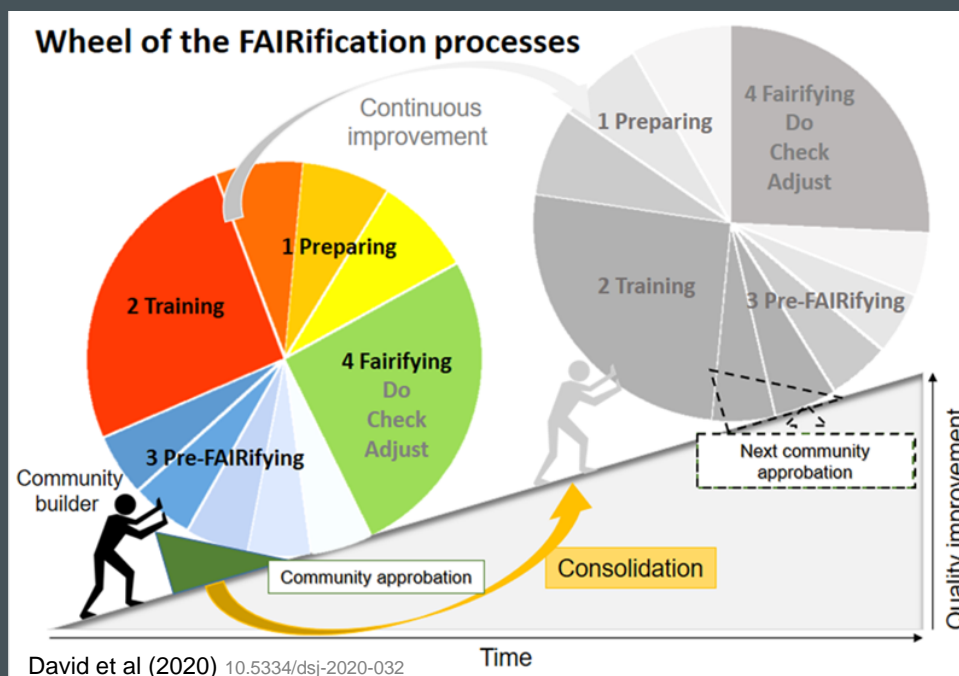


SESSION : PREPARATION A LA “FAIRISATION” DES DONNEES

SAVOIR DIAGNOSTIQUER LA GESTION DE DONNÉES COLLECTÉES/PRODUITES DANS LE CADRE D’UN PROJET

LE LONG DU CYCLE DE VIE DES DONNÉES

// ANIMATION : EMILIE LERIGOLEUR ET AMANDINE HÉNON

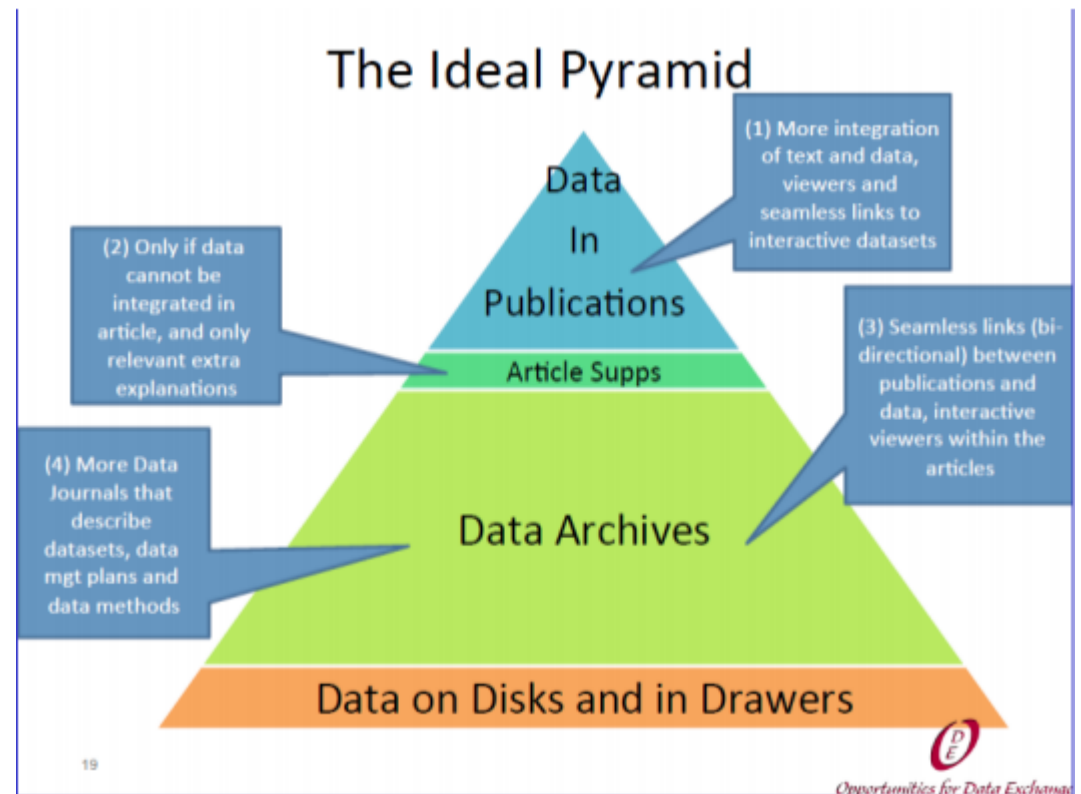
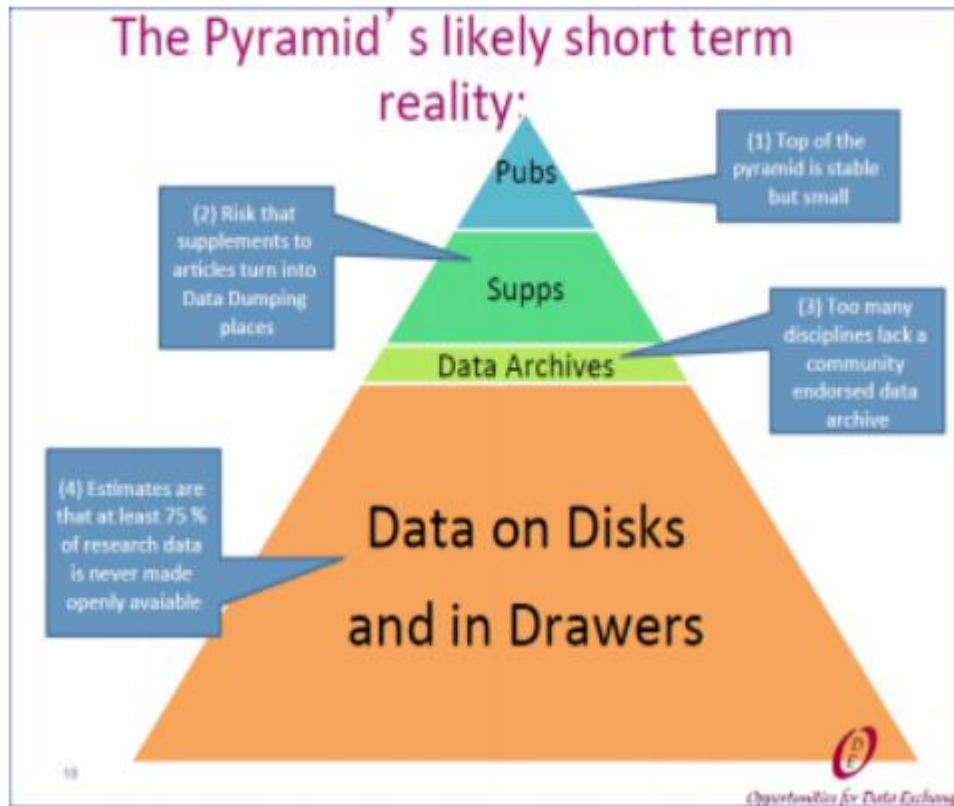


CONSTAT : PERTE DE DONNÉES

- De plus en plus de données sont produites par la Recherche mais on constate une **perte considérable** faute d'une gestion adaptée. Les raisons sont multiples : pas de sauvegarde, mauvais traitement de la donnée, départ en retraite ou changement de personnel...
Selon une étude de 2013, 17 % des données sont inaccessibles seulement 2 ans après leur publication et 80 % des données de la Recherche sont perdues après 20 ans !
Une bonne **gestion des données** est donc **primordiale**.
- Qu'est-ce que le **data management** ? C'est la prise en charge des données durant tout leur **cycle de vie**, depuis le début du projet scientifique jusqu'à sa finalisation, puis au-delà, pour permettre la réutilisation des données sur le long terme.

Source : <https://www.didaktic.fr/science-ouverte/la-gestion-des-donnees-de-la-recherche-et-leur-diffusion-quels-enjeux/>

CONSTAT : VERS UN MEILLEUR PARTAGE ET UNE MEILLEURE OUVERTURE DES DONNÉES



https://anfdonnees2017.sciencesconf.org/data/pages/20170706_dmp_puren.pdf

OBJECTIFS DE LA SESSION

- La **gestion des données** et leur **diffusion** sont les deux étapes clés de l'ouverture des données de la recherche. C'est **en amont**, avant même la collecte des données, qu'il faut envisager leur gestion pour qu'elles puissent être effectivement conservées et réutilisées sur le long terme.
- **En amont** de la FAIRisation, il y a une **phase de pré-FAIRisation** qui permet d'énoncer les objectifs de la FAIRisation et d'identifier les données et les métadonnées à rendre FAIR.
- Ce que l'on va aborder dans cette session :
 - C'est quoi le cycle de vie des données ?
 - Gestion des données : les concepts-clé
 - Anticiper la diffusion des données
 - Stratégie de la diffusion des données
 - TD Apprendre à planifier la gestion des données dans un objectif de FAIRisation
 - Vers un diagnostic de la gestion des données à FAIRiser

C'EST QUOI LE CYCLE DE VIE DES DONNÉES ? (1/5)

Définition

Le cycle de vie des données de recherche décrit le processus d'utilisation des données de leur création à la publication et à leur réutilisation ultérieure (source : INIST)



Objectifs

- permet de formaliser les différentes étapes de gestion des données
- aide à s'organiser, notamment dans un contexte multi-partenarial

L'utilisation d'un cycle de vie permet de passer d'une perspective à court terme à une perspective à long terme dans la gestion des données (sources : INIST, Doranum)

C'EST QUOI LE CYCLE DE VIE DES DONNÉES ? (2/5)

The collage features several diagrams illustrating the data lifecycle:

- Jata Lifecycle Management (DLM) | prion.com**: A circular diagram with stages: Create, Store, Use, Share, Archive, Destroy.
- DATA LIFECYCLE**: A linear flow diagram with icons for various stages.
- The Data Lifecycle: Enabling Business Growth – Axian, Inc. axian.com**: A linear flow diagram with stages: Create, Store, Use, Share, Archive, Destroy.
- Research data lifecycle | Research Dat... blogs.rtu.edu.sg**: A circular diagram with stages: Data Creation, Data Processing, Data Analysis, Data Preservation, Data Access, Data Review.
- Collect, Prepare, Publish, Maintain**: A circular diagram with four large colored circles.
- The Data Science Lifecycle**: A circular diagram with many small stages around a central lightbulb icon.
- The 16-Step Data Lifecycle for Communica... pmsucks.com**: A circular diagram with 16 small stages.
- Data Lifecycle | Library | Lancaster Uni... lancaster.ac.uk**: A circular diagram with stages: Data created/collected, Data processed, Data analyzed, Data shared, Data archived.
- What is research data management? | Libra... biblio.uottawa.ca**: A circular diagram with stages: Plan, Create, Process, Analyze, Preserve, Share, Reuse.
- UTILISATION, RÉUTILISATION, PLANIFICATION, PARTAGE, ACQUISITION, INTÉGRATION, CONTRÔLE DE QUALITE, ACCÈS, DIFFUSION, ARCHIVAGE**: A large circular diagram with these terms around the perimeter and **VALORISATION DES DONNÉES** in the center.
- Planification, Collecte, Analyse, Documentation, Stockage, Conservation, Réutilisation, Exposition**: A circular diagram with these stages and the text **INIST CNRS** at the bottom.

