Ce webinaire a pour objectif de vous expliquer à quoi correspondent les principes FAIR. Ces derniers s'appliquent tout au long du cycle de vie des données de recherche. Celui-ci sera détaillé en expliquant à chaque étape les bonnes pratiques à mettre en place pour respecter les principes FAIR.

Sommaire

cnrs Inist

01

Définition des données de recherche

Définition des données de recherche

« 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 les résultats** de recherche. » (Définition l'OCDE)

Si vous voulez en savoir plus : What is data?

cnrs Inist

4

Plusieurs définitions des données de la recherche existent.

« 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 les résultats de recherche**. » (*OCDE, Organisation de Coopération et de Développement Économiques. Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. 2007. https://doi.org/10.1787/9789264034020-en-fr*)

On peut également citer la définition issue de l'ARDC - Australian Research Data Commons (traduction Inist-CNRS) :

« Le terme de données de la recherche désigne les données sous forme de faits, d'observations, d'images, de résultats de programmes informatiques, d'enregistrements, de mesures ou d'expériences sur lesquels un argument, une théorie, un test ou une hypothèse, ou un autre produit de la recherche est basé [...]. Les données peuvent être numériques, descriptives, visuelles ou tactiles. Elles peuvent être brutes, nettoyées ou traitées, et peuvent être conservées dans tout format ou support [...]. » *(ARDC, Australian Research Data Commons. What Is Research Data. 2019)*

Selon leur contexte de création (capture ou production), leur exploitation, leur analyse et les traitements qu'elles subissent, les données de recherche peuvent être :

- de différentes **natures :** données brutes, dérivées, formatées, nettoyées, primaires, secondaires, traitées....
- de tous **types** : audio, vidéos, images, bases de données, codes sources, données géospatiales, photographies, langages de programmation, matérielles et physiques, modèles, visualisations, 3D, numériques, textuelles, numérisations, scans, qualitatives, quantitatives, statistiques...
- contenues dans divers **supports** : documents électroniques, programmes informatiques, carnets de laboratoire, supports papier, logiciels...

Sources *:*

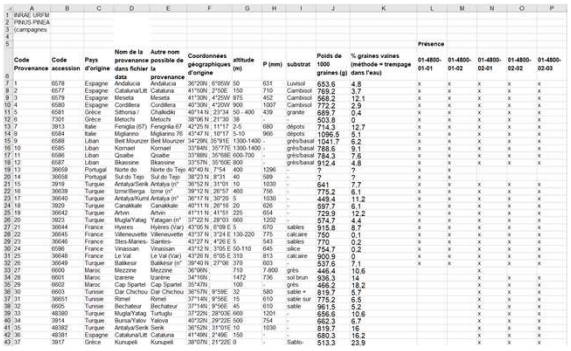
- *Rivet Alain, Bachèlerie Marie-Laure, Denis-Meyere Auriane, Tisserand Delphine. Traçabilité des activités de recherche et gestion des connaissances. Guide pratique de mise en place. 2018.* *https://qualite-en-recherche.cnrs.fr/gt/tracabilite-des-activites-de-recherche/*
- *UNIL, Université de Lausanne. Les données de recherche.* *https://www.unil.ch/openscience/fr/home/menuinst/open-research-data/les-donnees-de-recherche.html*

Ressource *:*

DoRANum. L'origine et la description des données de la recherche. 15 mars 2022. *https://doranum.fr/plan-gestion-donnees-dmp/origine-description-donnees-recherche_10_13143_e9zh-w908/*

Exemples de données de recherche

Photo d'une peinture pariétale représentant une scène de chasse



	Code Provenance	Code accession	Pays d'origine	Nom de la provenance dans fichier data	Autre nom possible de la provenance	Coordonnées géographiques d'origine	altitude (m)	P (mm)	substrat	Poids de 1000 graines (g)	% graines vaines (méthode + trempage dans l'eau)	Présence 01-4800-01-01	01-4800-01-02	01-4800-02-01	01-4800-02-02	01-4800-02-03
1	1	6578	Espagne	Andalucia	Andalucia	36°20N, 6°05W	50	631	Luvisol	653.6	4.8	x	x	x	x	x
2	2	6577	Espagne	Catalunia/Lib	Catalunia	41°10N, 2°10E	150	710	Cambisol	769.2	3.7	x	x	x	x	x
3	3	6579	Espagne	Meseta	Meseta	41°30N, 4°25W	975	452	Cambisol	568.2	12.1	x	x	x	x	x
4	4	6580	Espagne	Cordillera	Cordillera	40°30N, 4°20W	900	1007	Cambisol	772.2	2.9	x	x	x	x	x
5	5	6581	Grèce	Sithonia /	Chalkidiki	40°14 N, 23°34	50 - 400	439	granite	689.7	0.4	x	x	x	x	x
6	6	7301	Grèce	Metochi	Metochi	38°06 N, 21°39	38	-	-	503.8	0	x	x	x	x	x
7	7	3913	Italie	Fenoglia (57)	Fenoglia 67	42°25 N, 11°17	2-5	680	dépots	714.3	12.7	x	x	x	x	x
8	9	6584	Italie	Migliarino	Migliarino 75	43°47 N, 10°17	5-10	960	dépots	1096.5	9.1	x	x	x	x	x
9	9	6588	Liban	Beit Mounzer	Beit Mounzer	34°29N, 35°81E	1300-1400	-	grès/basal	1041.7	6.2	x	x	x	x	x
10	10	6585	Liban	Kornael	Kornael	33°64N, 35°77E	1300-1400	-	grès/basal	788.6	9.1	x	x	x	x	x
11	11	6586	Liban	Qsaibe	Qsaibe	33°88N, 35°66E	600-700	-	grès/basal	784.3	7.6	x	x	x	x	x
12	12	6587	Liban	Bkassine	Bkassine	33°57N, 35°65E	800	-	grès/basal	912.4	4.8	x	x	x	x	x
13	13	36658	Portugal	Norte do	Norte do Tejo	40°40 N, 7°54	400	1296	-	?	?	x				
14	14	36658	Portugal	Sul do Tejo	Sul do Tejo	38°23 N, 8°31	40	589	-	?	?	x				
15	15	3918	Turquie	Antalya/Serik	Antalya (n°	36°52 N, 31°01	10	1030	-	641	7.7	x	x	x	x	x
16	16	36639	Turquie	Izmir/Bergai	Izmir (n°	39°12 N, 26°57	400	795	-	775.2	6.1	x	x	x	x	x
17	17	36640	Turquie	Antalya/Kumi	Antalya (n°	36°17 N, 30°29	5	1030	-	449.4	11.2	x	x	x	x	x
18	18	3920	Turquie	Canakkale	Canakkale	40°11 N, 26°16	20	626	-	597.7	6.1	x	x	x	x	x
19	19	36642	Turquie	Artvin	Artvin	41°11 N, 41°51	225	654	-	729.9	12.2	x	x	x	x	x
20	20	3923	Turquie	Mugla/Yatag	Yatagan (n°	37°22 N, 28°03	600	1302	-	574.7	4.4	x	x	x	x	x
21	21	36644	France	Hyeres	Hyères (Var)	43°05 N, 6°09 E	5	670	sables	915.8	8.7	x	x	x	x	x
22	22	36645	France	Villeneuvette	Villeneuvette	43°37 N, 3°24 E	130-220	775	calcaire	750	0.1	x	x	x	x	x
23	23	36646	France	Stes-Maries-	Saintes-	43°27 N, 4°26 E	5	543	sables	770	0.2	x	x	x	x	x
24	24	6596	France	Virassoun	Virassoun	43°12 N, 3°05 E	50-110	645	silice	754.7	0.2	x	x	x	x	x
25	25	36648	France	Le Vat	Le Val (Var)	43°26 N, 6°05 E	310	813	calcaire	900.9	0	x	x	x	x	x
26	26	36649	Turquie	Balikesir	Balikesir (n°	39°40 N, 27°06	379	603	-	537.6	7.1	x	x	x	x	x
27	27	6600	Maroc	Mezzine	Mezzine	36°96N,	710	7.900	grès	446.4	10.6			x		
28	28	6601	Maroc	Izarene	Izarène	34°16N,	1472	795	sol brun	936.3	14			x	x	x
29	29	6602	Maroc	Cap Spartel	Cap Spartel	35°47N,	100	-	grès	466.2	18.2			x	x	x
30	30	6603	Tunisie	Dar Chichou	Dar Chichou	36°57N, 9°58E	32	580	sable +	815.7	5.7			x	x	x
31	31	36651	Tunisie	Rimel	Rimel	37°14N, 9°58E	15	610	sable sur	775.2	6.5			x	x	x
32	32	6605	Tunisie	Bechateur	Bechateur	37°14N, 9°56E	45	610	sable	961.5	5.2			x	x	x
33	33	48380	Turquie	Mugla/Yatag	Turtuglu	37°22N, 28°03E	660	1201	-	656.6	10.6			x	x	x
34	34	3914	Turquie	Burnaz/Yalov	Yalova	40°52N, 29°22E	500	754	-	682.9	6.7			x	x	x
35	35	48382	Turquie	Antalya/Serik	Serik	36°52N, 31°01E	10	1030	-	815.7	16			x	x	x
36	36	48381	Espagne	Catalunia/Lib	Catalunia	41°46N, 2°48E	150	-	-	680.3	16.2			x	x	x
37	37	3917	Grèce	Kumupeli	Kumupeli	38°07N, 21°25E	0	-	Sablo-	513.3	23.9			x	x	x

