

Success stories pour la publication de données “FAIR” dans le cadre du projet ANR de science ouverte SO-DRIIHM

Emilie Lerigoleur

GEODE
GEOGRAPHIE DE L'ENVIRONNEMENT

 emilie-lerigoleur

 0000-0002-0864-659X

IdRef : 253124867

 emilie.lerigoleur@univ-tlse2.fr


 **HAL**
science ouverte

Séminaire TechnicoTop, Univ. Sorbonne, 06/10/2023



TECHNIQUES AU TOP

Les 13 Observatoires Hommes-Milieux (OHM) du LabEx DRIIHM



◆ Dispositif CNRS-INEE depuis 2007, financé par un LabEx depuis 2012

- > ~ 1250 pers. / 200 labos / 133 org.
- > 866 projets annuels/thèses
- > 73 champs disciplinaires
- > données hétérogènes “longue traîne”
- > Personnel support: ~10 pers.= ~2 ETP

◆ Une infrastructure de données construite brique par brique : géocatalogue geodriihm, web SIG, photothèque

→ Constat actuel de faible partage et diffusion des données :-)

Pardo 2016

Le projet ANR SO-DRIIHM

- ★ ANR Appel Flash SO-DATA 2019
- ★ 8 partenaires / 98 k€
- ★ Octobre 2020 >>>> mars 2024

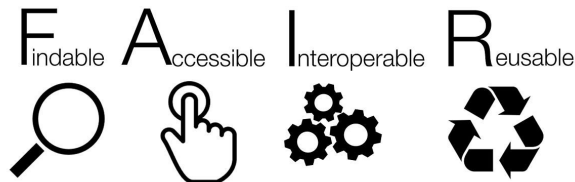
2 grands objectifs

- 1. Promotion de la science ouverte et accompagnement** auprès de la communauté DRIIHM
- 2. Co-construction d'une e-infrastructure** ergonomique et interopérable, intégrant les outils existants (co-design)



Accompagner les scientifiques :

- Plans de Gestion de Données pour les projets des OHM
- Diffusion de données aussi FAIR que possible
- Data paper



Les principes FAIR

Les chercheurs s'appuient sur les connaissances scientifiques antérieures, notamment sur les résultats publiés dans les articles scientifiques. La reproductibilité des résultats, ainsi que leur croisement, ne sont cependant envisageables qu'avec des données originelles et leurs conditions d'obtention. C'est pourquoi la science ouverte vise à faciliter l'accès aux publications scientifiques et aux données de la recherche. Cette facilitation s'accompagne d'un certain nombre de mesures pour rendre les données scientifiques facilement découvrables, accessibles, interopérables et réutilisables. Ce sont les principes FAIR : Findable, Accessible, Interoperable, Reusable.

Findable	PID	Métadonnées	Métadonnées avec PID	Entrepôt de données
Accessible	Protocole standard	Protocole libre et ouvert	Authentification	Accès aux métadonnées
Interoperable	Vocabulaire	Vocabulaire FAIR	Métadonnées liées	
Reusable	Métadonnées avec attributs	Licence	Provenance	Standards de la communauté

Focus sur la FAIRisation des jeux de données des OHM

14 dépôts dans des entrepôts de données

Données en écologie et environnement

6 datasets dans Data.InDoRES



1 dataset dans PNDB + 1 data paper



4 datasets dans GBIF (+1 en cours)