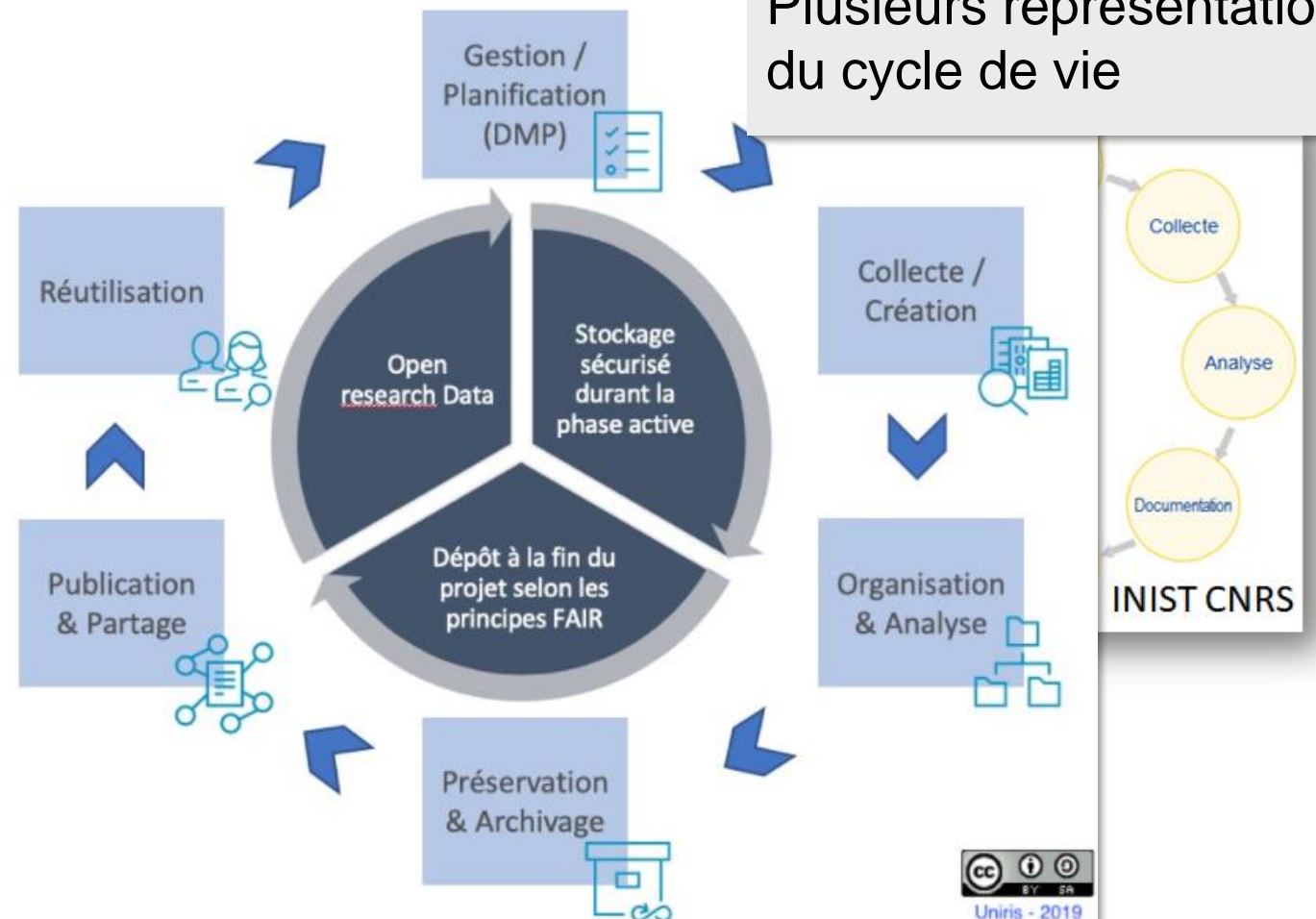
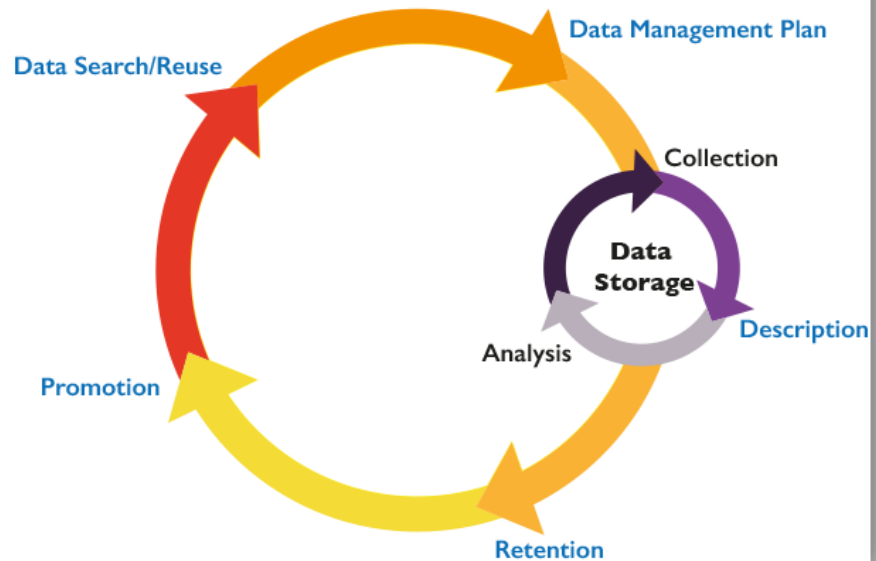
Plusieurs représentations du cycle de vie

C'EST QUOI LE CYCLE DE VIE DES DONNÉES ? (3/5)

Cycle de vie des données de recherche

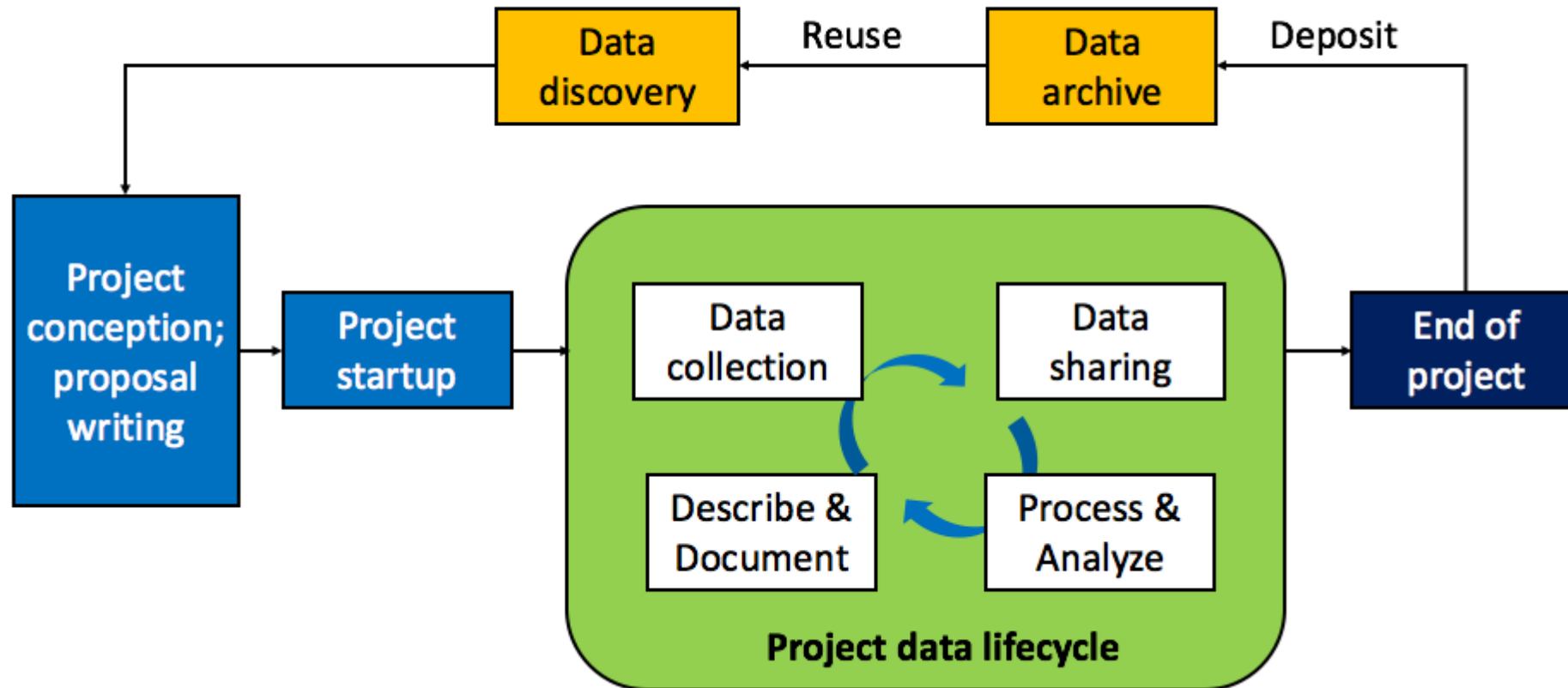
Plusieurs représentations du cycle de vie

The Research Data Management Lifecycle



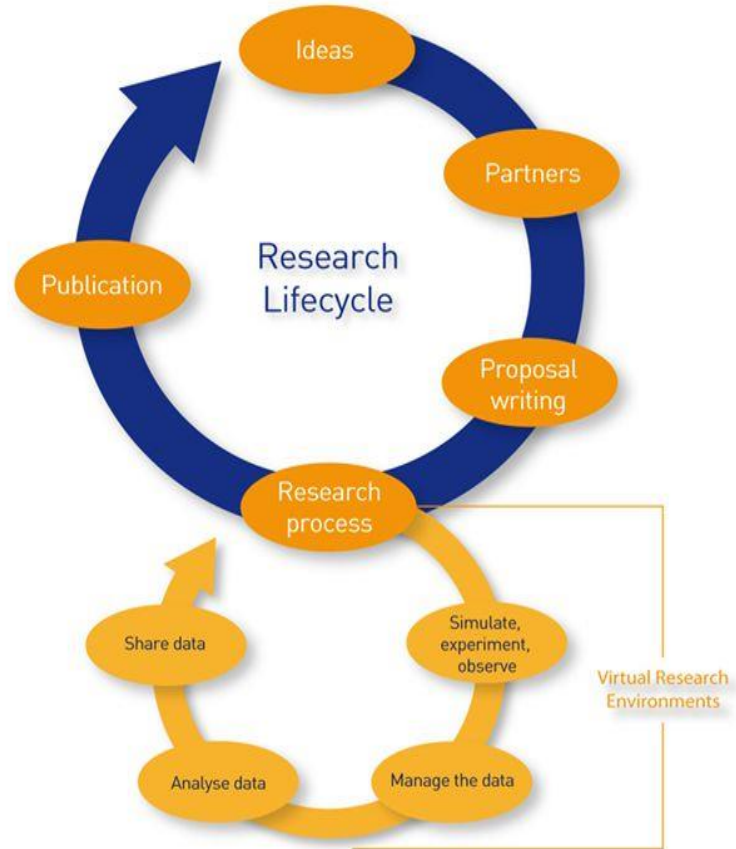
C'EST QUOI LE CYCLE DE VIE DES DONNÉES ? (4/5)