Données sur les traits phénotypiques de *Pinus pinea* mesurés dans les jardins communs du réseau INRAE de génétique forestière pour la recherche et l'expérimentation (GEN4X)

cnrs Inist

6

Voici des exemples concrets de données de recherche dans différentes disciplines :

- Cette photo d'une peinture pariétale représentant une scène de chasse dans le désert du Wadi Ram, en Jordanie est un exemple de donnée pour un chercheur en archéozoologie qui étudie le rôle du monde animal au sein des sociétés humaines sans écriture.

  Source de l'image : https://pixabay.com/fr/photos/jordan-wadi-d%C3%A9sert-sable-paysage-4158477/

- Données sur les traits phénotypiques de *Pinus pinea* mesurés dans les jardins communs du réseau INRAE de génétique forestière pour la recherche et l'expérimentation (GEN4X). Ce jeu de données est composé d'observations telles que la hauteur totale, la circonférence, le nombre de cônes, la survie et la fourchaison, réalisées jusqu'en 2012 dans cinq tests de provenance de *Pinus pinea* du réseau français de jardins communs GEN4X (réseau de génétique forestière pour la recherche et l'expérimentation).

  Fady Bruno, Vauthier Denis, Michotey Celia. Phenotypic trait dataset of Pinus pinea from five common gardens of the INRAE GEN4X network. 2021. https://doi.org/10.15454/MQKQ1O

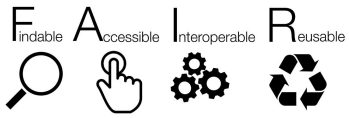
02

Principes FAIR

Principes FAIR

4 principes à respecter pour garantir une utilisation optimale des données de recherche et des métadonnées associées, à la fois **par les hommes et par les machines**

- **F** (Findable) = Facile à trouver
- **A** (Accessible) = Accessible
- **I** (Interoperable) = Interopérable
- **R** (Reusable) = Réutilisable



Principes admis par les différentes communautés scientifiques au niveau international, ainsi que par les financeurs (ex : Commission européenne, ANR, etc.)

Applicables tout au long du cycle de vie des données

cnrs Inist

8

Tout au long de cette présentation des petites bulles vous permettront de repérer à quel(s) principe(s) FAIR correspondent principalement les bonnes pratiques décrites.

Ressource *:*

DoRANum. Les principes FAIR. 4 décembre 2019. https://doranum.fr/enjeux-benefices/principes-fair_10_13143_z7s6-ed26/

Principe F

F

Le principe F facilite la découverte des données par les humains et les systèmes informatiques :

- Catalogues ou entrepôts
- Identifiant pérennes
- Métadonnées riches

Facile à trouver

cnrs Inist

9

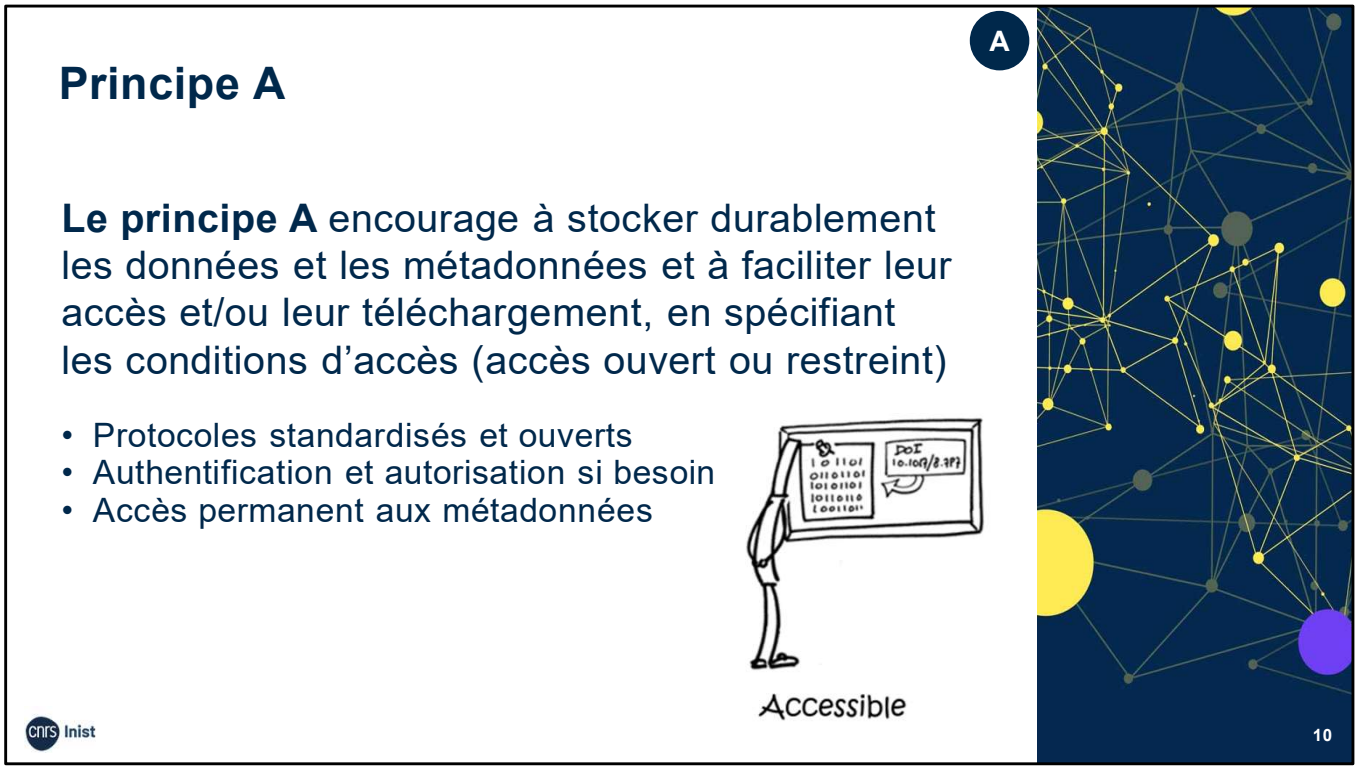
Le principe F signifie que vos données doivent être faciles à trouver par des humains et par des systèmes informatiques.