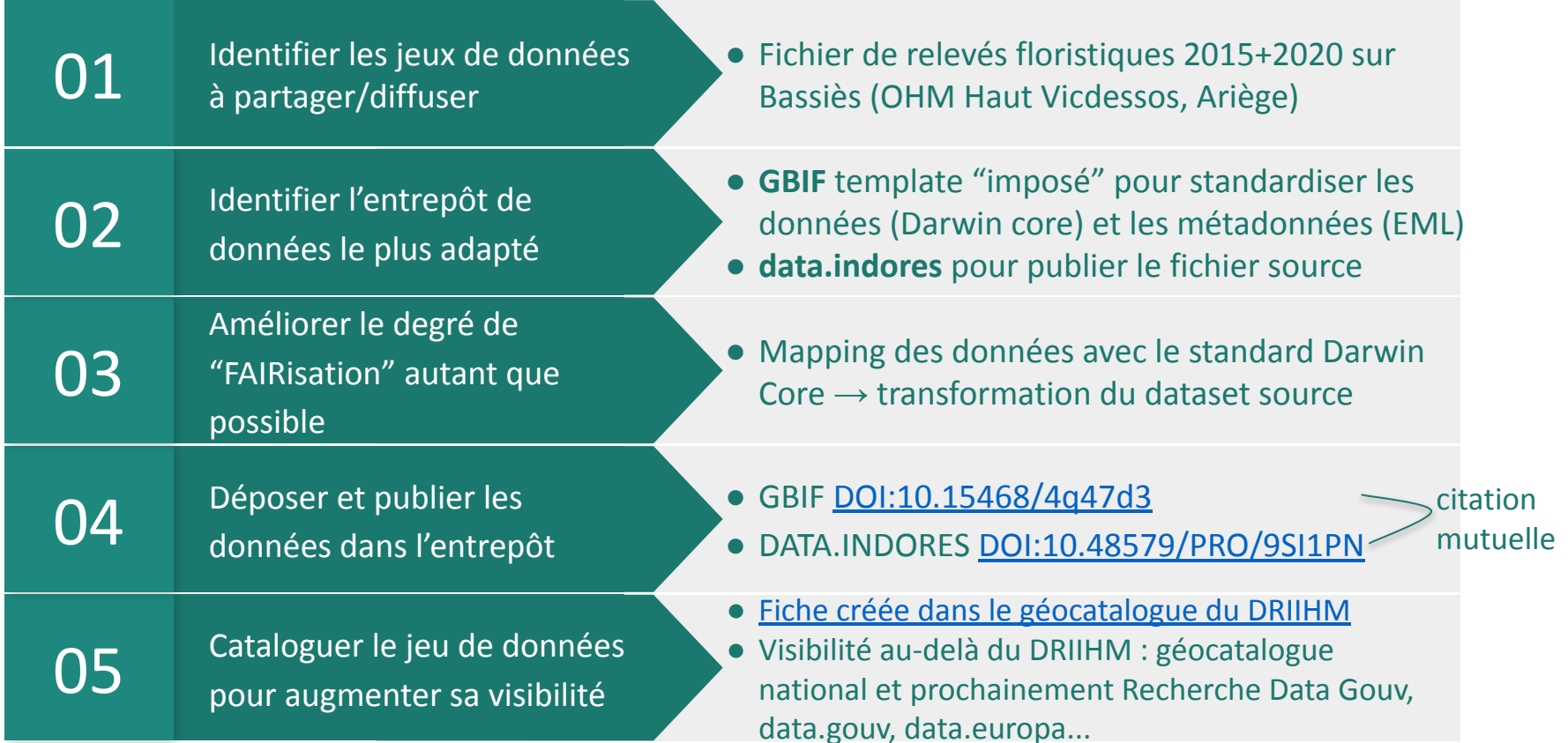


Données SHS

3 datasets dans Nakala



Exemple 1 : FAIRisation d'un fichier de relevés floristiques



citation mutuelle

Ex. 1 / Illustration de l'étape 3 de transformation du jeu de données source avec standardisation selon Darwin Core



TableurBotaHarmonisé_Base.ods - LibreOffice Calc

Jeu de données d'origine

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1																PASTSERV.2020									
2					AD.PS1.csv	AD.PS19.csv	AD.PS2.csv	AD.PS20.csv	AD.PS21.csv	AD.PS22.csv	AD.PS23.csv	AD.PS4.csv	AD.PS5.csv	AD.PS6.csv	AD.PS7.csv	AD.PS9.csv	AD.PS10.csv	AD.PS11.csv	AD.PS12.csv	AD.PS13.csv	AD.PS14.csv	AD.PS15.csv	AD.PS16.csv		
3					TBF1	TBF8	TBF2	TBF9	TBF10	TBF11	TBF12	TBF3	TBF4	TBF5	TBF6	TBF7	TMF2	TMF3	TMF4	TMF5	TMF6	TMF7	TMF8		
4	Inventaire	Nom valid	NOMBDD	Nb occurrence	BF	BF	BF	BF	BF	BF	BF	BF	BF	BF	BF	BF	MF	MF	MF	MF	MF	MF	MF	MF	
5	PastServ	Abies alba		19	0.03	14		14				3	0.03			14	32	3	32	57	14	32	57	32	14
6	PastServ	Acer campestre		2	14									14											
7	PastServ	Acer monspessulanum		1	0.3																				
8	PastServ	Aceras anthr	Orchis anthr	1	0.03																				
9	SebHau	Achillea millefolium		1																					
10	SebHau	Agrostis capillaris		4																					
11	SebHau	Agrostis stolonifera		2																					
12	PastServ	Ajuga reptans		2	0.03									0.03											
13	PastServ	Alchemilla alpina		6							0.03														
14	SebHau	Allium schoenoprasum		1																					
15	PastServ	Anemone sp		1		0.03																			