- Articulation dans le cycle de vie d'un projet de recherche



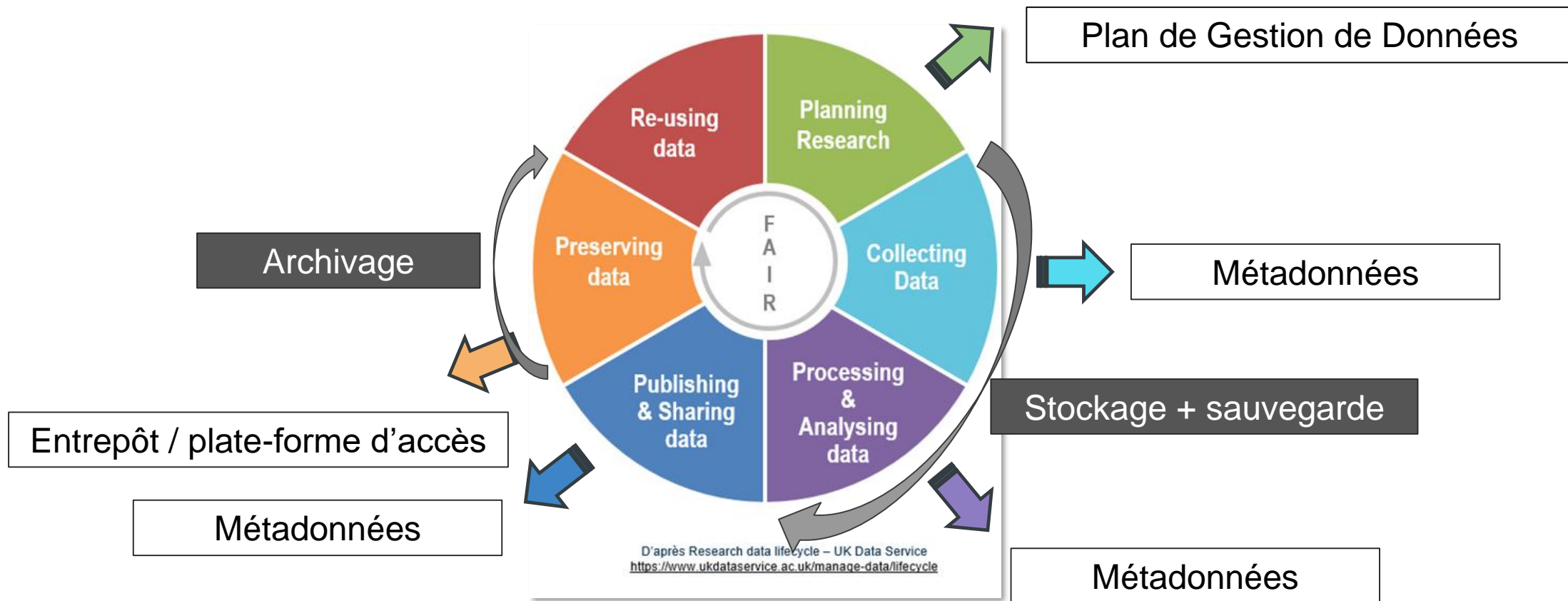
C'EST QUOI LE CYCLE DE VIE DES DONNÉES ? (5/5)

Research
lifecycle -
JISC



<http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/whatwedo/campaigns/res3/jischelp.aspx>

GESTION DES DONNÉES LE LONG DU CYCLE DE VIE DES DONNÉES : LES CONCEPTS-CLÉ



CONCEPTS-CLÉ :

LES DATA MANAGEMENT PLANS (DMP) (1/3)

Description des données et collecte ou réutilisation des données existantes

Documentation et qualité des données

Stockage et sauvegarde pendant le processus de recherche

Exigences légales et éthiques, codes de conduite

Partage des données et conservation à long terme

Responsabilités et ressources en matière de gestion des données

Définition

Un Plan de Gestion de Données (PGD), ou *Data Management Plan* (DMP) en anglais, est un document formalisé - un livrable du projet pour la plupart des appels à projets actuellement - explicitant la manière dont seront obtenues, documentées, analysées, disséminées et utilisées les données produites au cours et à l'issue d'un processus ou d'un projet de recherche. ¹

Le DMP doit commencer à être rédigé dès le début du projet (voire en amont). Il facilitera la bonne gestion des données tout au long du cycle de vie.

Il existe plusieurs modèles dont celui de Science Europe sur lequel se base celui de l'ANR.

¹ Source : [Guide des bonnes pratiques sur la gestion des données de la recherche](#)

CONCEPTS-CLÉ :

LES DATA MANAGEMENT PLANS (DMP) (2/3)

Trafic

Exemple

Projet Trafic
Fondation de France

Modèle de DMP : ANR
6 blocs de questions

3 types de jeux de données

Renseignements sur le projet Produits de recherche Modèle choisi Rédiger Partager Demande d'assistance conseil Télécharger

tout développer | tout réduire

1. Description des données et collecte ou réutilisation de données existantes (2 questions) +
2. Documentation et qualité des données (2 questions) +
3. Stockage et sauvegarde pendant le processus de recherche (2 questions) +
4. Exigences légales et éthiques, codes de conduite (3 questions) +
5. Partage des données et conservation à long terme (4 questions) +
6. Responsabilités et ressources en matière de gestion des données (2 questions) -

donnees_AIS_brutes donnees_AIS_traitees donnees_cetaces

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

B *I* ☰ ☷ 🔗 📄

Les métadonnées de Kalila, et les données associées, sont et restent la propriété pleine et entière:

- des structures fournisseurs de données : OMMAG, Breach Antilles, Guadeloupe Evasion Découverte, Cétacés Caraïbes et Aventures Marines, sanctuaire AGOA, IFAW ;
- des trois porteurs de projet (Laurent Bouveret OMMAG, Jean-Luc Jung (UMR 7205 SYEB), Iwan Le Berre (UMR6554 LETG) ainsi que de Lorraine Coché, Bénédicte Madon et Romain

Enregistrer

Recommandations Commentaires

ANR

- Décrire les rôles et les responsabilités concernant les activités de gestion des données, par exemple : saisie des données, production des métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données. Nommer la(es) personne(s) responsable(s) impliquée(s) dans la mesure du possible.

CONCEPTS-CLÉ :

LES DATA MANAGEMENT PLANS (DMP) (3/3)

- Ne pas voir les DMP comme une contrainte mais comme une aide à l'élaboration de la stratégie de gestion des données.
- [DMP Opidor](#) est un service de l'INIST proposant des modèles de DMP mais aussi une aide à la rédaction.
- Des tutoriels et une aide existent sur [Dorandum](#).
- Conseils pour FAIRiser ses données.

Le DMP est un document qui évoluera durant tout le cycle de vie de la donnée. Il permettra de remplir facilement les métadonnées lors d'un dépôt dans un entrepôt ou dans un catalogue de métadonnées.

CONCEPTS-CLÉ :

LES MÉTADONNÉES (1/5)

Définition

Les **métadonnées** sont les informations sur les données.

Elles se retrouvent tout au long du cycle de vie.

Elles viennent documenter vos données, vos protocoles, le matériel, les termes utilisés et permettent une meilleure compréhension des données.



- La documentation de la donnée permettra une bonne compréhension pour tous les acteurs, une bonne contextualisation. Elle facilitera aussi la réutilisation.
- Les métadonnées sont généralement structurées selon différents **standards**, génériques ou thématiques.

CONCEPTS-CLÉ :

LES MÉTADONNÉES (2/5)

Exemple 1 de standard de métadonnées :
Dublin Core pour des métadonnées généralistes

Élément	Élément (anglais)	Commentaire
1. Titre (métadonnée)	Title	Nom donné à la ressource
2. Créateur (métadonnée)	Creator	Nom de la personne, de l'organisation ou du service responsable de la création du contenu de la ressource
3. Sujet (métadonnée) ou mots clés	Subject	Thème du contenu de la ressource (mots clés, expressions, codes de classification)
4. Description (métadonnée)	Description	Présentation du contenu de la ressource (résumé, table des matières, représentation graphique du contenu, texte libre)
5. Éditeur	Publisher	Nom de la personne, de l'organisation ou du service responsable de la mise à disposition ou de la diffusion de la ressource
6. Contributeur	Contributor	Nom de la personne, de l'organisation ou du service responsable de contributions au contenu de la ressource
7. Date (métadonnée)	Date	Date de création ou de mise à disposition de la ressource
8. Type	Type	Nature ou genre de la ressource (catégories, fonctions, genres généraux, niveaux d'agrégation du contenu)
9. Format	Format	Manifestation physique ou numérique de la ressource
10. Identifiant de la ressource	Identifier	Référence univoque à la ressource dans un contexte donné (URI, ISBN)
11. Source	Source	Référence à une ressource dont la ressource décrite est dérivée (URI)
12. Langue (métadonnée)	Language	Langue du contenu intellectuel de la ressource
13. Relation (métadonnée)	Relation	Référence à une ressource apparentée
14. Couverture (métadonnée)	Coverage	Couverture spatio-temporelle de la ressource (domaine d'application)
15. Gestion de droits (métadonnée)	Rights	Informations sur les droits associés à la ressource (IPR, copyright, etc.)

https://fr.wikipedia.org/wiki/Dublin_Core

CONCEPTS-CLÉ : LES MÉTADONNÉES (3/5)

Exemple 2 de standard de métadonnées :
Ecological Metadata Language ('EML) pour des données d'écologie



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_METACAT_PRODUCTION. doi:10.48502/8bb5-pk85.

Downloads 0 Citations 0 Views 0 Copy Citation Assessment report

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

Name	File type	Size	Download All
Metadata: Kakila database of marine mammal observation data in the AGOA sanctuary - French Antilles.xml	EML v2.2.0	140 KB	Download
BDD_Kakila_v2_20210420_observateur.tsv	More info text/csv	11 KB	Download
BDD_Kakila_v2_20210420_observation.tsv	More info text/csv	463 KB	Download
BDD_Kakila_v2_20210420_sortie.tsv	More info 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

<https://data.pndb.fr/view/doi:10.48502/8bb5-pk85>

CONCEPTS-CLÉ :

LES MÉTADONNÉES (4/5)

Exemple 3 de standard de métadonnées :
norme ISO 19 139 (INSPIRE) pour les données géographiques

Catalogue de métadonnées InDoRES Rechercher Visualiser S'identifier Français

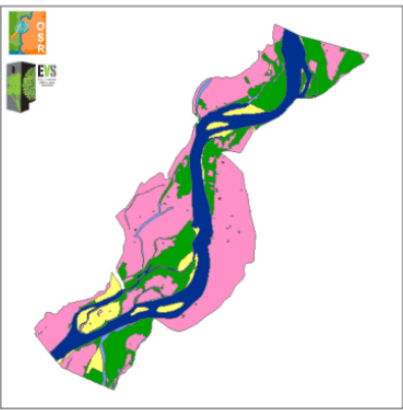
Retour à la recherche Précédent Suivant

Cartographie de l'occupation du sol des Vieux Rhône (1860, 1950, 2006)

Couches vectorielles de polygones représentant les types d'occupation du sol des Vieux Rhône et de leurs marges fluviales entre Chautagne et Beaucaire : agriculture, forêt alluviale, banc de galets, surface en eau, etc. en 1860, 1950 et 2006.

Télécharger Mode affichage

Aperçu



large_thumbnail

Étendue spatiale

- Vieux Rhône de Chautagne à Beaucaire.

Étendue temporelle

Site web de l'Observatoire des Sédiments du Rhône
Site de l'OSR <http://www.graie.org/> Ouvrir le lien

Atlas de synthèse du schéma directeur
Schéma directeur de ré-activation de la dynamique fluviale des marges du Rhône
http://www.graie.org/osr/IMG/pdf/Atlas_synthese.pdf Ouvrir le lien

Rapport de synthèse du schéma directeur
Schéma directeur de ré-activation de la dynamique fluviale des marges du Rhône
http://www.graie.org/osr/IMG/pdf/Rapport_de_synthese.pdf Ouvrir le lien

<http://www.indores.fr/geonetwork/srv/fre/catalog.search#/metadata/a/578a1f9b-7c0b-4ca4-acdf-80e386d73b0b>

CONCEPTS-CLÉ :

LES MÉTADONNÉES (5/5)



Plan de gestion de données évolutif

Dictionnaires de données, thésaurus, ontologies, protocoles de collecte (standards de données)

Protocoles d'analyse/de traitement

Fiches de métadonnées des jeux de données diffusés (standards de métadonnées)

CONCEPTS-CLÉ :

LA SAUVEGARDE, LE STOCKAGE ET L'ARCHIVAGE (1/2)

Définitions

- **Archivage** : notion de pérennité et de conservation sur le long terme.