Bonnes pratiques :

- Déposer vos jeux de données dans un entrepôt de données
- Leur attribuer un identifiant pérenne
- Les décrire précisément à l'aide de métadonnées

Source *:*

- *INRAe, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. Produire des données FAIR. https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair*



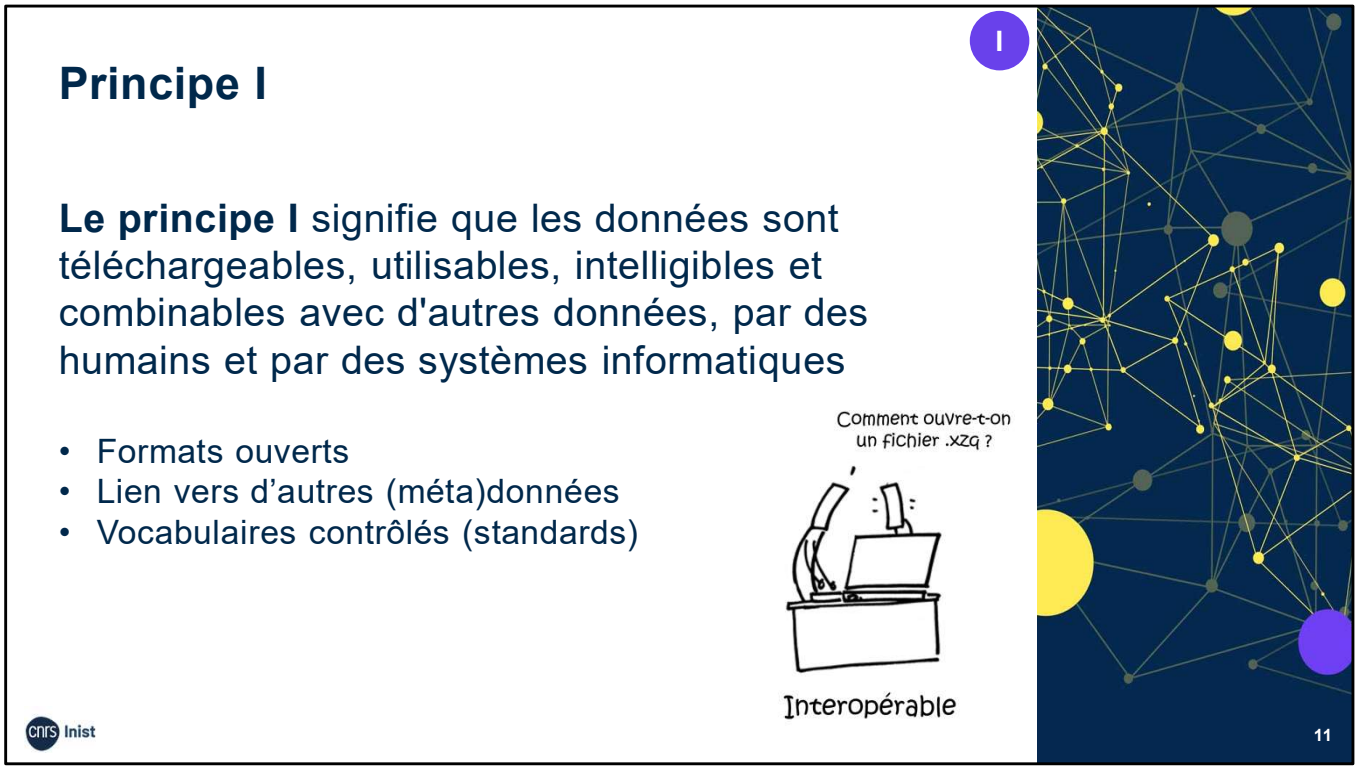
Le principe A encourage à stocker durablement les données et les métadonnées et à faciliter leur accès et/ou leur téléchargement, en spécifiant les conditions d'accès (accès ouvert ou restreint).

Bonnes pratiques :

- Protocoles de communication standards : les données doivent être rendues accessibles soit via une URL, soit via un téléchargement de fichier(s). L'idée est de ne pas avoir besoin de recourir à des outils spécialisés ou propriétaires pour pouvoir accéder aux données, exemples : Skype, Microsoft Exchange…
- Préférer le dépôt dans des entrepôts certifiés qui proposent un accès ouvert.
- Définir qui a le droit d'accéder aux (méta)données, par quelle procédure (ouverture totale ou demande d'accès par authentification ou demande d'autorisation)
- Les métadonnées restent accessibles même si les données ne le sont pas ou plus.

***Sources** :*

- *INRAe, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. Produire des données FAIR. https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair*
- *Université Paris-Saclay. Produire des données FAIR. https://www.universite-paris-saclay.fr/recherche/science-ouverte/produire-des-donnees-fair*



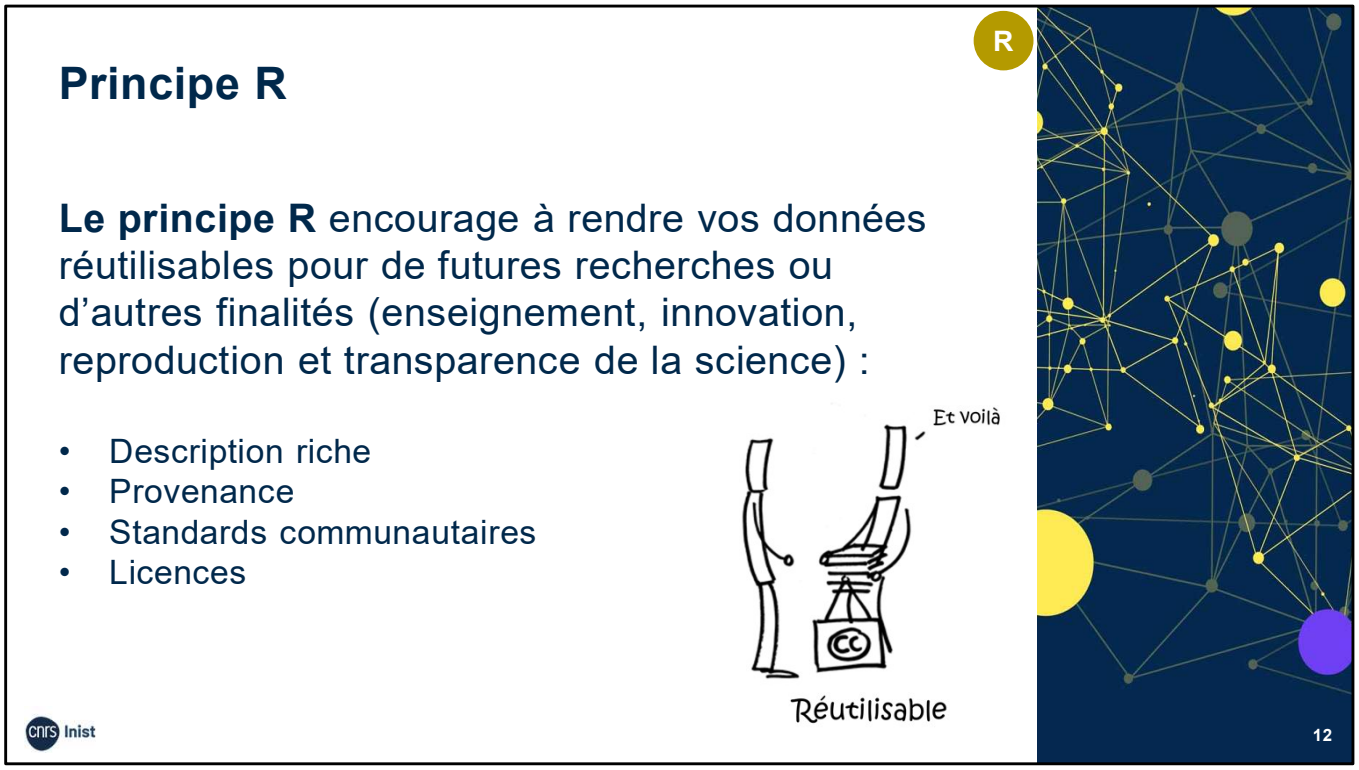
Le principe I signifie que les données sont téléchargeables, utilisables, intelligibles et combinables avec d'autres données, par des humains et des systèmes informatiques.

Bonnes pratiques :

- Utiliser des formats ouverts et indépendants.
- Utiliser des vocabulaires standards, des thésaurus, des ontologies… Le but étant de « parler le même langage ».
- Contextualiser les données en indiquant les liens vers d'autres données et métadonnées ou vers des publications.

Source *:*

- *INRAe, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. Produire des données FAIR. https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair*



Le principe R encourage à rendre vos données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction et transparence de la science). Les métadonnées et les données doivent être bien décrites afin de pouvoir être reproduites et/ou combinées dans différents contextes.

Bonnes pratiques :

- Les données sont très précisément décrites avec le plus d'informations possibles (métadonnées).
- Leur provenance est renseignée.
- Elles répondent à des standards communautaires pertinents pour le domaine.
- Il est important de leur attribuer une licence d'utilisation.