event_ipt_template_v2_example_data.xlsx - LibreOffice Calc

Template GBIF

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	eventID	samplingProtocol	samplingEffort	sampleSizeValue	sampleSizeUnit	eventDate	eventTime	startDayOfYear	eventRemarks	country	countryCode	locality	locationID	decimalLatitude	decimalLongitude	geodeticDatum
2	994-tr009-s00	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2012-10-11 09:28:02Z		284	No occurrence	Israel	IL	Sde boker	tr009-s00	30.8539	34.7692	WGS84
3	3502-tr056-s6	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2015-10-19 12:25:02Z		291		Israel	IL	Nahal Kof	tr056-s6	31.29077	34.81112	WGS84
4	3502-tr056-s9	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2015-10-19 12:25:02Z		291		Israel	IL	Nahal Kof	tr056-s9	31.28945	34.81151	WGS84

2 feuilles : tableau des échantillonnages et tableau des occurrences par échantillonnage

Ex. 1 / Illustration de l'étape 4 de dépôt dans le GBIF via l'outil IPT

The screenshot shows the IPT web interface for a resource named 'pastserv_mode-respyr_vegetation_data'. The interface is divided into several sections:

- Overview:** 'pastserv_mode-respyr_vegetation_data'. Description: 'Resource overview. Upload or connect to source data, map data to Darwin Core, complete metadata, publish the resource and register it with the GBIF network.' Includes a 'Delete' button.
- Source Data:** A section for managing source data files and SQL sources. It includes a 'Database' dropdown, a 'Connect' button, and a list of files with their sizes, row counts, and columns. Two files are listed: 'associated_occurrences_pastserv_mode-respyr_20220503 [file]' and 'sampling_events_pastserv_mode-respyr_20220503 [file]'. Each file has an 'Edit' button.
- Darwin Core Mappings:** A section for mapping source data to Darwin Core terms. It includes a 'Darwin Core Event' dropdown, an 'Add' button, and a list of mappings. Two mappings are shown: 'Darwin Core Event' (11 terms mapped to 'sampling_events_pastserv_mode-respyr_20220503') and 'Darwin Core Occurrence' (12 terms mapped to 'associated_occurrences_pastserv_mode-respyr_20220503'). Each mapping has an 'Edit' button.
- Metadata:** A section for entering resource metadata. It includes a 'Choisir un fichier' button, an 'Auc... oisi' button, and a text area for 'Your resource metadata.'

➤ Import des 2 templates Darwin core

➤ Mapping avec le standard Darwin core (validateur)

➤ Saisie des métadonnées et licence ouverte CC BY



Ex. 1 / Résultat : un dataset en open data entreposé et catalogué

▼ Jeu de données standardisé
Darwin core et EML

Vegetation surveys in mountainous area (Pyrenees, France)

Published by UMR 5502 Géographie de l'Environnement (CNRS/Université Toulouse 2)
Mazier F • Plancher C • Haunold S • Bonnet C • Belly H • Houet T

1,757 OCCURRENCES 3 CITATIONS

Global Biodiversity Information Facility

The data file contains absolute cover percentage of plant species from vegetation survey in the Vicedosses mountainous area (Pyrenees, France). The data have been collected in 2015 and 2020 within the framework of two projects: ANR-10-JCJC-1804 MODE-RESPYR (Modeling Past and future land cover changes in the Pyrenees) and the PASTSERV project funded by the Observatoire Hommes-Milieux Pyrenees Haut Vicedosses (Labex DRIIHM ANR-11-LABX0010). All observed vascular plant species were listed and the cover of each species was estimated using the seven degrees of the Braun-Blanquet scale (r, +, 1, 2, 3, 4, 5). These codes were then converted to absolute percentage cover (van der Maarel 1979).

