdentifiers

Material Entry

dwcassociatedSequences | dwcdisposition | dwcmaterialEntryID | dwcmaterialEntryRemarks | dwcpreparations | dwcverbatimLabel

Material Sample

dwcmaterialSampleID

Event

dwcday | dwcEarliestDateCollected | dwcendDayOfYear | dwcEndTimeOfDay | dwceventAttributes | dwcendDate | dwceventId | dwceventRemarks | dwceventType | dwcfieldNotes | dwcfieldNumber | dwchabitat | dwclatestdDateCollected | dwcmonth | dwcparentEventID | dwcsampleSizeUnit | dwcsampleSizeValue | dwcsamplingEffort | dwcsamplingProtocol | dwcstartDayOfYear | dwcstartTimeOfDay | dwcverbatimEventDate | dwcyear

Location

dwccontinent | dwccordinatePrecision | dwccordinateUncertaintyInMeters | dwccountry | dwccountryCode | dwccountry | dwcdecimalLatitude | dwcdecimalLongitude | dwcfootprintSpatialFit | dwcfootprintSRID | dwcfootprintWKT | dwcgeodeticDatum | dwcgeoreferencedBy | dwcgeoreferencedDate | dwcgeoreferenceProtocol | dwcgeoreferenceRemarks | dwcgeoreferenceSources | dwchigherGeography | dwchigherGeographyID | dwcistland | dwcislandGroup | dwclocality | dwclocationAccordingTo | dwclocationAttributes | dwclocationID | dwclocationRemarks | dwcmaximumDepthInMeters | dwcmaximumDistanceAboveSurfaceInMeters | dwcmaximumElevationInMeters | dwcminimumDepthInMeters | dwcminimumDistanceAboveSurfaceInMeters | dwcminimumElevationInMeters | dwcmunicipality | dwcpointRadiusSpatialFit | dwcSamplingLocationID | dwcSamplingLocationRemarks | dwcstateProvince | dwcverbatimCoordinates | dwcverbatimCoordinateSystem | dwcverbatimDepth | dwcverbatimElevation | dwcverbatimLatitude | dwcverbatimLocality | dwcverbatimLongitude | dwcverbatimSRID | dwcverticalDatum | dwcwaterBody