***Source** :*

- *INRAe, Institut national de recherche pour l'agriculture, l'alimentation et l'environnement. Produire des données FAIR. https://science-ouverte.inrae.fr/les-donnees-et-le-numerique-scientifiques/produire-des-donnees-fair*

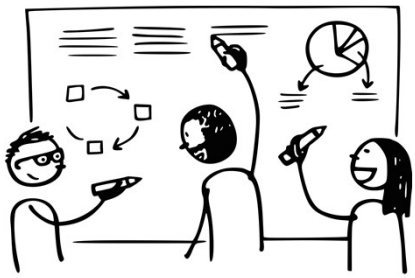
Afin de mieux saisir à quoi servent les principes FAIR, voici une liste de ce qu'ils apportent concrètement au chercheur ainsi qu'à la communauté scientifique.

Pour le chercheur

- Bonnes pratiques à adopter.
- C'est aussi un gain de temps dans la gestion du projet de recherche.
- Le chercheur gagne en visibilité.
- Ses données sont plus accessibles.
- L'intégrité scientifique est facilitée.
- Cela favorise les collaborations.

Qu’apportent les Principes FAIR ?
Pour la communauté scientifique

- Accès facile à des données publiques et réutilisables
- Gain de temps et d’argent : ne pas recréer des données déjà existantes
- Meilleure reproductibilité de la recherche



cnrs Inist

14

Pour la communauté scientifique

- L’accès à des données publiques et réutilisables est facilité, ainsi que l’accès à des corpus utiles pour d’autres domaines.
- L’Interopérabilité des données est meilleure.
- Il y a également un gain de temps et d’argent puisque des données déjà existantes ne sont pas recréées.
- Cela permet une meilleure reproductibilité de la recherche.

03

Cycle de vie des données de recherche

15

Cycle de vie des données de recherche

C'est l'ensemble des étapes

- de gestion
- de conservation
- de diffusion des données de recherche associées aux activités de recherche

cnrs Inist

16

« Le cycle de vie des données de la recherche est l'ensemble des étapes de gestion, conservation, diffusion et réutilisation des données scientifiques, associées aux activités de recherche. »

Deboin Marie-Claude. Découvrir de nouveaux métiers liés aux données de la recherche. CIRAD. 5 p. 5 octobre 2018. https://doi.org/10.18167/coopist/0061

Nous allons détailler les 6 étapes dans notre présentation.

Source de l'image *:*

D'après Research data lifecycle – UK Data Service.

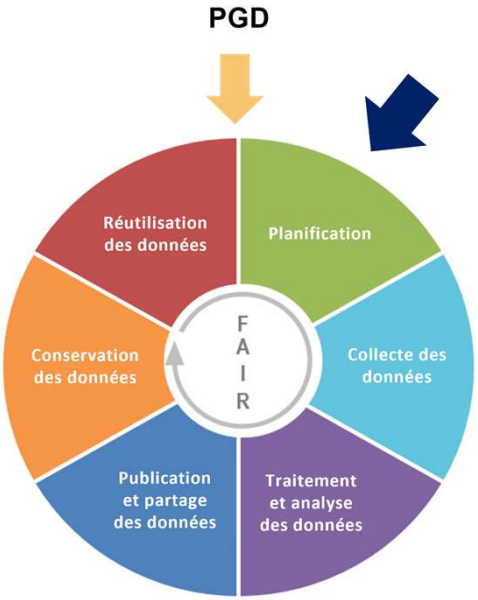
Ressource *:*

DoRANum. Le cycle de vie des données de recherche. 4 février 2021. https://doranum.fr/enjeux-benefices/le-cycle-de-vie-des-donnees-de-recherche_10_13143_gzj2-j593/

Étape 1 :
Planification (PGD/DMP)

17

Place dans le cycle de vie des données



cnrs Inist

- Le DMP ou PGD est un document de quelques pages, qui permet de définir et de mettre en place les actions indispensables pour gérer les données qui vont être collectées ou produites au cours d'un projet.
- Il explique comment sont gérées les données depuis leur création ou collecte jusqu'à leur partage et leur archivage.
- Le DMP est évolutif : sa rédaction démarre dès le début d'un projet, mais il va être mis à jour régulièrement en fonction de l'évolution du projet. Vous n'avez pas à répondre à toutes les questions du DMP dans sa phase initiale mais il est bon de réfléchir à tous les points abordés dans le DMP. Ce document est là pour faciliter la gestion des données de recherche de votre projet.
- Surtout, le DMP aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée.

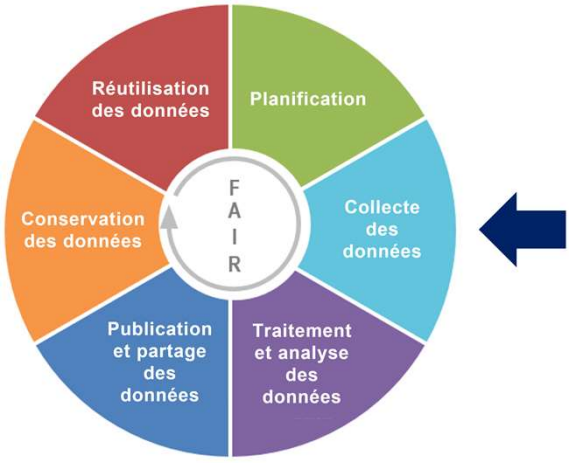
***Ressource** :*

- *DoRANum. Cours introductif sur le Plan de Gestion de Données (PGD). 27 avril 2022. https://doranum.fr/plan-gestion-donnees-dmp/cours-introductif-sur-le-plan-de-gestion-de-donnees_10_13143_t3j4-vn03/*

Étape 2 :
Création, collecte et description des données de recherche

20

Place dans le cycle de vie des données



cnrs Inist

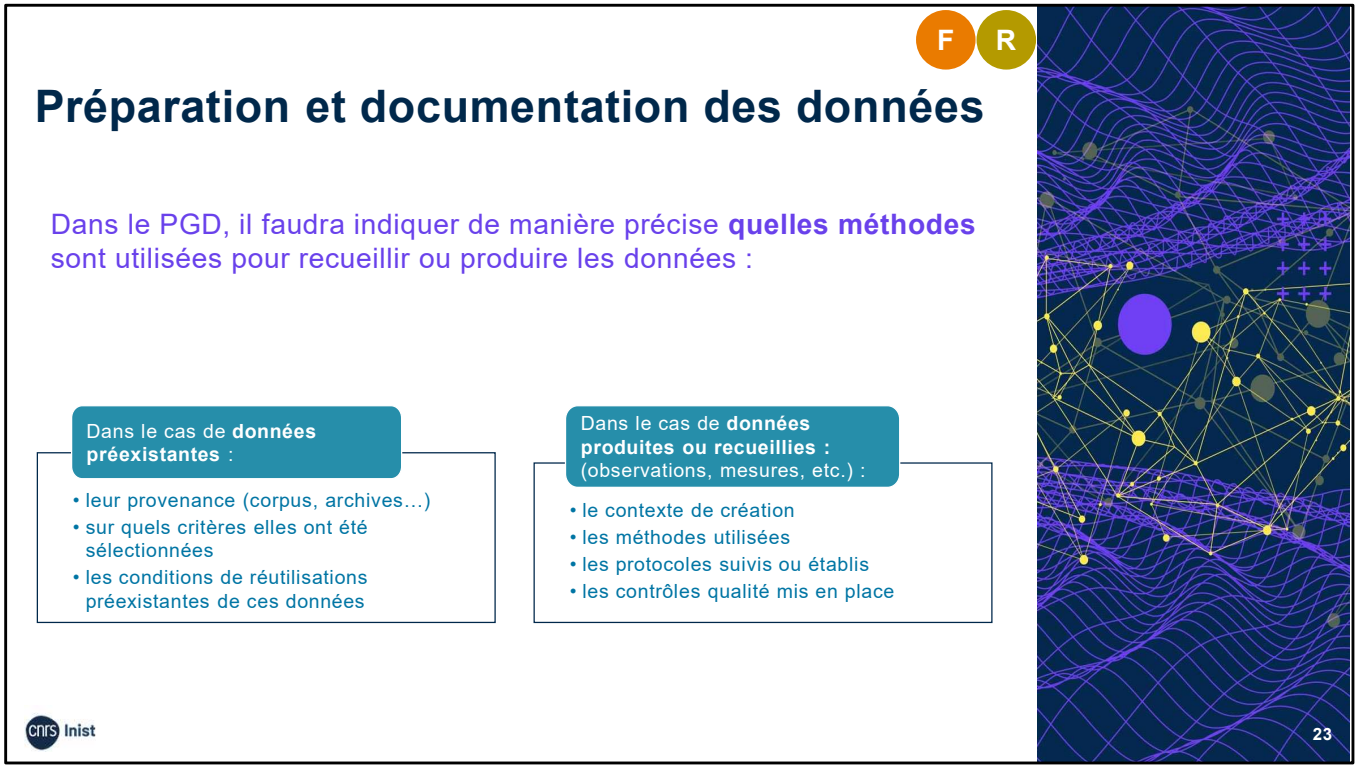
21

Selon le projet, les données de la recherche peuvent être :

- Produites ou recueillies : ce sont des données créées spécifiquement durant le projet. Elles sont élaborées, générées lors d'activités de recherche (collectes sur le terrain, observations, des mesures...).