Archivage

Attention, l'archivage appelle à utiliser des formats précis et a un coût. Le CINES fournit cette expertise

Source : [Guide des bonnes pratiques sur la gestion des données de la recherche](https://www.ukdataservice.ac.uk/manage-data/lifecycle)



- **Stockage** : espace où sont entreposées les données, où elles sont dupliquées mais aussi méthodes et technologies employées.

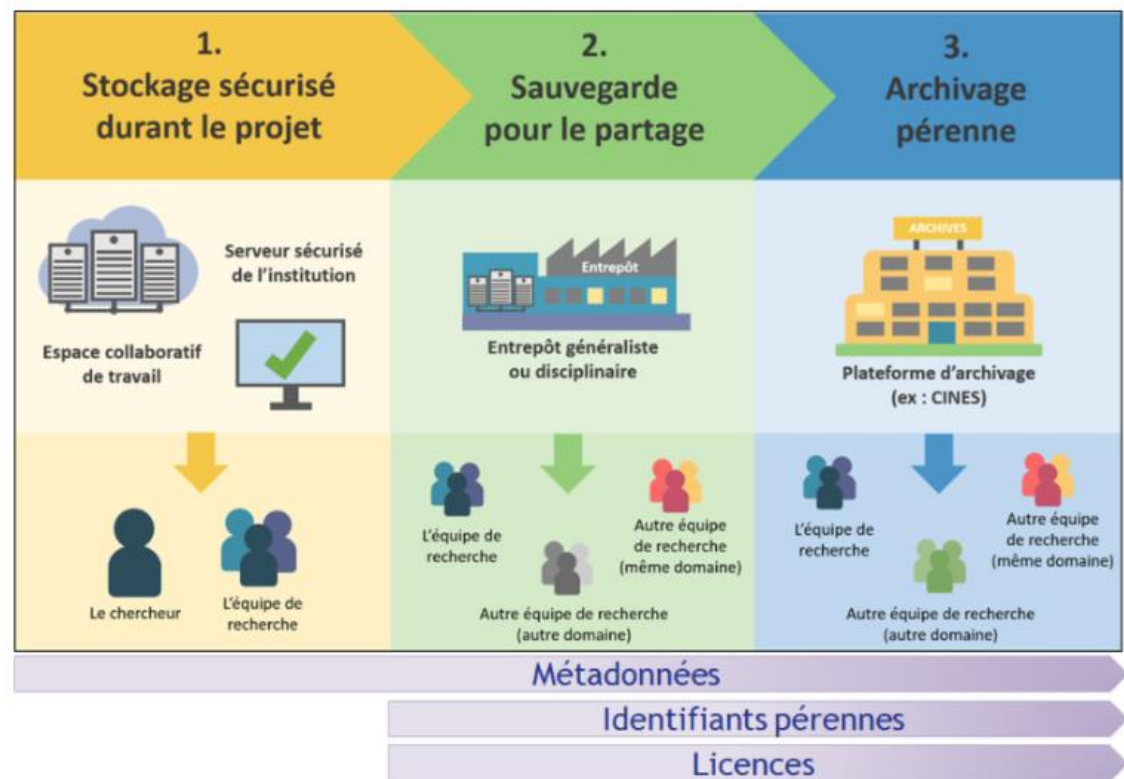
Stockage + sauvegarde

- **Sauvegarde** : dépôt des données sur un support numérique. Elle peut être faite sur un ordinateur personnel, un disque dur externe ou sur un serveur.

CONCEPTS-CLÉ :

LA SAUVEGARDE, LE STOCKAGE ET L'ARCHIVAGE (2/2)

Stockage et sauvegarde pendant le processus de recherche



Source : Doranum

CONCEPTS-CLÉ :

ENTREPÔTS, PLATE-FORMES, INFRASTRUCTURES DE DONNÉES

Définitions

- **Entrepôt** : permet de stocker, d'accéder et d'utiliser des données. Il peut être thématique (data.indores), institutionnel ou générique (Zenodo). Assure la sécurisation des données. L'entrepôt assure le stockage des données sur le moyen terme. [L'entrepôt en résumé](#)
- **Plate-forme d'accès ou infrastructure de données** : permet la gestion et l'accès aux données souvent couplée à des services associés (visualisation des données...).



Le **choix d'un entrepôt** se fera selon la thématique de recherche, les habitudes de la communauté. Il permet de déposer ses données de manière sécurisée, d'avoir un DOI associé. Il existe des annuaires d'entrepôts : [Re3data](#) ou encore [Cat Opidor](#).

La **plate-forme d'accès** ou l'**infrastructure de données** est complémentaire à l'entrepôt ou peut s'y substituer. Elle réunit des compétences scientifiques et techniques et offre des informations plus détaillées sur les données. Elle peut être totalement indépendante du stockage des données ou au contraire proposer ce service.

ANTICIPER LA DIFFUSION DES DONNÉES : QUELLES DONNÉES DIFFUSER ? (1/3)



Plan de gestion de données
évolutif



Première question à se poser :

**Quelles données
diffuser ?**

ANTICIPER LA DIFFUSION DES DONNÉES :

QUELLES DONNÉES DIFFUSER ? (2/3)

Anticiper autant que possible la diffusion de tout ou partie des données dans le cadre d'un projet de recherche (ou d'une structure) :

■ **Savoir identifier lesquelles**

- Données brutes « curées » ? Données issues de traitements/analyses ?
- Données d'intérêt au partage, au moins entre partenaires du projet > communauté > monde
- Données sans restrictions contraignantes ?
- Données ayant un accord consensuel de diffusion !

ANTICIPER LA DIFFUSION DES DONNÉES :

QUELLES DONNÉES DIFFUSER ? (3/3)

Anticiper autant que possible la diffusion de tout ou partie des données dans le cadre d'un projet de recherche (ou d'une structure) :

- **Savoir identifier lesquelles**
- **Les identifier le plus tôt possible pour pouvoir préparer/formater les données** dès leur création, comme l'infrastructure de données ou l'entrepôt éventuellement le requiert. Exemples :
 - utiliser le modèle pivot de Theia-OZCAR (v1.1) en vue d'une diffusion vers l'IR OZCAR
 - utiliser le standard Darwin Core pour les données d'occurrences de biodiversité pour une diffusion vers PNDB et/ou GBIF
- **Voir comment les rendre les plus réutilisables/reproductibles possibles**
 - cf. ci-dessus
 - réaliser un dictionnaire de données et/ou utiliser des vocabulaires contrôlés, thésaurus et ontologies autant que possible (Cf. Session Interopérabilité et Réutilisation)

ANTICIPER LA DIFFUSION DES DONNÉES :

LES DONNÉES QUE J'EXPLOITE SONT-ELLES CONCERNÉES PAR L'OPEN DATA ?

Ouvrir les données autant que possible, les fermer autant que nécessaire (EU, Prog. H2020)

Communication interdite par principe :

- Données présentant des risques pour la protection du secret de la défense nationale
- Données présentant des risques pour la sûreté de l'État, la sécurité publique, la sécurité de l'établissement
- Secrets professionnels

Communication obligatoire :

- Données géographiques (directive européenne INSPIRE),
- Données relatives à des émissions de substances dans l'environnement

Communication sous conditions

- Données présentant des risques pour la protection du potentiel scientifique et technique de la nation
- Le cas des zones à régime restrictif (ZRR)
- Données protégées par le droit d'auteur et autres droits de propriété intellectuelle
- Données relatives aux personnes, à la vie privée
- Données statistiques
- Données liées à un contrat avec un tiers non soumis à une obligation de service public

STRATÉGIE DE LA DIFFUSION DES DONNÉES :

COMMENT RENDRE AISÉMENT TROUVABLE MON JEU DE DONNÉES ? (1/3)

- **Quand on diffuse un jeu de données, on souhaite qu'il soit :**
 - [F] facilement trouvable et donc bien visible
 - [A] accessible en ligne à moyen terme (avec ou sans restriction)
 - [I] si possible aisément interopérable, donc comparable à d'autres données de même type dont les croisements/analyses seraient réalisés par soi ou d'autres personnes ou des machines
 - [R] si possible aisément réutilisable par soi-même et par d'autres

STRATÉGIE DE LA DIFFUSION DES DONNÉES :

COMMENT RENDRE AISÉMENT TROUVABLE MON JEU DE DONNÉES ? (2/3)

- **Quand on diffuse un jeu de données, on souhaite qu'il soit :**
- **Trouver un jeu de données facilement :**
 - **grâce à ses métadonnées descriptives pertinentes**, en plusieurs langues et/ou à l'aide d'ontologie/thésaurus donc standardisés (cas des mots-clés par exemple)
==> Soigner la description intrinsèque du jeu de données
 - **grâce à son exposition dans une ou plusieurs infrastructures :**
 - les plus généralistes (forte popularité) et/ou
 - connues/reconnues de la communauté disciplinaire et/ou
 - à l'échelle internationale ou au moins nationale
==> Augmenter la visibilité permet d'augmenter la « trouvabilité » du jeu de données

STRATÉGIE DE LA DIFFUSION DES DONNÉES : COMMENT RENDRE AISÉMENT TROUVABLE MON JEU DE DONNÉES ? (3/3)

Modes de diffusion optimisant la trouvabilité, la réutilisation, etc., des données :

- data papers,
- dépôts dans des entrepôts ouverts,
- données accessibles dynamiquement via les revues

	Availability	Findability	Interpretability	Re-usability
Data presented within articles	+	+/-	++	-
Data in Journal Supplements, unrestricted	++	++	+++	+/-
Data in Journal Supplements, but restricted	+	++	+++	-
Data in public archives, linked to publications	+++	++++	+++	+++
Journals storing data	+	++	+++	++
Journals making data interactive	++	+++	++++	++++
Data Publications	++++	++++	++++	++++

Table 6: Rating the different ways to publish research data against the Treloar criteria

	Citability	Curation	Preservation
Data presented within articles	+++	+++	+++
Data in Journal Supplements, unrestricted	++++	+/-	+/-
Data in Journal Supplements, but restricted	++	++	+
Data in public archives, linked to publications	++++	++++	++++
Journals storing data	+++	+++	+++
Journals making data interactive	++++	+++	++++
Data Publications	++++	++++	++++

Table 7: Rating the different ways to publish research data against the criteria of citability, curation and preservation.

Définition

- **Findability**, ou **trouvabilité** ou **repérabilité** en français : possibilité et facilité par laquelle les internautes trouvent une information sur Internet et à l'intérieur d'un site ou d'une page Web

Reilly, S., Schallier, W., Schrimpf, S., Smit, E., & Wilkinson, M. (2011). Report on integration of data and publications. Zenodo.
<https://doi.org/10.5281/zenodo.8307>

STRATÉGIE DE LA DIFFUSION DES DONNÉES :

COMMENT RENDRE INTEROPÉRABLE MON JEU DE DONNÉES ?