Un standard de données : Le Darwin Core

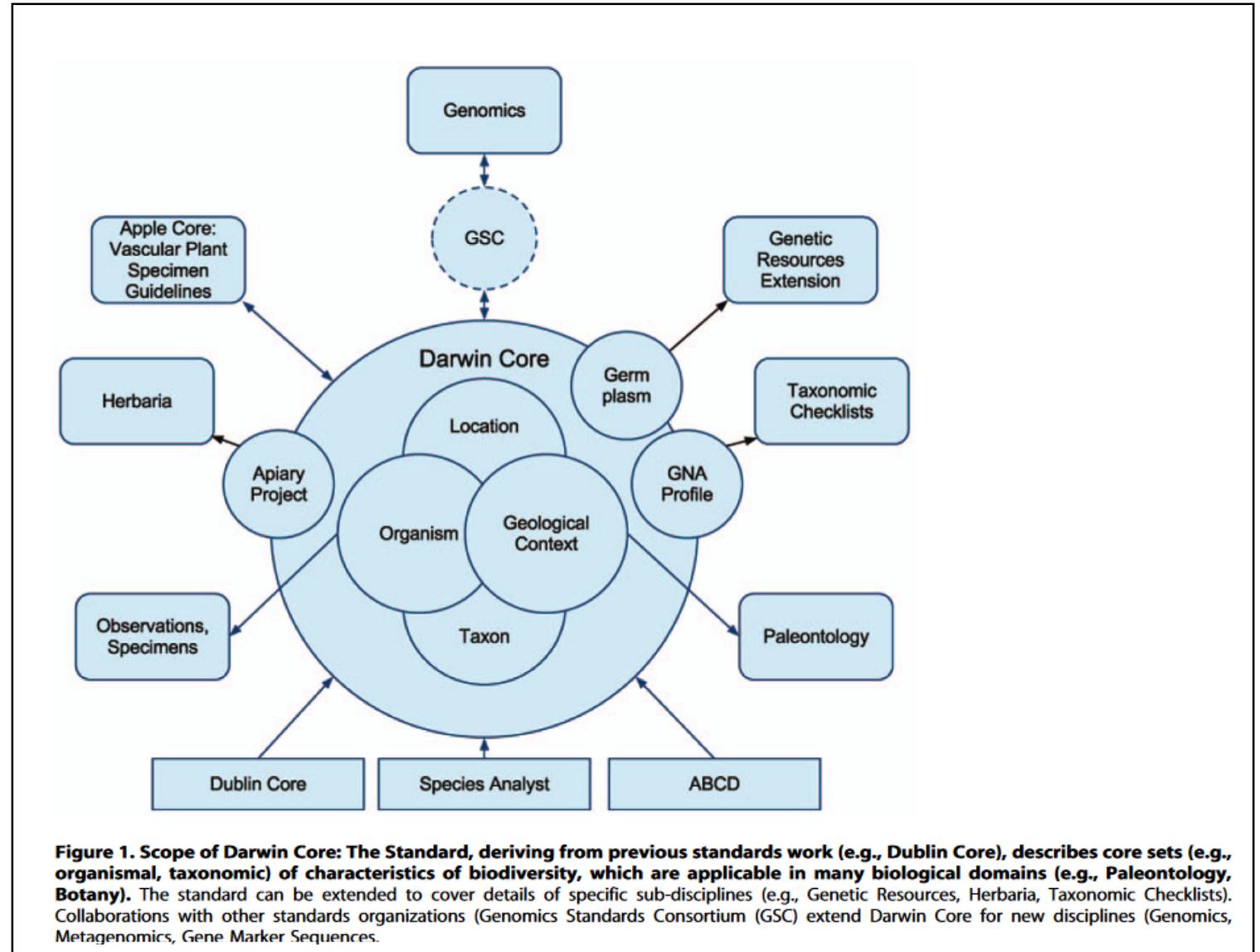


Biodiversity
Information
Standards
TDWG

Darwin Core

Le développement du Darwin Core s'étend sur plus d'une décennie.

Il a été adopté par l'éditeur Pensoft comme référence à privilégier pour la diffusion de données taxonomiques pour plusieurs de ces revues (ZooKeys ; Biodiversity Data Journal ; etc.).



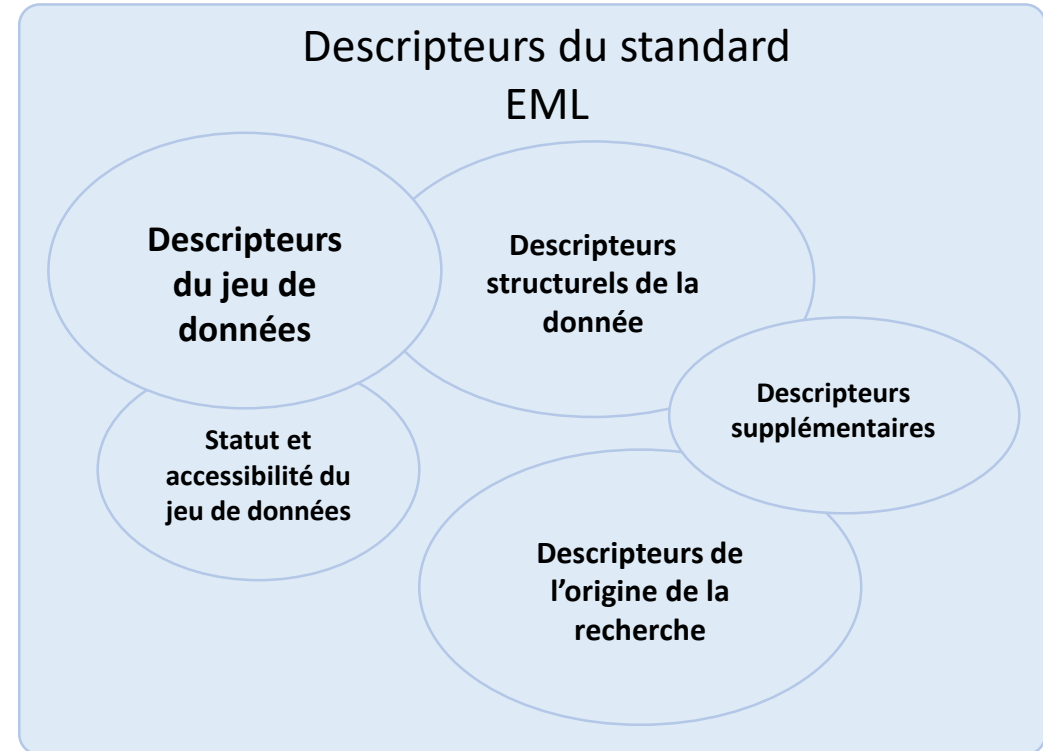
(Wieczorek et al., 2012)

Un standard de métadonnées : l'EML



Jeux de données
-
avec une structure propre

+



Source (Garnier et al., 2016)

Permet de décrire n'importe quel jeu de données en écologie.

La nature et le contenu des champs décrivant les données sont laissés libres, pourvu qu'ils soient clairement définis.

N'a pas d'impact sur la structure des données en elles-mêmes.