- **Préexistantes** : ce sont des données déjà existantes (provenant de corpus, archives...) qui sont utilisées pour le projet. Les données utilisées peuvent avoir été recueillies initialement dans un autre contexte que celui de la recherche mais elles sont utilisées comme données de recherche dans le cadre du projet.

Selon que les données sont produites ou préexistantes, leur description ne contiendra pas les mêmes informations.



Il est important de **bien documenter les données** de manière à ce qu'elles soient faciles à trouver et réutilisables.

Ces informations devront apparaitre dans les métadonnées et/ou dans un fichier Readme. Elles sont indispensables pour que d'autres chercheurs puissent reproduire les résultats (reproductibilité et réplicabilité).

La **reproductibilité** est la capacité, par une équipe différente, de reproduire une expérience, sans se fier au dispositif expérimental et aux codes logiciels développés par l'équipe d'origine.

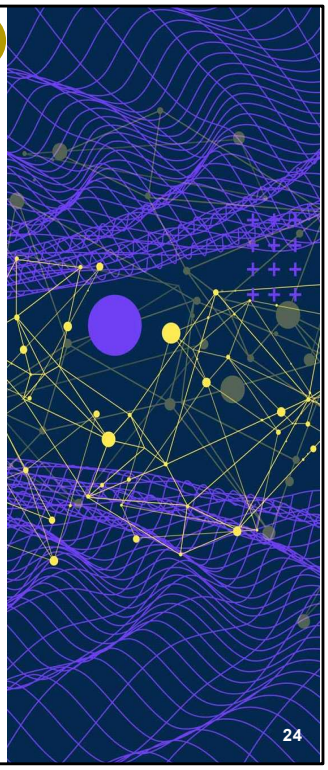
La **réplicabilité** est la capacité, par une équipe différente, de reproduire une expérience en ré-utilisant le même dispositif expérimental décrit (y compris les codes logiciels).

F R

Préparation et documentation des données

Il est impératif de **bien préparer et documenter ses données** afin d'optimiser le stockage, le partage, l'archivage et la réutilisation

Attention aux **données sensibles, personnelles ou confidentielles** : prendre les précautions nécessaires afin de respecter les règles juridiques et éthiques en vigueur