Project ID: ANR-10-JCJC-1804/ANR-11-LABX0010
Publication date: September 14, 2022
Metadata last modified: September 14, 2022
Hosted by: GBIF France
Licence: CC BY 4.0
How to cite: DOI 10.15468/4q47d3

1,757 Occurrences 100% With taxon match 100% With coordinates 100% With year

1,757 GEOREFERENCED RECORDS

125 EVENTS

Event ID	Event date	Sampling protocol	Occurrence count
181F	1 January 2020	DOI 10.1007/97800032021 and described in Gillet F 2009	51 (0 absent)
181S	1 January 2020	DOI 10.1007/97800032021 and described in Gillet F 2009	37 (0 absent)

▼ Jeu de données source

OHM Pyrénées Haut Vicedosses (CNRS)

data.InDoRES > DRIIHM (Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes-Milieux) > OHM Pyrénées Haut Vicedosses >

Vegetation surveys in mountainous area (Pyrenees, France)

Version 1.0

Mazier, Florence; Plancher, Clara; Haunold, Sébastien; Bonnet, Colline; Belly, Hugo; Houet, Thomas, 2022, "Vegetation surveys in mountainous area (Pyrenees, France)", <https://doi.org/10.48579/PRO/9S11PN>, data.InDoRES, V1

Citer le jeu de données

Pour en savoir plus, consultez le document Data Citation Standards.

Access Dataset

Contact Owner Share

Statistiques du jeu de données

4 Téléchargements

Description

The data file contains absolute cover percentage of plant species from vegetation survey in the Vicedosses mountainous area (Pyrenees, France). The data have been collected in 2015 and 2020 within the framework of two projects: ANR-10-JCJC-1804 MODE-RESPYR (Modeling Past and future land cover changes in the Pyrenees) and the PASTSERV project funded by the Observatoire Hommes-Milieux Pyrenees Haut Vicedosses (Labex DRIIHM ANR-11-LABX0010). All observed vascular plant species were listed and the cover of each species was estimated using the seven degrees of the Braun-Blanquet scale (r, +, 1, 2, 3, 4, 5). These codes were then converted to absolute percentage cover (van der Maarel 1979). The version of the data file formatted according to the Darwin Core standard has been uploaded to GBIF (<https://doi.org/10.15468/4q47d3>). This file is the original raw data file. (2022-05-03)

Le fichier de données contient le taux de couverture absolue des espèces végétales provenant de relevés de végétation dans la vallée montagnarde du haut Vicedosses (Pyrénées, France). Les données ont été collectées en 2015 et 2020 dans le cadre de deux projets : ANR-10-JCJC-1804 MODE-RESPYR (Modélisation des

▼ Fiche de métadonnées du géoDRIIHM

OHM Pyrénées Haut Vicedosses Recherche Visualiser S'identifier Francais

Retour Précédent Suivant

Télécharger Mode attachage

Vegetation surveys in mountainous area (Pyrenees, France)

Aperçu

The data file contains absolute cover percentage of plant species from vegetation survey in the Vicedosses mountainous area (Pyrenees, France). The data have been collected in 2015 and 2020 within the framework of two projects: ANR-10-JCJC-1804 MODE-RESPYR (Modeling Past and future land cover changes in the Pyrenees) and the PASTSERV project funded by the Observatoire Hommes-Milieux Pyrenees Haut Vicedosses (Labex DRIIHM ANR-11-LABX0010). All observed vascular plant species were listed and the cover of each species was estimated using the seven degrees of the Braun-Blanquet scale (r, +, 1, 2, 3, 4, 5). These codes were then converted to absolute percentage cover (van der Maarel 1979). The version of the data file formatted according to the Darwin Core standard has been uploaded to GBIF (<https://doi.org/10.15468/4q47d3>). The original raw data file is accessible in the CNRS-INEE data.indores repository <https://doi.org/10.48579/PRO/9S11PN>.

Étendue spatiale

Téléchargements et liens

- Projet OHM PASTSERV 2020 [Ouvrir le lien](https://www.driihm.fr/projets/projets-2020/details/149/2266)
- Source file on data.indores.fr [Ouvrir le lien](https://data.indores.fr/dataset.xhtml?persistentId=doi:10.48579/PRO/9S11PN)
- File based on Darwin core standard published on GBIF website [Ouvrir le lien](#)

Événements

Publication 15-09-2022 02:00

Form par

Mis à jour : il y a 5 jours

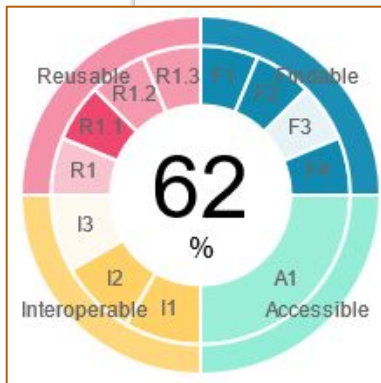
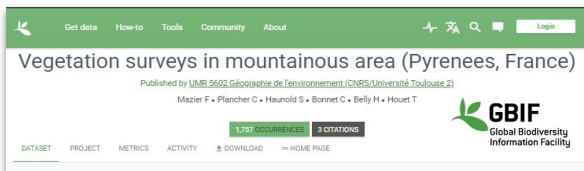
Ex. 1 / Evaluation du niveau de FAIRitude (FAIRness)