- Interopérabilité = accessibilité et réutilisation des données facilitées
- Choisir un entrepôt en adéquation avec les recommandations de son établissement ou des tutelles en s'assurant :
 - de l'interopérabilité du système avec les infrastructures de plus haut niveau (futur recherche.data.gouv, EOSC...)
 - des conditions d'accès et d'utilisation satisfaisantes (licence CC-BY...)
- Utiliser des thésaurus, référentiels ou vocabulaires contrôlés pour :
 - les données (ex : Darwin Core pour les données d'occurrences de biodiversité, référentiel taxonomique TAXREF, modèle pivot de Theia-OZCAR)
 - les métadonnées (mots-clé : thésaurus tels que GEMET, OZCAR-Theia thesaurus)

Cf. session Interopérabilité et Réutilisation



TD - APPLICATION

APPRENDRE À PLANIFIER LA GESTION DE DONNÉES : LES BONNES QUESTIONS À SE POSER EN VUE DE LA FAIRISATION DES DONNÉES

TD/ APPRENDRE À PLANIFIER LA GESTION DE DONNÉES : LES BONNES QUESTIONS À SE POSER EN VUE DE LA FAIRISATION DES DONNÉES

Principe du TD

- Etape 1. Explication de la méthode
- Etape 2. Illustration à l'aide d'un jeu de données exemple
- Etape 3. TD : chaque stagiaire répond aux questions sur le cycle de vie des données en lien avec son jeu de données, avec l'aide des co-animateurs (groupes de travail possibles)
 - outil de TD : <https://docs.google.com/presentation/d/17uTvFzgf8j-GCfZ-aJ7XhaJES8r4OkAr/edit?usp=sharing&oid=107545774364631245781&rtpof=true&sd=true>
 - Chaque stagiaire copie-colle les diapos 4 à 8 de ce support et répond aux questions
- Etape 4. Débriefing final

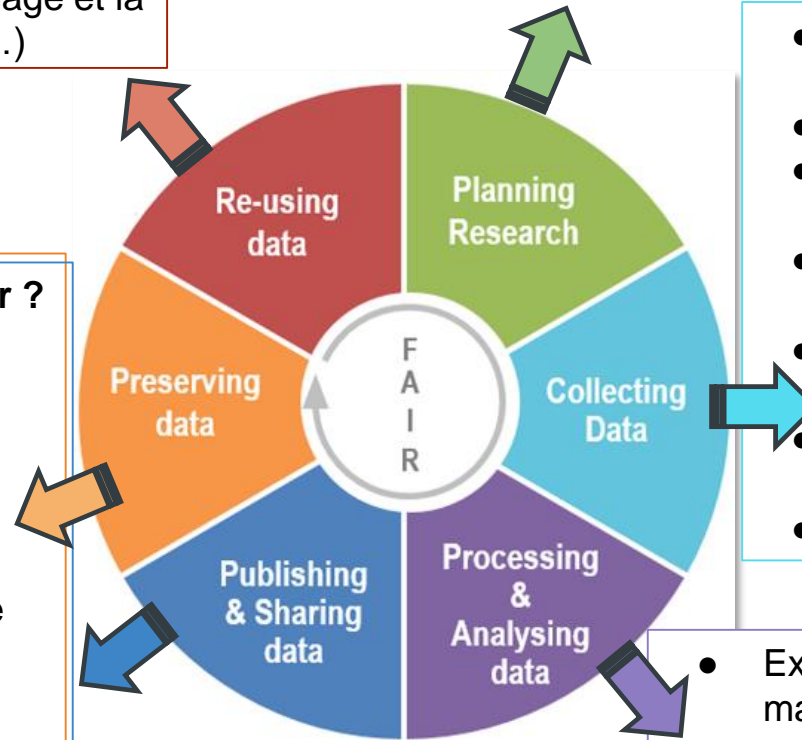
- **Comment optimiser la (ré)utilisation des données diffusées ?** (réfléchir dès les phases amont sur le choix des variables, la standardisation des contenus, le langage et la documentation des scripts associés...)

- **Qui fait quoi où quand comment pourquoi combien ? rôles**
- **Quel cadre légal et éthique** pour les données ?
À qui appartiennent les données, notamment celles à FAIRiser ?

- **Quelles données diffuser/préserver ?**
- Quel **entrepôt** ou plate-forme ou infrastructure de données choisir ?
- **À qui s'adressent les données ?**
(Selon le public ciblé, la diffusion sera différente : science participative, collectivité ou communauté scientifique exclusivement)
- À partir de quand les données seront accessibles (« **embargo** ») ?
- Quelle **licence** choisir ?
- À quelle **fréquence** seront mises à jour les données ?

- **Comment** seront collectées/acquises et gérées les données ?
- **Où** seront-elles stockées et sauvegardées ?
- Avec quelles **mesures de sécurité/protection** s'il s'agit de données personnelles ou sensibles ?
- Faut-il mettre en place une **base de données** ?
L'interface est-elle déjà existante ?
- Quelle est la **volumétrie** approximative des données collectées et/ou acquises ?
- De quels **types** et de quels **formats** les données seront collectées et/ou acquises ?
- Existe-t-il un **protocole de collecte** ? diffusable ?

- Existe-t-il un **standard de données** adapté aux données manipulées et reconnu par la communauté disciplinaire ? Si oui, faut-il l'utiliser et pour quels jeux de données ?
- Faut-il créer un **dictionnaire des données** pour tout ou partie des données acquises/collectées/produites ?
- Existe-t-il un **protocole de traitement** ? diffusable ?
- Faut-il **documenter les scripts** de traitement ?
- Quelles sont les mesures de **contrôle qualité** ?



D'après Research data lifecycle – UK Data Service
<https://www.ukdataservice.ac.uk/manage-data/lifecycle>

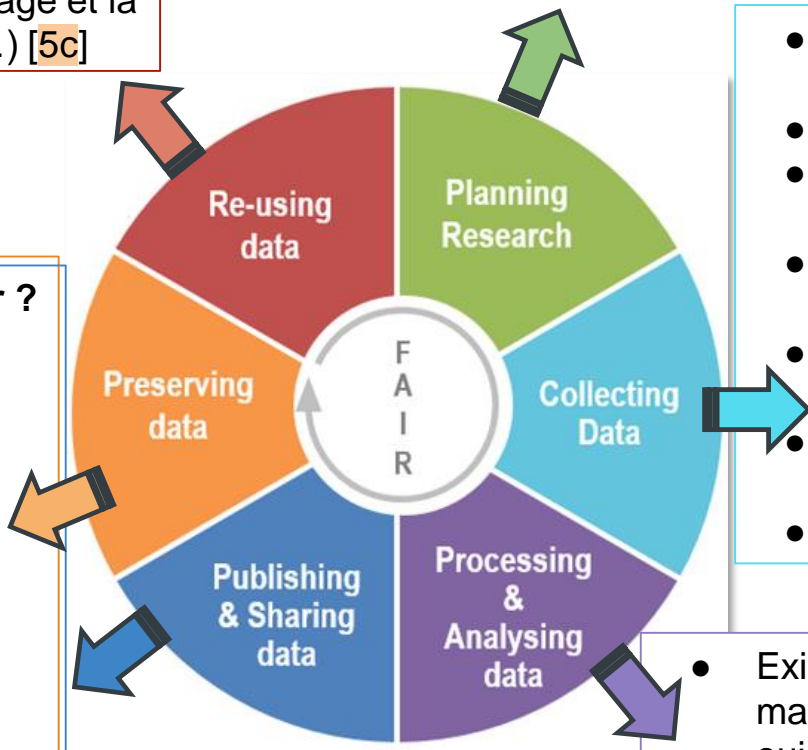
[code question
modèle ANR DMP]

- **Comment optimiser la (ré)utilisation des données diffusées ?** (réfléchir dès les phases amont sur le choix des variables, la standardisation des contenus, le langage et la documentation des scripts associés...) [5c]

- **Qui fait quoi où quand comment pourquoi combien ? rôles** [6a,b]
- **Quel cadre légal et éthique** pour les données ?
À qui appartiennent les données, notamment celles à FAIRiser ? [4a,b,c]

- **Quelles données diffuser/préserver ?**
- Quel **entrepôt** ou plate-forme ou infrastructure de données choisir ? [5a,b,d]
- **À qui s'adressent les données ?** (Selon le public ciblé, la diffusion sera différente : science participative, collectivité ou communauté scientifique exclusivement) [5a]
- À partir de quand les données seront accessibles (« **embargo** ») ? [5a]
- Quelle **licence** choisir ? [5a]
- À quelle **fréquence** seront mises à jour les données ? [5a]

- **Comment** seront collectées/acquises et gérées les données ? [1a+3a]
- **Où** seront-elles stockées et sauvegardées ?
- Avec quelles **mesures de sécurité/protection** s'il s'agit de données personnelles ou sensibles ? [3b]
- Faut-il mettre en place une **base de données** ? L'interface est-elle déjà existante ?
- Quelle est la **volumétrie** approximative des données collectées et/ou acquises ? [1b]
- De quels **types** et de quels **formats** les données seront collectées et/ou acquises ? [1b]
- Existe-t-il un **protocole de collecte** ? diffusable ? [2a]



D'après Research data lifecycle – UK Data Service
<https://www.ukdataservice.ac.uk/manage-data/lifecycle>

On retrouve l'ensemble des questions présentes dans le modèle de DMP de l'ANR (v4)

- Existe-t-il un **standard de données** adapté aux données manipulées et reconnu par la communauté disciplinaire ? Si oui, faut-il l'utiliser et pour quels jeux de données ?
- Faut-il créer un **dictionnaire des données** pour tout ou partie des données acquises/collectées/produites ? [2a]
- Existe-t-il un **protocole de traitement** ? diffusable ? [2a+5c]
- Faut-il **documenter les scripts** de traitement ? [2a+5c]
- Quelles sont les mesures de **contrôle qualité** ? [2b]

TD/ APPRENDRE À REPRÉSENTER SUR LE CYCLE DE VIE DES DONNÉES

Jeu de données exemple : Base de données KAKILA (Guadeloupe, mammifères marins)"

Présentation du jeu de données exemple

- Jeu de données = **Base de données Kakila** d'occurrences de mammifères marins au large de la Guadeloupe (**données d'observations** issues de **science participative**)
- Contexte de recherche :
 - suivi de l'écosystème marin environnant, impacts et changements
 - rechercher des proxys permettant d'évaluer les impacts des activités humaines sur l'écosystème marin ==> **les mammifères marins** = taxons emblématiques et intégratifs sur le plan écologique
 - approches moléculaires et de science participative
 - couplage avec données de fréquentation portuaire et de transports maritimes (données AIS)
 - plusieurs projets en lien (« GenMamGua » OHM/DRIIHM - « TRAFIC » Fondation de France)
- Sources :
 - Jean-Luc Jung, Iwan Le Berre, Lorraine Coché, Bénédicte Madon, Eric Foulquier, et al.. Baleines et dauphins : des belles espèces sentinelles à étudier dans le cadre d'un OHM littoral, mais d'un abord bien complexe. *International Symposium of LabEx DRIIHM*, Labex DRIIHM; CNRS InEE, Sep 2021, Toulouse, France. [10.34972/driihm-dab5a4](#). [hal-03353044](#)
 - Madon B., Lerigoleur E., communications personnelles



Source des données		Nombre d'observations	Total
Association	OMMAG	2626	4704 observations, 2000 – 2019 21 espèces
	BREACH	199	
Whale-watchers	Aventures Marines	527	
	Cétacés Caraïbes	310	
	Guadeloupe Evasion Découverte	829	
Campagne Scientifique	IFAW	56	
OFB	AGOA	155	