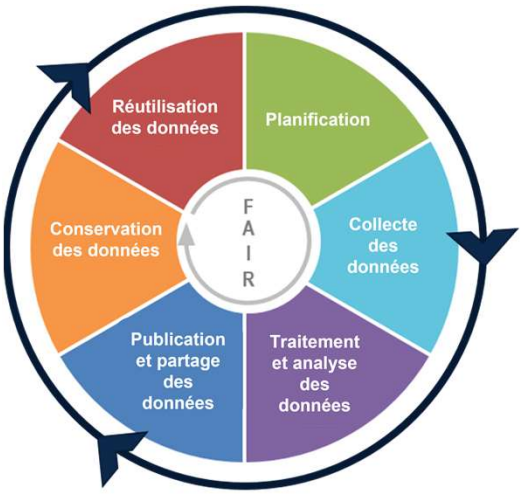
cnrs Inist

24

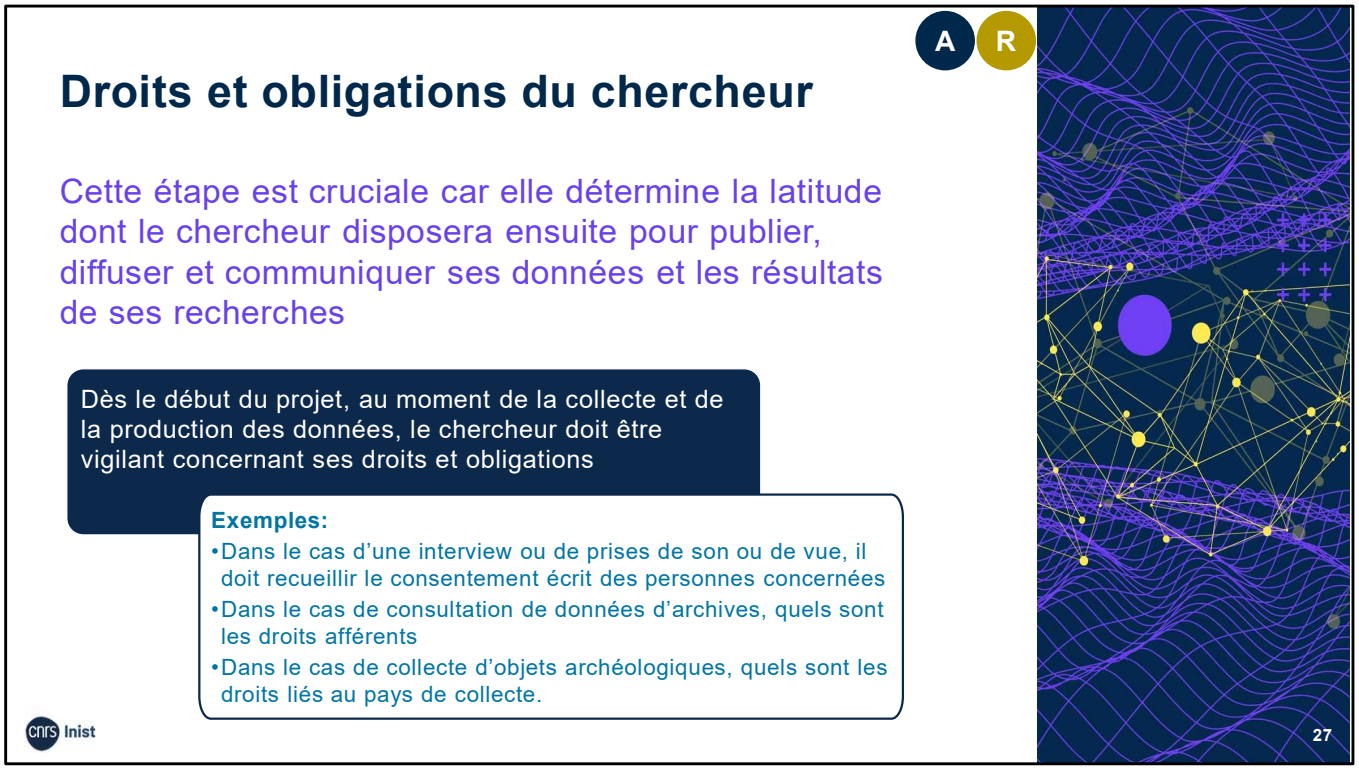
Aspects juridiques et éthiques

25

Place dans le cycle de vie des données

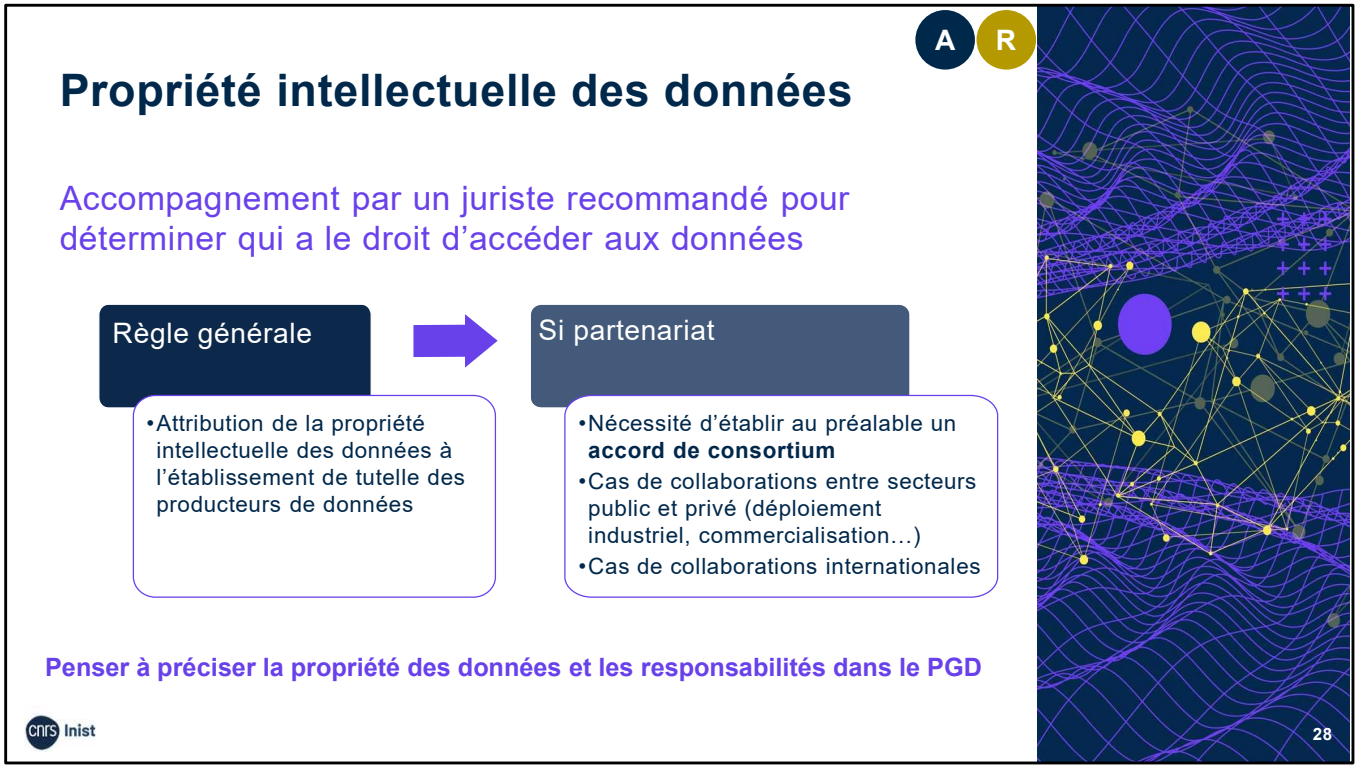


cnrs Inist



Ressource *:*

- *Delplanque Catherine, Lamrini Nawale, Leclère Fabrice, Maurel Lionel, et al. Guide : Règlement Général pour la Protection des Données. 2019.* *https://www.u-plum.fr/guide-reglement-general-pour-la-protection-des-donnees/*



Attention, **le PGD n'a aucune valeur juridique**.

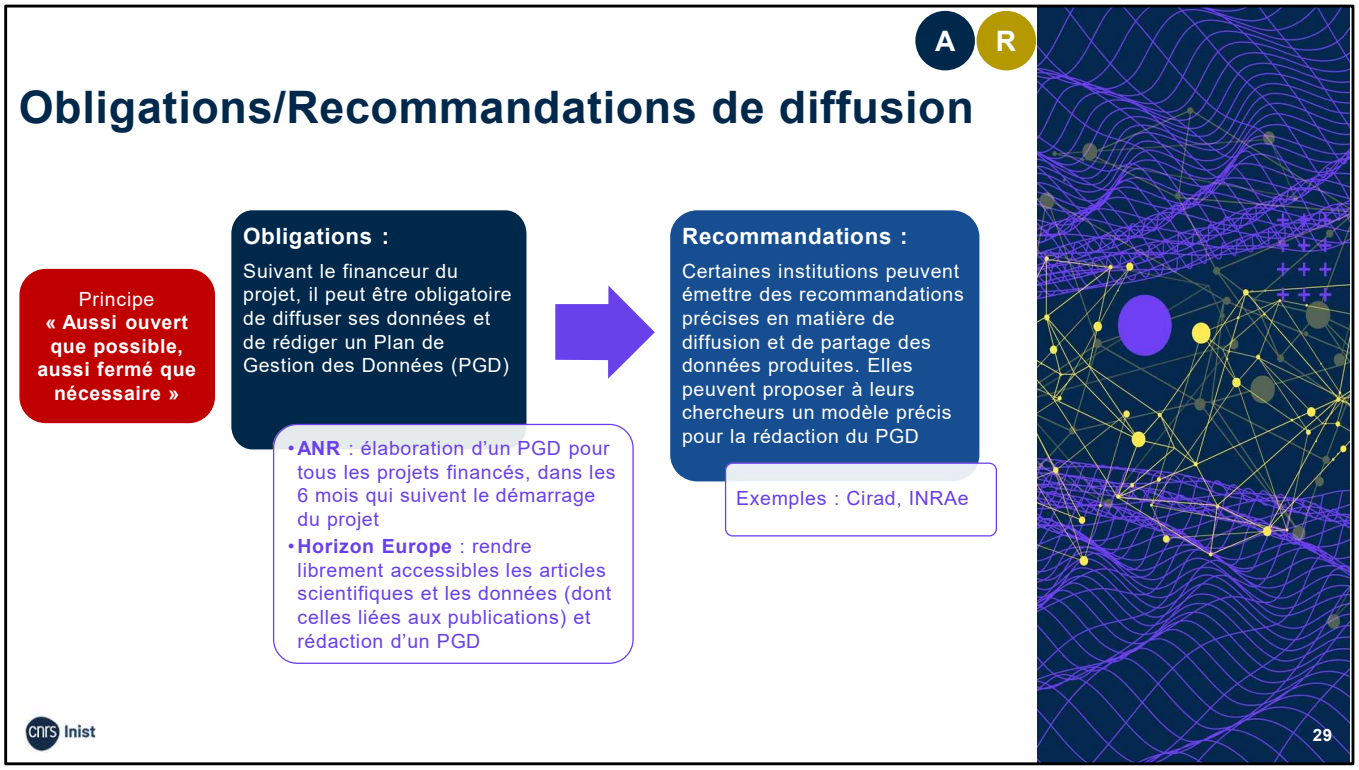
Il ne s'agit pas du même droit que les publications (droit d'auteur principalement).

Les données relèvent d'un régime lié au droit des bases de données. Dans ce cas, **le droit de propriété appartient légalement au « producteur » de la base de données**, compris au sens de la personne qui réalise l'investissement financier et matériel nécessaire à la constitution de la base. Il s'agira donc en général de **l'établissement de tutelle des chercheurs** qui **sera considéré comme le titulaire effectif du droit de propriété**.

Mais si ce droit existe formellement, il ne peut plus être opposé aux droits des ré-utilisateurs des données (principe d'ouverture des données). En effet, la loi pour une République numérique a explicitement « neutralisé » le droit des bases de données des administrations pour faire primer le principe de libre réutilisation. Il en résulte que **les données produites par les chercheurs sont** bien **comprises dans le principe d'ouverture par défaut**.