outil utilisé : [F-UJI](#)
FAIR assesment tool



▼ Jeu de données standardisé

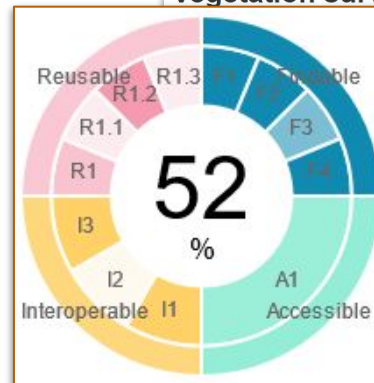
Darwin core et EML [DOI:10.15468/4q47d3](https://doi.org/10.15468/4q47d3)



	Score earned:	Fair level:
Findable:	6 of 7	moderate
Accessible:	1 of 3	initial
Interoperable:	3 of 4	moderate
Reusable:	5 of 10	moderate

▼ Jeu de données source

[DOI:10.48579/PRO/9SI1PN](https://doi.org/10.48579/PRO/9SI1PN)



	Score earned:	Fair level:
Findable:	6.5 of 7	advanced
Accessible:	1 of 3	initial
Interoperable:	3 of 4	moderate
Reusable:	2 of 10	initial

Event ID	Event date	Sampling protocol	Occurrence count
101F	1 January 2020	10.15007/BF00003021 and described in Gillet F. 2000	51 (0 absent)
101S	1 January 2020	10.15007/BF00003021 and described in Gillet F. 2000	37 (0 absent)

(Labex DRIIHM ANR-11-LABX0010). All observed vascular plant species were listed and the cover of each species was estimated using the seven degrees of the Braun-Blanquet scale (r, +, 1, 2, 3, 4, 5). These codes

NB: Attention, utiliser avec précaution ces outils d'évaluation qui font des vérifications automatisées. Tous les contenus ne sont pas finement vérifiés !

Exemple 2 : Les données de l'enquête en ligne "Partagez votre expérience de partage de données !" (juin 2021)

UNIVERSITÉ Toulouse Clean Jaurès **SONDAGE**

Load unfinished survey Language: English ▾

Share your experience of data sharing!

DRIIHM - Interdisciplinary Research Device on Human-Environment Interactions

We are asking you because you are involved in at least one of the Human-Environment Observatories (OHM) of the LabEx DRIIHM.

Data sharing is a major challenge for the scientific community today. Sharing is both making available the data we produce, but also accessing data useful to our research.

We are currently designing a web platform at the DRIIHM level in order to facilitate data and to accompany you in this process. This project is led by the engineers/researchers of the DATADRIIHM group within the framework of the DRIIHM ANR Open Science project (SO-DRIIHM).

It's obviously essential that this platform adapts to your practices and meets your needs. We would like you to share your experience of data sharing by answering the questionnaire via <https://enquetes.univ-tlse2.fr/index.php/968188?lang=en>

Your participation requires about 15 minutes. It is essential to the success of the project.

Deadline to answer: July 9th!

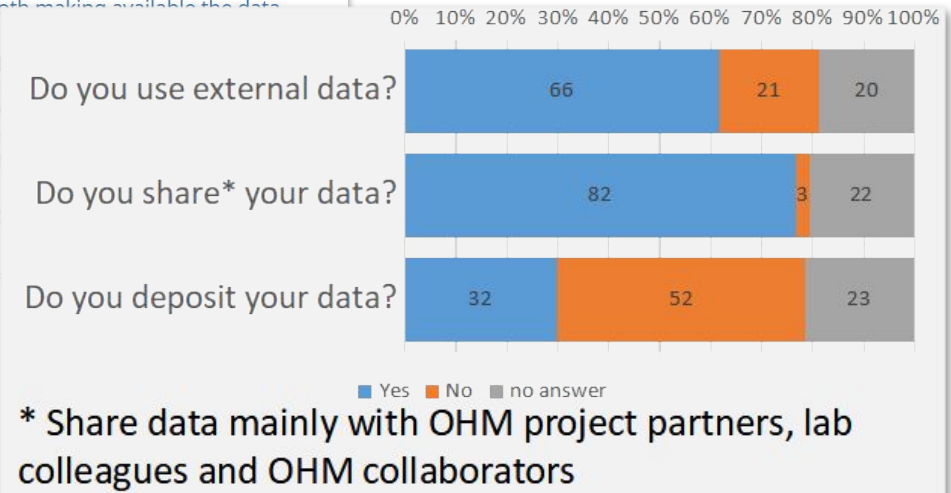
We warmly thank you for your answers and for your involvement in the service of the scientific community.