Sept sources de données différentes

Tableaux, non homogénéisés entre-eux

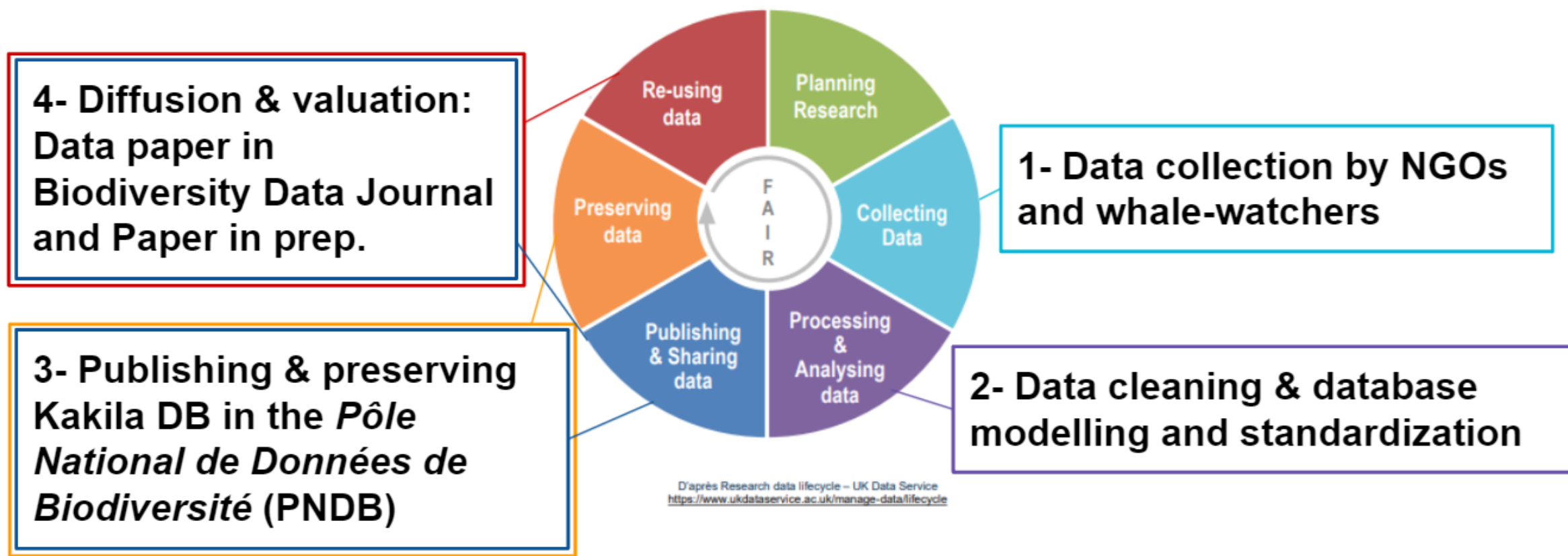
The image shows a complex data table with multiple columns and rows. At the top, there are logos for OMMAG and CÉTACÉS CARAÏBES. The table has several sections of headers, including 'Date', 'Heure', 'GPS', 'Espèce', 'Nombre', 'Direction', 'Vitesse', 'Polarisation', 'Comportement', and 'Observations'. The data rows contain numerical and categorical information, such as coordinates, species names, and counts. The table is partially obscured by a grid overlay.

Extrait des jeux de données transmis

Jean-Luc Jung, Iwan Le Berre, Lorraine Coché, Bénédicte Madon, Eric Foulquier, et al.(2021) [\(10.34972/driihm-dab5a4\)](#), [\(hal-03353044\)](#)

Creation of Kakila

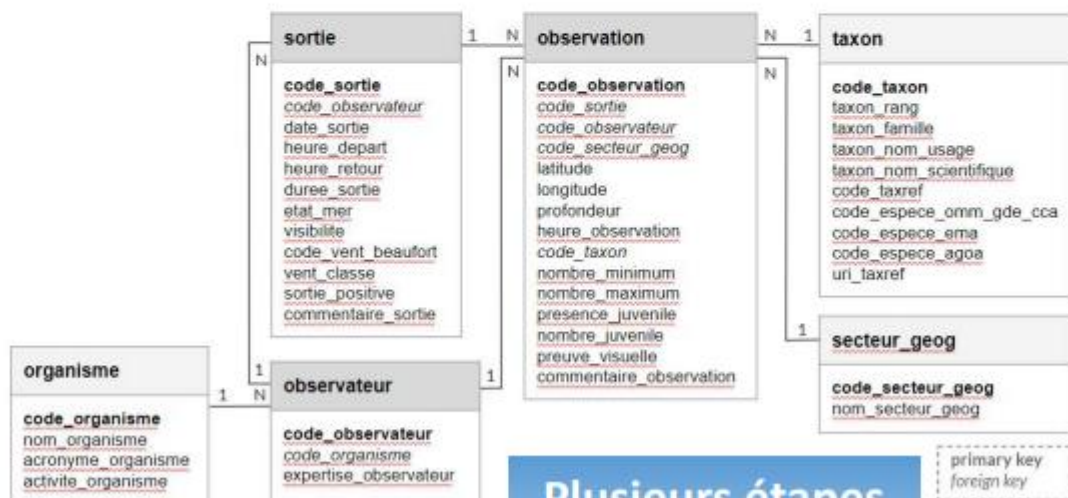
steps and challenges towards a FAIR, open and citizen-science database



Madon B., Lerigoleur E., communications personnelles

Harmonisation et organisation des données

Organisation logique
en six tables de
données



Plusieurs étapes
de curation des
données

Table 3.

Download as [CSV](#) [XLSX](#)

Data dictionary - metadata repository - of the Kakila DB. Datasets and Column labels are also presented in the "Data resources" part. The Darwin core data standards are described in [Wieczorek et al. \(2012\)](#).

Datasets and Column labels	Definition	Data type	Darwin Core term code	Darwin Core term definition
Dataset "sortie" (Trip)				
code_sortie	Code of the boat trip carried out by an organisation and reported by an observer	Text	eventID	An identifier for the set of information associated with an Event (something that occurs at a place and time). May be a global unique identifier or an identifier specific to the data set.
date_sortie	Date of the trip.	Date	eventDate	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded.

Création d'un dictionnaire de données
(incluant lorsque pertinent, des équivalences
Darwin core)

Définition et mise en oeuvre d'une stratégie de bancarisation

- Intégration des données dans le Pôle National de Données de Biodiversité
- Utilisation de Metashark (description de données écologiques), *Ecological Metadata Language*

The screenshot shows the PNDDB Data Catalog interface. At the top, there are navigation tabs: DATA, SUMMARY, ABOUT, Jump to, DOI or E, Go, and SIGN IN. Below the navigation, there is a search bar and a list of datasets. The selected dataset is titled 'Lorraine Coché, Elie Amaud, Bouvenet 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:nid:PNDB_METACAT_PRODUCTION, doi:10.48502/8bb5-pk85'. Below the title, there are statistics for Downloads (0), Citations (0), and Views (0). There are also buttons for 'Copy Citation' and 'Assessment report'. The main content area shows a table of files in the dataset:

Name	File type	Size	Download All
Metadata: Kakila database of marine mammal observation data in the AGOA sanctuary - French Antilles.xml	EML v2.2.0	140 KB	Download
BOD_Kakila_v2_20210420_observateur.tsv	text/csv	11 KB	Download
BOD_Kakila_v2_20210420_observation.tsv	text/csv	463 KB	Download
BOD_Kakila_v2_20210420_sortie.tsv	text/csv	217 KB	Download

At the bottom of the table, there is a link to 'Show 6 more items in this data set'.

The screenshot shows the header of the PNDDB website. It includes the PNDDB logo, navigation links (Outils, Animations, Projets, Gouvernance, FAQ, Contacts), and a language selector (Langue FR) and a button for 'Accéder au portail de données'. Below the header, there is a large banner with the PNDDB logo and the text 'Pôle National de Données de Biodiversité'. Underneath the banner, there is a sub-header 'Un pôle de données au service des scientifiques produisant, gérant et analysant des données de biodiversité' and three buttons: 'Accéder au portail de données', 'Générer de la métadonnée via MetaShARK', and 'Analyser des données via Galaxy-E'. Below this, there is a paragraph of text and a list of four points:

En 2018, le Ministère de l'Enseignement supérieur, de la recherche et de l'innovation a inscrit sur sa feuille de route la création d'une nouvelle infrastructure intitulée Pôle National de données de biodiversité (PNDB). Les missions du PNDB s'inscrivent dans une approche FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), et consistent à :

1. fournir un accès aux jeux de données et de métadonnées, à des services associés et à des produits dérivés des analyses;
2. promouvoir l'animation scientifique pour identifier les lacunes et favoriser l'émergence de dispositifs portés par des communautés d'utilisateurs et producteurs;
3. faciliter le partage des pratiques avec les autres communautés de recherche, favoriser le partage des données et leur réutilisation, s'insérer dans la réflexion de la future infrastructure Système Terre;
4. favoriser la cohérence avec les efforts nationaux, européens et internationaux relatifs à l'accès et à l'exploitation des données de recherche sur la biodiversité, à la promotion de produits et services.

Le PNDB est porté par le Muséum national d'Histoire naturelle, plus particulièrement par l'UMS 2056 Patrinat, unité MNHN CNRS et CFB. Le projet est en lien étroit avec la FRB et plusieurs de ses institutions associées (JOFB, BRGM, CIRAD, CNRS, Brenner, INERIS, INRA, IRD, IRSTEA, MIP-H, Univ. Montpellier).

Portail de données MetaShark Galaxy-E

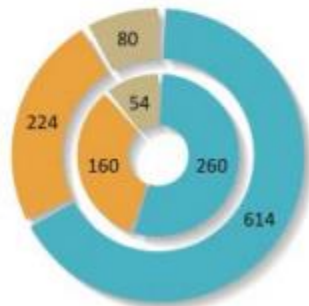
Yvan Le Bras, UMS Patrinat et al.

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_METACAT_PRODUCTION.
 doi:10.48502/8bb5-pk85

Article views

Unique views: 474 | Total views: 918

■ HTML ■ PDF ■ XML



Biodiversity Data Journal 9: e69022
 doi: 10.3897/BDJ.9.e69022

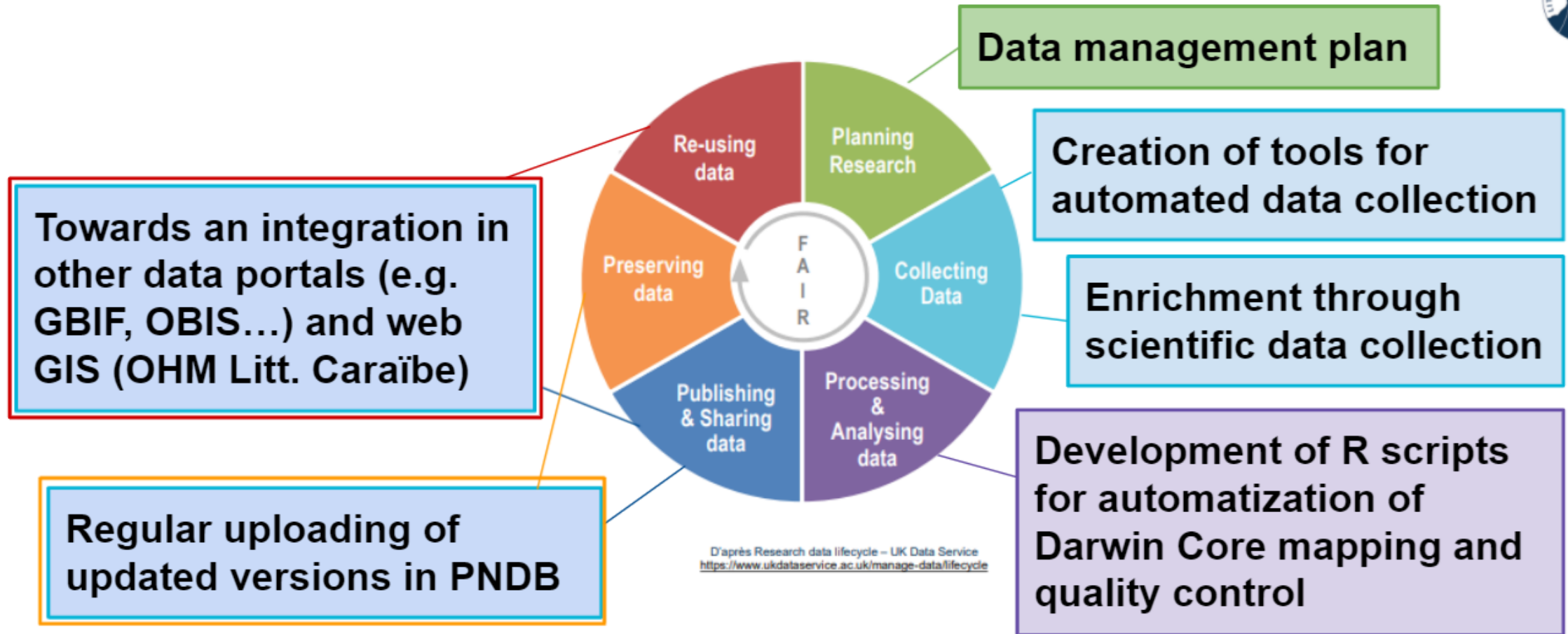


Data Paper

Kakila database: Towards a FAIR community approved database of cetacean presence in the waters of the Guadeloupe Archipelago, based on citizen science