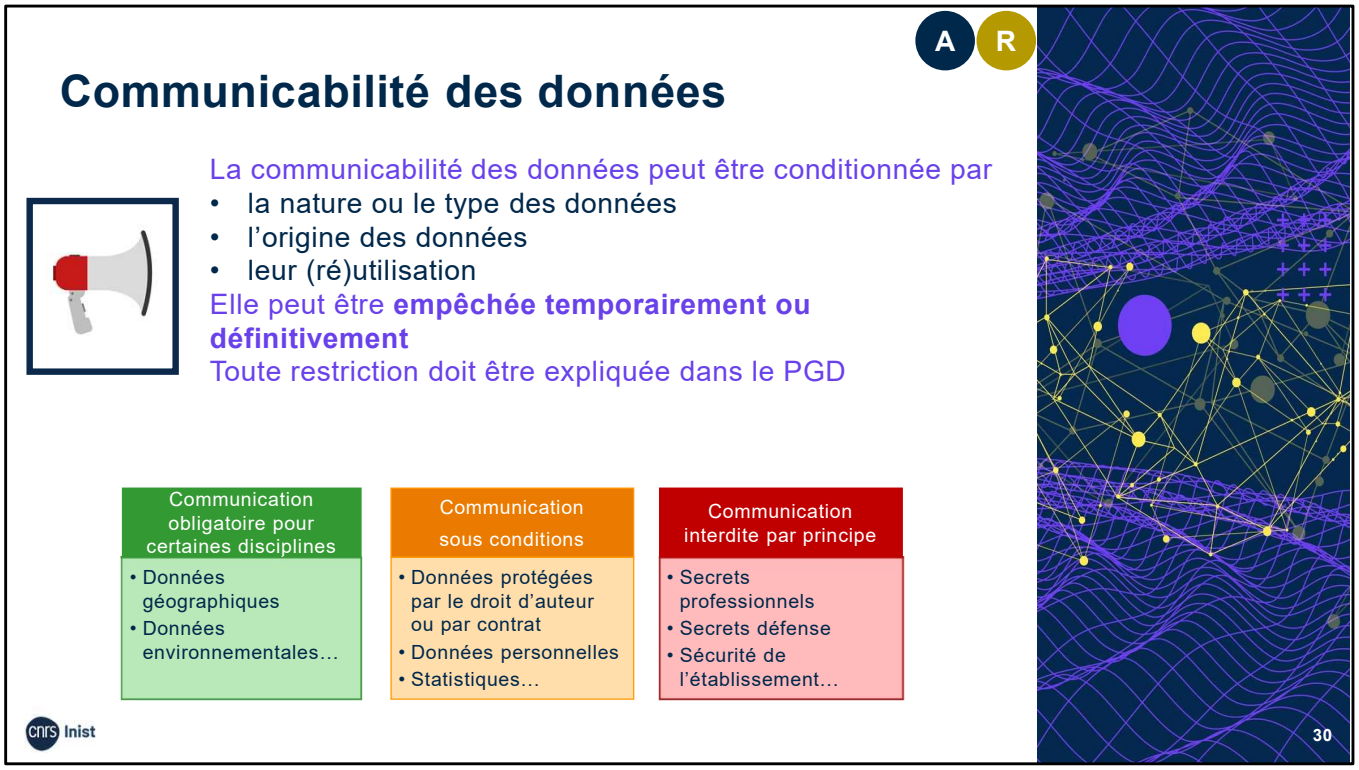
Ressources *:*

- *DoRANum. Questions juridiques liées aux données de recherche : interview de Lionel Maurel. https://doranum.fr/aspects-juridiques-ethiques/questions-juridiques-liees-aux-donnees-de-la-recherche_10_13143_xjgm-hb78/*
- *Ginouvès Véronique, Gras Isabelle, et al. La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques. https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques*
- *Bordignon Frédérique, Boistel Romain, Du Pasquier Delphine. Qui a les droits, quelles obligations ? https://espacechercheurs.enpc.fr/sites/default/files/logigramme_a_plat_2.pdf*



Ressources *:*

- *Becard Nicolas, Castets-Renard Céline, Chassang Gauthier, Dantant Martin, Freyt-Caffin Laurence, Gandon Nathalie, Martin Caroline, Martelletti Andrea, Mendoza-Caminade Alexandra, Morcrette Nathalie, Neirac Claire. Ouverture des données de recherche. Guide d'analyse juridique en France. Décembre 2017.* *http://dx.doi.org/10.15454/1.481273124091092E12*
- *Bordignon Frédérique, Boistel Romain, Du Pasquier Delphine. Qui a les droits, quelles obligations ?* *https://espacechercheurs.enpc.fr/sites/default/files/logigramme_a_plat_2.pdf*
- *Ginouvès Véronique, Gras Isabelle, et al. La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques.* *https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques*



Données de différentes **natures** : brutes, dérivées, formatées, nettoyées, primaires, secondaires, traitées…

Données de différents **types** : archives, audios, vidéos, bases de données, codes sources, données géospatiales, images, photographies, langages de programmation, matérielles et physiques, modèles, visualisations, 3D, numériques, textuelles, numérisations, scans, qualitatives, quantitatives, statistiques…

L'**éthique** nous invite à réfléchir aux valeurs qui motivent nos actes et à leurs conséquences et fait appel à notre sens moral et à celui de notre responsabilité.

La **déontologie** réunit les devoirs et obligations imposés à une profession, une fonction ou une responsabilité.

L'**intégrité scientifique** concerne la « bonne » conduite des pratiques de recherche. Elle désigne l'ensemble des règles et valeurs qui doivent régir l'activité de recherche pour en garantir le caractère honnête et scientifiquement rigoureux. Le respect de ces règles est une condition indispensable du maintien du lien de confiance accordée par la société aux acteurs de la recherche.

Sources *:*

- *CNRS, Centre national de la recherche scientifique. Responsabilité de recherche. 8 janvier 2024. https://www.cnrs.fr/fr/le-cnrs/responsabilites/responsabilite-de-recherche*
- *ANR. L'intégrité scientifique. https://anr.fr/fr/lanr/engagements/lintegrite-scientifique/*

Des normes, chartes, déclarations, codes et politiques en éthique et en intégrité scientifique encadrent les pratiques pour l'ensemble de nombreux acteurs internationaux de la recherche.

Ressources *:*

- *ANR, Agence nationale de la recherche. Politique en matière d'éthique et d'intégrité scientifique. Août 2014. https://anr.fr/fileadmin/documents/2014/Politique-ethique-integrite-scientifque-aout-2014.pdf*
- *Ginouvès Véronique, Gras Isabelle et al. La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques : https://hal-amu.archives-ouvertes.fr/page/guide-de-bonnes-pratiques*
- *Questions éthique et droit en SHS : https://ethiquedroit.hypotheses.org/*

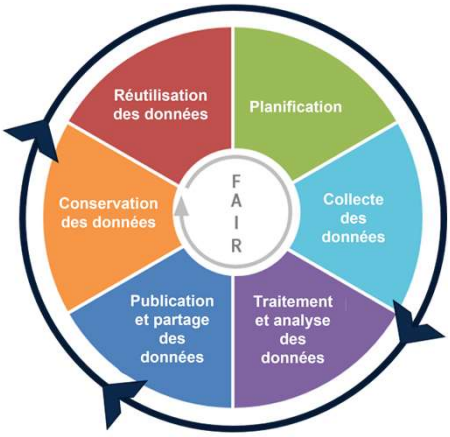
En résumé

Pour être sûr que les données sont accessibles et réutilisables, veillez à respecter

Les aspects juridiques	Les aspects éthiques
• Propriété intellectuelle • Accords de consortium • Obligations de diffusion • Communicabilité des données • ...	• Comité d'éthique • Normes • Chartes

Métadonnées

Place dans le cycle de vie des données



Il est recommandé de renseigner les métadonnées au fur et à mesure de l'avancée du projet

Au moment du partage puis de l'archivage pérenne, des métadonnées spécifiques seront à renseigner

Définition – utilisation des métadonnées

- Les métadonnées permettent de décrire plus précisément les données
- Elles sont à renseigner au fur et à mesure pour chaque jeu de données
- La boite de conserve = jeu de données / l'étiquette = métadonnées

Sans métadonnées

Avec métadonnées

cnrs Inist

35

Renseigner les métadonnées au fur et à mesure pour chaque jeu de données, pour toutes les étapes du cycle de vie des données.

Ressource *:*

DoRANum. Cours introductif sur les métadonnées. 27 avril 2022.

https://doranum.fr/metadonnees-standards-formats/cours-introductif-sur-les-metadonnees_10_13143_vwce-g965/

Métadonnées embarquées et enrichies

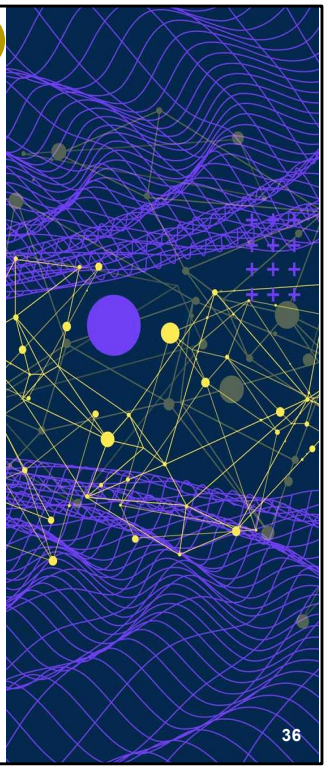
Compléter les métadonnées embarquées par des métadonnées enrichies

Métadonnée embarquée : produite automatiquement par les appareils (de prise de vue ou de son, de mesure…)

Ex : données GPS, type d'appareil, date, calibrage technique, etc.