Sincerely
The DATADRIIHM group



There are 34 questions in this survey.

To continue, please first accept our questionnaire policy
[Show policy](#)

- ★ 107 répondants
- ★ 34 questions
- ★ anonymisation



Ex. 2 : Comment rendre autant que possible FAIR des données d'enquête ?

01	Identifier les jeux de données à partager/diffuser	<ul style="list-style-type: none"> Données de l'enquête "Partagez votre expérience de partage de données !" + questionnaire FR/EN 	 
02	Identifier l'entrepôt de données le plus adapté	<ul style="list-style-type: none"> Nakala (IR Huma Num) 	
03	Améliorer le degré de "FAIRisation" autant que possible	<ul style="list-style-type: none"> FAIRiser les métadonnées autant que possible via Nakala F et A OK pour les données, R a minima (résumé + trame), I → inadapté ?! 	
04	Déposer et publier les données dans l'entrepôt	<ul style="list-style-type: none"> questionnaire: DOI:10.34847/nkl.bd39hw08 données: DOI:10.34847/nkl.db933xfo 	<p>lien entre les 2</p>
05	Cataloguer le jeu de données pour augmenter sa visibilité	<ul style="list-style-type: none"> Fiche dans le géocatalogue geodriihm Visibilité : géocatalogue national et prochainement Recherche Data Gouv, data.gouv, data.europa... 	

Ex. 2 : Importance de lier les dépôts et de créer des collections



Rechercher dans NAKALA... | Q

Fr Se connecter

Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes-Milieus (DRIIHM)

DOI : 10.34847/nkl.add0599 Publique

Créée le 13/04/2021

Lauréat de la deuxième vague de l'appel à projet Laboratoire d'Excellence (LabEx) dans le cadre du programme "Investissements d'avenir", le LabEx DRIIHM, Dispositif de Recherche Interdisciplinaire sur les Interactions Hommes-Milieus, regroupe les Observatoires Hommes-Milieus (OHM, au nombre de 13), outils créés par le CNRS-INEE en 2007. Ce LabEx mis en place initialement pour 8 ans (février 2012 ...

Voir plus

Afficher les détails de la collection

Filter par :	Type	Licence	Année de création	Trier par :	Date de dépôt (décroissante)
				Auteurs	Date de dépôt
<input checked="" type="checkbox"/>	Données de l'enquête "Partagez votre expérience de partage de données !" (juin 2021) - projet ANR SO-DRIIHM			Pascal Dayre et al.	04/07/2023
<input checked="" type="checkbox"/>	Questionnaire de l'enquête "Partagez votre expérience de partage de données !" (juin 2021) - projet ANR SO-DRIIHM			Pascal Dayre et al.	04/07/2023

➤ Création d'une **collection** pour le Dispositif DRIIHM pour rassembler tous les dépôts

➤ Lien entre les 2 dépôts : questionnaire (trame) et les réponses anonymisées des enquêtés

Ex. 2 : Décrire le plus finement possible les données à l'aide des métadonnées proposées par Nakala

Rechercher dans NAKALA... Fr Se connecter

Données de l'enquête "Partagez votre expérience de partage de données !" (juin 2021) - projet ANR SO-DRIIHM FR | EN
[DOI:10.34847/nkl.db933xfo](https://doi.org/10.34847/nkl.db933xfo)
 version 2 Publique

Auteurs : Pascal Dayre, Emilie Lerigoleur et Bertrand Morandi

FR | EN

Le partage des données est aujourd'hui un défi. Le partage c'est à la fois accéder à des données, c'est aussi mettre à disposition les données que l'on produit.
 L'enquête vise à mieux comprendre les pratiques d'utilisation, de partage et de dépôt de données, pour mieux co-concevoir la future e-infrastructure DATA-DRIIHM.
 L'enquête a utilisé un questionnaire en ligne, basé sur l'application Lime...