Lorraine Coché[‡], Elie Arnaud[§], Laurent Bouveret[!], Romain David[¶], Eric Foulquier[‡], Nadège Gandilhon[#], Etienne Jeannesson[¶], Yvan Le Bras[§], Emilie Lerigoleur[¶], Pascal Jean Lopez^{¶, *}, Bénédicte Madon[‡], Julien Sananikone[!], Maxime Sèbe^{‡, †}, Iwan Le Berre[‡], Jean-Luc Jung^{¶, †}

Discussion & Perspectives for Kakila



Madon B., Lerigoleur E., communications personnelles

TD/ APPRENDRE À PLANIFIER LA GESTION DE DONNÉES : LES BONNES QUESTIONS À SE POSER EN VUE DE LA FAIRISATION DES DONNÉES

Principe du TD

- ~~Etape 1. Explication de la méthode~~
- ~~Etape 2. Illustration à l'aide d'un jeu de données exemple~~ ⇒ Place au TD !
- Etape 3. TD : chaque stagiaire répond aux questions sur le cycle de vie des données en lien avec son jeu de données, avec l'aide des co-animateurs (groupes de travail possibles)
 - outil de TD : <https://docs.google.com/presentation/d/17uTvFzgf8j-GCfZ-aJ7XhaJES8r4OkAr/edit?usp=sharing&oid=107545774364631245781&rtpof=true&sd=true>
 - Chaque stagiaire copie-colle les diapos 4 à 8 de ce support et répond aux questions
- Etape 4. Débriefing final

BILAN - CONCLUSION

Des concepts-clé liés à la gestion des données :

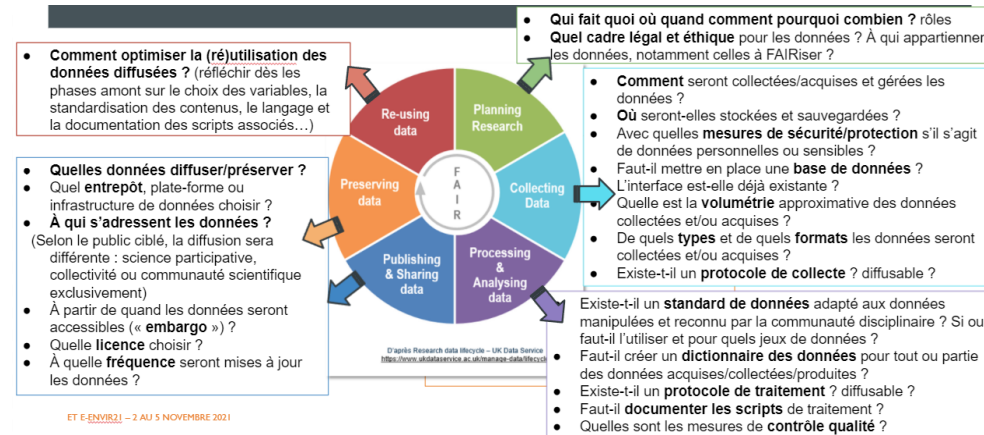
- cycle de vie des données
- plan de gestion de données
- métadonnées
- entrepôt, plate-forme, infrastructure
- stockage/sauvegarde/archivage
- standards (données et métadonnées)

Les différentes stratégies pour valoriser ses données de recherche :

- Publier ses données dans un entrepôt,
- Fournir ses données sous la forme de matériel supplémentaire à la publication,
- Publier ses données dans une publication scientifique spécifique décrivant les données, ou Data Papers (article de données),
- Publier dans le Web des données.

Source : <https://datapartage.inrae.fr/Partager-Publier>

Une méthode pour apprendre à anticiper la gestion des données et à se poser les bonnes questions à l'aide du DMP et du cycle de vie des données



→ Penser à “re-boucler” le cycle de vie des données autant de fois que nécessaire (le DMP reste évolutif) = prendre du recul

→ Apprendre à diagnostiquer une gestion de données existantes à l'aide d'outils d'évaluation

APPRENDRE À FAIRE UN DIAGNOSTIC SUR LA GESTION DES DONNÉES

EXEMPLES D'OUTILS PERMETTANT DE RÉALISER UN DIAGNOSTIC (1/3)

Exemple 1

Comment opérationnaliser et évaluer la prise en compte du concept 'FAIR' dans le partage des données: vers une grille simplifiée d'évaluation du respect des critères FAIR

David, Romain, Mabile, Laurence, Yahia, Mohamed, Cambon-Thomsen, Anne, Archambeau, Anne-Sophie, Bezuidenhout, Louise, Bekaert, Sofie, Bertier, Gabrielle, Bravo, Elena, Carpenter, Jane, Cohen-Nabeiro, Anna, Delavaud, Aurélie, De Rosa, Michele, Dollé, Laurent, Grattarola, Florencia, Murphy, Fiona, Pamerlon, Sophie, Specht, Alison, Tassé, Anne-Marie, ... Zilioli, Martina. (2018). Comment opérationnaliser et évaluer la prise en compte du concept 'FAIR' dans le partage des données: vers une grille simplifiée d'évaluation du respect des critères FAIR. (1.1). Journée Nationale de la Science Ouverte (JNSO 2018), Paris France. Zenodo. <https://doi.org/10.5281/zenodo.2551500>

How to operationalize and to evaluate the FAIRness in the crediting and rewarding processes in data sharing: a first step towards a simplified assessment grid.

1) FINDABLE (8 essential criteria)

Indexed identifier ?
Identification
Are each data/dataset identified by an indexed and independent identifier?
 Never/NA If Mandatory Sometimes Always

Unique, global, persistent ID?
Identification
Are the data identifiers unique, global and persistent?
 Never/NA If Mandatory Sometimes Always

ID scheme?
Identification
Has any identifying schema been used for data (e.g. DOI)?
 Never/NA If Mandatory Sometimes Always

Persistent metadata / data link ?
Metadata traceability
Are the metadata linked to the dataset through a persistent identifier?
 Never/NA If Mandatory Sometimes Always

Metadata & authority linked ?
Metadata traceability
Are the metadata of each dataset linked to a unique authority (responsible for the datasets at a given time)?
 Never/NA If Mandatory Sometimes Always

Datasets linked to authority ?
Metadata traceability
Are all datasets linked to an authority (legal entity) through a unique and persistent identifier over time (e.g. institution, association or established body)?
 Never/NA If Mandatory Sometimes Always

Standards/dictionary for data description?
Metadata description and searchability
If relevant, has the researcher used valid and updated standards for data describing? If so, are the data standards and particularly versioning data standards recommended by community-approved or appropriate authorities specified? If no standards exist, has the researcher created a well described data dictionary?
 Never/NA If Mandatory Sometimes Always

Data format/type description?
Metadata description and searchability
Are the types and formats of data generated / collected well described?
 Never/NA If Mandatory Sometimes Always

Result for Findable: .../8 Never/NA .../8 If Mandatory .../8 Sometimes .../8 Always

2) ACCESSIBLE (3 essential criteria)

Data repositories?
Repository
Does the researcher use data repositories for the storage of data?
 Never/NA If Mandatory Sometimes Always

Efficient and rich services for various uses & users?
Data security and services
Does the researcher use efficient and rich services to access data (online browser, downloading, visualization and system support) (different types of user support)?
 Never/NA If Mandatory Sometimes Always

Data access restriction justification?
Access restriction
In case of a non open access, is the restriction properly justified by the researcher?
 Never/NA If Mandatory Sometimes Always

Result for Accessible: .../3 Never/NA .../3 If Mandatory .../3 Sometimes .../3 Always

3) INTEROPERABLE (2 essential criteria)

Standard vocabularies, thesaurus, ontologies or data dictionary?
Identification
Are standard vocabularies, thesaurus or ontologies used at data level (e.g. in a long term archive)?
 Never/NA If Mandatory Sometimes Always

Interoperability criteria explained?
Identification
Are the interoperability criteria explained?
 Never/NA If Mandatory Sometimes Always

Result for Interoperability: .../2 Never/NA .../2 If Mandatory .../2 Sometimes .../2 Always

4) REUSABLE (5 essential criteria)

Relevant actions for data reuse potential?
Data potential
Have relevant actions been undertaken by the researcher to enhance the data reuse potential?
 Never/NA If Mandatory Sometimes Always

Provenance for raw and transformed data?
Data traceability
Are the provenance of raw and transformed data well described?
 Never/NA If Mandatory Sometimes Always

Result for Reusable: .../5 Never/NA .../5 If Mandatory .../5 Sometimes .../5 Always

1) FINDABLE (8 essential criteria)

Indexed identifier ?
Identification
 Never/NA If Mandatory Sometimes Always

Are each data/dataset identified by an indexed and independent identifier ?

Unique, global, persistent ID?
Identification
 Never/NA If Mandatory Sometimes Always

Are the data identifiers unique, global and persistent ?

ID scheme?
Identification
 Never/NA If Mandatory Sometimes Always

Has any identifying schema been used for data (e.g. DOI)?

Persistent metadata / data link ?
Metadata traceability
 Never/NA If Mandatory Sometimes Always

Are the metadata linked to the dataset through a persistent identifier?

Metadata & authority linked ?
Metadata traceability
 Never/NA If Mandatory Sometimes Always

Are the metadata of each dataset linked to a unique authority (responsible for the datasets at a given time)?

Datasets linked to authority ?
Metadata traceability
 Never/NA If Mandatory Sometimes Always

Are all datasets linked to an authority (legal entity) through a unique and persistent identifier over time (e.g. institution, association or established body)?

Standards/dictionary for data description?
Metadata description and searchability
 Never/NA If Mandatory Sometimes Always

If relevant, has the researcher used valid and updated standards for data describing? If so, are the data standards and particularly versioning data standards recommended by community-approved or appropriate authorities specified? If no standards exist, has the researcher created a well described data dictionary?

Data format/type description?
Metadata description and searchability
 Never/NA If Mandatory Sometimes Always

Are the types and formats of data generated / collected well described?

Result for Findable: .../8 Never/NA .../8 If Mandatory .../8 Sometimes .../8 Always

APPRENDRE À FAIRE UN DIAGNOSTIC SUR LA GESTION DES DONNÉES

EXEMPLES D'OUTILS PERMETTANT DE RÉALISER UN DIAGNOSTIC (2/3)

2) ACCESSIBLE (3 essential criteria)

Data repositories?

Repository Never/NA If Mandatory Sometimes Always
Does the researcher use data repositories for the storage of data?

Efficient and rich services for various uses & users?

Data security and services Never/NA If Mandatory Sometimes Always
Does the researcher use efficient and rich services to access data (various formats, visualisations, practical tools and systems adapted to different types of use and users)?

Data access restriction justification?

Access restriction Never/NA If Mandatory Sometimes Always
In case of a non legal restricted access, is the restriction properly justified by the researcher?

Result for Accessible: .../3 Never/NA .../3 If Mandatory .../3 Sometimes .../3 Always

3) INTEROPERABLE (2 essential criteria)

Standard vocabularies, thesaurus, ontologies or data dictionary?

Identification Never/NA If Mandatory Sometimes Always
Are standard vocabularies, thesaurus or ontologies used for all data types present in datasets, to enable interdisciplinary interoperability between well defined domains? If not, is a well-defined open data dictionary provided?

Interoperability criteria explained?

Identification Never/NA If Mandatory Sometimes Always
Are the interoperability criteria explained?

Result for Interoperability: .../2 Never/NA .../2 If Mandatory .../2 Sometimes .../2 Always

4) REUSABLE (5 essential criteria)

Relevant actions for data reuse potential?

Data potential Never/NA If Mandatory Sometimes Always
Which relevant actions have been undertaken by the researcher to enhance the data reuse potential?

Provenance for raw and transformed data?

Data traceability Never/NA If Mandatory Sometimes Always
Are the provenance and type of all data properly specified (origin of raw, primary, transformed, secondary...)?

Information on methods and tools that permit the understanding, integrity of data?

Reusability tools Never/NA If Mandatory Sometimes Always
Does the researcher provide information on methods and tools that permit the understandability, integrity, value and readability of data intended to be kept on the long-term? (e.g. versioning, archival and long term reuse issue for protocols, softwares, required methods and contexts to create, read and understand data)

Data sharing arrangements meet data ethics and protection?

Reusability right Never/NA If Mandatory Sometimes Always
Do the data reuse control and data sharing arrangements meet the data protection and "local/national ethics requirements?"

Legal reuse restriction properly justified?

Reusability right Never/NA If Mandatory Sometimes Always
In case of a legal reuse restriction (such as personal data, state and public security, national defense secret, confidentiality of external relations, information systems security, secrets in industrial and commercial matters), is the restriction properly justified?

Result for Reusable: .../5 Never/NA .../5 If Mandatory .../5 Sometimes .../5 Always

TOTAL FAIR simple criteria evaluation results:

.../18 'Never/NA' .../18 'If Mandatory' .../18 'Sometimes' .../18 'Always'

*advices will be provided according to the criteria predominantly obtained

APPRENDRE À FAIRE UN DIAGNOSTIC SUR LA GESTION DES DONNÉES

EXEMPLES D'OUTILS PERMETTANT DE RÉALISER UN DIAGNOSTIC (3/3)

Exemple 2

Diagnostic FAIR d'un observatoire (Theia- OZCAR)

FAIR_test_V5 .XLSX

Fichier Édition Affichage Insertion Format Données Outils Aide

100% | € % .0 .00 123 | Par défaut ... | 11 | B I S A | | | | | | | | | | | | | | | |

A1	A	B	E	F	G	I	J	K	L	M	N	O	P	Q	R	
			Votre observatoire													
2	Service offert par Theia/OZCAR ?	Diagnostic FAIR d'un observatoire	type	pondération	score	réponse	Question complémentaire	Réponse ouverte / commentaire libre		Les réponses sont comptabilisées selon qu'elles se rapportent aux différentes lettres de FAIR, on y a ajouté des questions diverses (D), plutôt d'ordre organisationnel.						
3		Nom de l'observatoire					entrez le nom ici			Score de l'observatoire : remplir case I3 SVP						
4		1- Questions générales sur la gestion des données de votre observatoire									F	A	I	R	D	
5		Avez-vous établi un Plan de Gestion des Données pour votre observatoire ?	D	2	0	??					oui	0	0	0	0	
6		Comment les utilisateurs peuvent-ils avoir connaissance des données produites par votre observatoire (métadonnées) ? (plusieurs réponses possibles)	F	4	0						partiellement	0	0	0	0	
7		- Par le bouche à oreille et les contacts inter-personnels	F	0	0	?					non	0	0	0	0	
8		- Données listées sur une page web	F	1	0	?					sans réponse	0	0	0	0	
9		- Catalogue des données (ex : GeoNetwork)	F	4	0	?					SCORE	0%	0%	0%	0%	
10		- Portail web d'accès aux données (ex : BDOH, HYSAE, etc...)	F	4	0	?				NB : il est nécessaire de répondre aux questions complémentaires pour que les points FAIR et D soient comptabilisés dans ce tableau						
11		- Autre (si autre préciser)	F	1	0	?										
12		Comment sont mises à disposition les données de vos observatoires ? (Plusieurs réponses possibles)	A	4	0											
13		- Contact direct avec le producteur des données	A	0	0	?										
14		- Téléchargement direct de fichiers depuis le web (page web, site ftp, ...)	A	2	0	?										
15		- Données téléchargeables sans métadonnées standardisées (ex : BDOH, HYSAE, etc...)	A	3	0	?										

<https://docs.google.com/spreadsheets/d/1OHtc0XNfVsAXPLIORARRZSfcdxQrKa5h/edit?rtpof=true#gid=1868094237>



JEUX DE DONNÉES EXEMPLES

PRÉSENTATION





Jeu de données météo Bénin

[Lien vers la fiche de métadonnées](#)

[Lien vers les données](#)

Leader du projet : Sylvie Galle

- ▶ Visualisation : **AMMA-CATCH**
- ▶ Jeu de données météo au Bénin.
- ▶ Les données possèdent un identifiant pérenne.
- ▶ Le jeu de données est un relevé de mesures issues de capteur.



Jeu de données Cartographie de l'occupation ancienne des sols du massif des Pyrénées

[Lien vers la fiche de métadonnées](#)

[Lien vers les données](#)

Auteur : Wilfried Heintz

- ▶ Entrepôt : **data.inrae**
- ▶ Cartographie de l'occupation ancienne du sol du massif des Pyrénées réalisée à partir de la digitalisation des minutes de cartes d'Etat Major. (2012-04-01)
- ▶ Le jeu de données possède un identifiant pérenne.
- ▶ Le format des données est un shapefile contenant une couche vecteur polygone.



Jeu de données Relevés végétation

[Lien vers la fiche de métadonnées](#) ou encore [ici](#)

[Lien vers les données](#)

Auteur : Audrey Alignier

- ▶ Entrepôt : **B2SHARE**
- ▶ Jeu de données d'occurrences de plantes vasculaires au sein de la Zone Atelier Armorique (entre 1994 et 2017).
- ▶ Le jeu de données possède un identifiant pérenne.
- ▶ Le format des données est un tableur Excel (format xlsx) où l'on retrouve des colonnes telles que "Lieu", "Année", "Espèce" ou encore "Abondance".
- ▶ Le protocole utilisé est également documenté et des liens vers de la documentation sont listés.

REMERCIEMENT

cnrs

- CNRS soutien financier
- R. OMNES (conseillère formation, DR10) suivi de la formation
- P. RUPP (ICUBE), C. DIDIERJEAN (EDYTEM) gestion administrative et financière
- Le comité organisateur : BED/RZA, OZCAR/THEIA, UMS BBEES, PNDB, avec la participation de la MISHA
- Centre Saint-Thomas, hébergement

- Retrouvez les informations sur : <https://e-envir-21.sciencesconf.org/>



Maison Interuniversitaire des
Sciences de l'Homme
- Alsace -



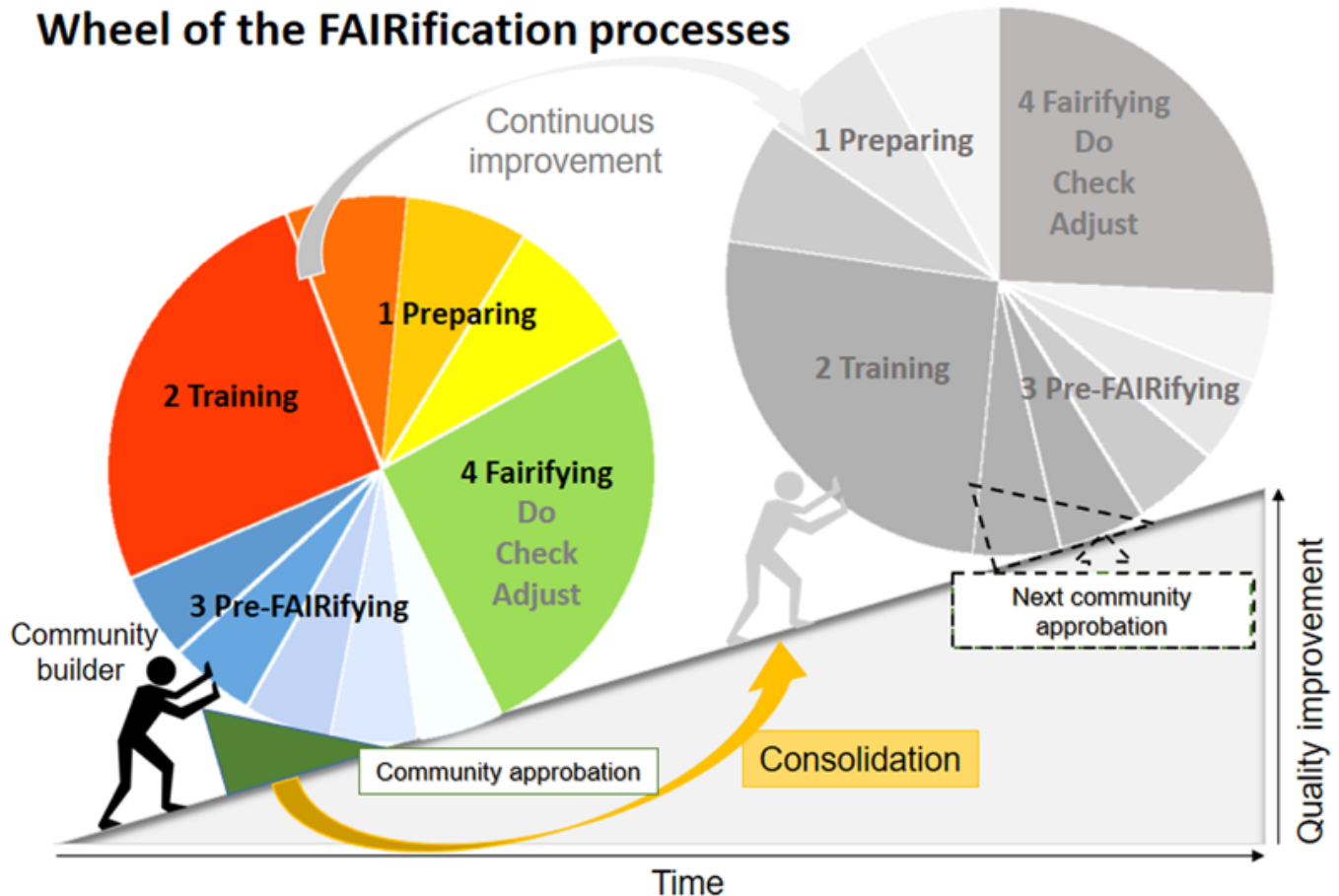
ANNEXES

FAIRISATION OU FAIRIFICATION, QUELLE DIFFÉRENCE ? C'EST QUOI LA PRÉ-FAIRISATION ? (1/2)

- La **FAIRisation** (ou FAIRization) signifie l'ensemble du processus qui amène à produire des données FAIR.
- Quant à la **FAIRification**, elle correspond à la réalisation technique de la FAIRisation.
- **En amont** de cette phase de FAIRification, il y a une **phase de pré-FAIRisation** qui permet d'énoncer les objectifs de la FAIRification et d'identifier les données et les métadonnées à rendre FAIR.
- Cette distinction a été formulée par le groupe RDA-SHARC-IG (Research Data Alliance - SHARing Rewards and Credit -Interest Group). Elle permet de mieux appréhender les différentes étapes d'un projet de FAIRisation, bien qu'elle ne soit pour l'instant pas adoptée par la communauté scientifique.
- Les termes de FAIRisation et FAIRification sont souvent utilisés de façon interchangeable et dépendent de l'interlocuteur. On peut néanmoins remarquer que GO FAIR, l'une des structures de référence, n'emploie que le terme FAIRification.
- Sources :
 - Evaluation tool of FAIR criteria literacy and compliance to foster research data sharing. 2. Open Science FAIR Conference: “ Synergies for Sustainable, Open and Responsible Research ”, Sep 2019, Porto. (hal-02164148)
 - http://phn-wiki.ish-lyon.cnrs.fr/doku.php?id=fairdata:fairisation_fairification

FAIRISATION OU FAIRIFICATION, QUELLE DIFFÉRENCE ? C'EST QUOI LA PRÉ-FAIRISATION ?

Process	Steps
Preparing FAIRification	Explain FAIRification
	Define constraints
	Define advantages
Training	Increase FAIR literacy Convince partners
Pre-FAIRifying	Building shared strategy
	Define community
	Define objects and variables
	Select items to be identified
	Analyse common denominators
FAIRifying	Do: Downward levelling
	Check: first interoperations
	Adjust: Identifying gaps and new expectation



Romain David, Laurence Mabile, Alison Specht, Sarah Stryeck, Mogens Thomsen, et al.. FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. *CODATA Data*

ENJEUX DE LA DIFFUSION DES DONNÉES :

PYRAMIDE DE PUBLICATION DES DONNÉES