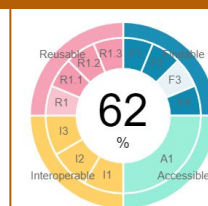
Voir plus

Fichiers	Visualisation
DRIIHM_enquete_pl...	Avez-vous déjà déposé vos données sur un site en ligne ? Pourquoi ?
DRIIHM_enquete_pl...	os données
DRIIHM_enquete_pl...	taire
	e] Pourquoi les avez-vous déposées ?
	Est-ce en lien avec un article ?

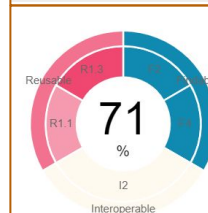
➤ Ici en 2 langues : français + anglais

Critères FAIR "classiques"

outil utilisé : F-UJI
FAIR assesment tool



	Score earned:	Fair level:
Findable:	6 of 7	moderate
Accessible:	1 of 3	initial
Interoperable:	4 of 4	advanced
Reusable:	4 of 10	initial



	Score earned:	Fair level:
Findable:	3 of 3	advanced
Accessible:	Not tested in this metrics	
Interoperable:	0 of 1	incomplete
Reusable:	2 of 3	initial

Critères FAIR "SHS"

Discussion - conclusion sur la diffusion de données aussi FAIR que possible



- ★ Les données « FAIR » sont par essence facilement trouvables, accessibles, interopérables et réutilisables, cool, non ? Mais les données de la recherche (hors exception RGPD, secret...) sont-elles vraiment toutes “ouvrables” et “réutilisables” ?
- ★ Quelques bonnes pratiques découvertes “par expérience” :
 - Identifier les données à diffuser le plus tôt possible
 - anticiper le choix de l’entrepôt de données : attention, déposer des données dans un entrepôt ne veut pas dire rendre FAIR à 100% des données, c’est déjà bien mais souvent insuffisant !
 - anticiper la structure (si possible standardisée) des données
 - documenter dans le Plan de Gestion de Données les métadonnées décrivant ces données, à l’aide si possible de métadonnées standardisées
- ★ “FAIR” # open : transparence le plus tôt possible sur les conditions d’accès / utilisation des données
 - importance du choix de la licence
 - importance du choix de l’entrepôt
 - importance de l’accompagnement dans la démarche de « FAIRisation » des données
- ★ Des constats et questionnements :
 - Besoins de RH, de formations pragmatiques, d’outils informatiques conviviaux et simples
 - Quid du coût financier de la science ouverte ? de la FAIRisation des données ?
 - Quid du coût énergétique de la science ouverte ? Quelle démarche écoresponsable préconiser ?

Suite et perspectives du projet ANR SO-DRIIHM

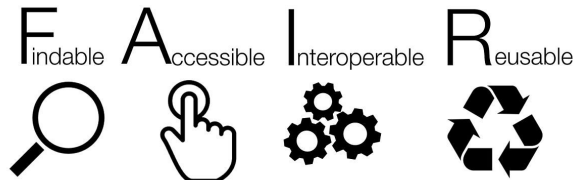
2 grands objectifs

1. **Promotion de la science ouverte et accompagnement** auprès de la communauté DRIIHM
2. **Co-construction d'une e-infrastructure** ergonomique et interopérable, intégrant les outils existants (co-design)



Vers la création d'un portail "aiguilleur de la science ouverte"

- Aider la recherche, les dépôts et la valorisation des données
- Aider à la FAIRisation des données
- Aider au support des données pour les communautés de support



Merci pour votre attention

Des questions ?

Quelques ressources et références

→ Les productions dans le cadre de SO-DRIIHM dans HAL : <https://hal.science/search/index?q=so-driihm>

Merci à tous les participants au projet : Fanny Arnaud, Birane Cissé, Pascal Dayre, Priscilla Duboz, Arnaud Jean-Charles, Iwan Le Berre, Mathieu Massaviol, Corinne Pardo, Jean-Claude Raynal, Marie-Laure Trémélo

- + les stagiaires et contractuels: Ghita, Chloé, Louis...
- + tous les chercheurs que l'on a accompagnés !

<https://hal.science/hal-04128197>

Success stories of FAIR data publication on behalf of the ANR SO-DRIIHM Open Science project

E. Lerigoleur^{1*}, L. Bouveret², A. Drogé³, J.-L. Jung⁴, I. Le Berre⁵, M. Massaviol⁶, F. Mazier⁷, P. L. Nélige⁸, A. Niang⁹, C. Pardo⁹, M.-L. Trémélo⁹

¹ORIS - IRECO (UMR0062), ²ORIS - IRECO (UMR0062), ³ORIS - IRECO (UMR0062), ⁴ORIS - IRECO (UMR0062), ⁵ORIS - IRECO (UMR0062), ⁶ORIS - IRECO (UMR0062), ⁷ORIS - IRECO (UMR0062), ⁸ORIS - IRECO (UMR0062), ⁹ORIS - IRECO (UMR0062)

SO-DRIIHM project in brief

The ANR SO-DRIIHM project encourages the sharing and opening of data within, between and beyond OSHs.

- Promote Open Science and enhance Open Science practices, showing the challenges and benefits of structuring, indexing, sharing and opening research data.
- Enhance the DRIIHM existing infrastructure to a more user-friendly and interoperable infrastructure with new functionalities that will be prioritized according to users' expectations and motivations (co-design process).

What is FAIR Data? F.A.I.R. D.E.S.

- Findable:** Metadata and data should be findable for both humans and computers.
- Accessible:** Once found, users need to know how the data can be accessed.
- Interoperable:** Data needs to work with applications or workflows for analysis, storage and processing.
- Reusable:** The goal of FAIR is to optimize data reuse via comprehensive well-described metadata.

3 success stories in making data as FAIR as possible

- TRAFIC project core team:** J.-L. Jung (responsible), L. Cédric (trainee in ecology), I. Le Berre & E. Lerigoleur (data stewards), B. David & B. Maron (data experts), OSMAS & other NGOs and whale watchers.
- Wild animals project core team:** P. L. Nélige (responsible), A. Drogé & A. Niang (PhD students), I. Le Berre & B. Maron (data stewards), E. Lerigoleur (data steward).
- PASTSERV project core team:** A. Niang (responsible), S. Fouquet, C. Pardo, C. Barret, H. Bally, T. Houtot (data producers and data analyst), E. Lerigoleur (data steward).

1. Identify the datasets useful for dissemination

- The Kakile database (atollcan biodiversity in Guadeloupe)** is linked tables (xlsx format).
- 3 tables (xlsx format):** Large mammals biodiversity in Ferlo (North of Senegal), occurrence data, trapping data of large mammals in Ferlo and bird biodiversity in Ferlo (occurrence data).
- 1 table combining absolute cover percentage of plant species from vegetation survey in the Villeduane mountains area (Pyrénées, France)** (xlsx format).

2. Identify the most appropriate data repository // Check conditions

- GBIF** portal offers:
 - A national data portal in the field of biodiversity.
 - A permanent Identifier (DOI).
 - Open Access (eg. CC BY).
 - A detailed description of the data using the EML (Ecological Metadata Language) metadata standard and the possibility to specify provenance and the data dictionary (available).
 - Interoperability with other thematic portals.
 - Free of charge, many formats accepted.
- GBIF** portal offers:
 - An international biodiversity data portal.
 - A DOI.
 - A detailed description of the data using EML standard metadata.
 - Use of the Darwin Core standard (DwC) to structure and interconnect tax files with a common nomenclature or targeted data format, but this standardization simplifies the publication process, and it also makes it easy for users to discover, search, evaluate and compare datasets.
 - Interoperability with other thematic portals.
 - Free of charge, open formats.
- GBIF + Data Commons** portal offers:
 - GBIF has been chosen to publish the dataset in the standard Darwin Core format.
 - The Data Commons portal (Data Commons) has been chosen to publish the source file with a structure commonly used by the research community it offers.
 - An international data repository (Data Commons) for long-term environmental data.
 - A DOI.
 - Open Access (eg. CC BY).
 - Standardized core metadata (Darwin Core, DOI).
 - Free of charge, many formats accepted.

3. Deposit & dissemination // Documenting and cataloguing data thanks to metadata

- Default tool: Metadata (Metadata Entry Application for Review and Knowledge) hosted by ORIS** (<https://www.oris.fr/Tools/OrisMet>)
- Default tool: IPT (Integrated Publishing Toolkit) hosted by the GBIF France** (<https://www.gbif.org/fr/data-deposit>)
- Default tool: IPT (Integrated Publishing Toolkit) hosted by the GBIF France** (<https://www.gbif.org/fr/data-deposit>)

4. Check the FAIRness level of the data using F-UII tool

Net so hard to achieve! A DOI for each dataset. Guaranteed medium-term preservation. Fairly good level of data FAIRness.

Pay particular attention to the choice of data repository, which will affect the completion of the FAIR criteria.

Interested in data sharing? Interested in making your data FAIR? Would you like to contribute to the SO-DRIIHM project as a data leader for our future open science publication platform? Contact us @ osd@iris.gov.fr or erl@iris.gov.fr