Métadonnée enrichie : ajoutée par l'auteur

Ex : mots-clés, sujet, auteur, laboratoire ou organisme, nom du projet, licence, etc.



Un **schéma de métadonnées** est une **construction organisée d'info**rmations. C'est une **liste structurée** et **composée d'éléments descriptifs reliés entre eux**.

L'utilisation d'un standard de métadonnées, notamment disciplinaire, est un élément clé pour atteindre un haut degré de respect des principes FAIR :

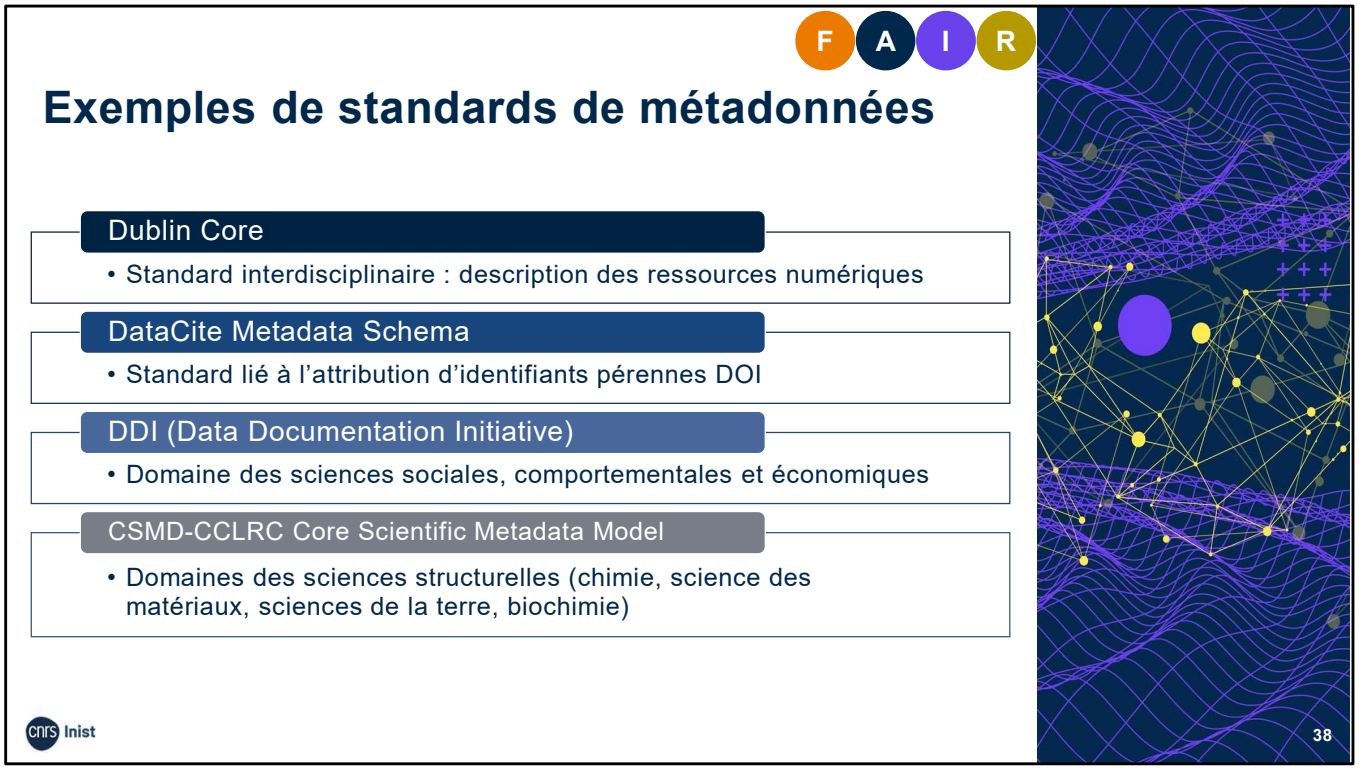
- **Facile à trouver** : une donnée n'est souvent trouvable que par les éléments de métadonnées indexés dans le moteur de recherche consulté.
- **Accessible** : notamment grâce aux métadonnées.
- **Interopérable** : grâce au standard commun qui facilite les traitements informatiques.
- **Réutilisable** : grâce à la provenance décrite dans les métadonnées, la licence attribuée et au recours à un standard disciplinaire.

Attention : opter pour un schéma de métadonnées plutôt qu'un standard rend vos données beaucoup moins interopérables et moins FAIR. Seules les métadonnées renseignées selon un standard seront interopérables. Les métadonnées spécifiques le seront moins voire pas du tout.

N'hésitez pas à vous faire aider par un documentaliste ou un informaticien.

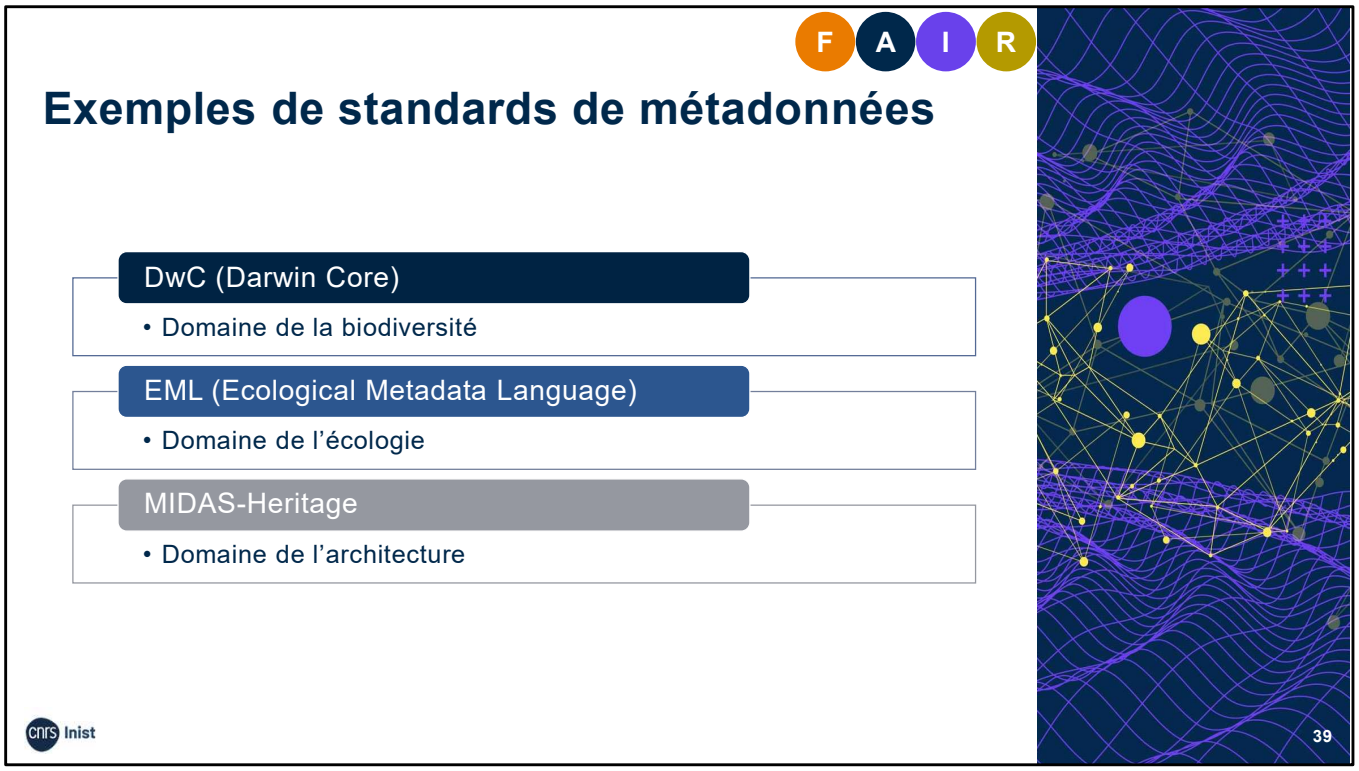
Ressource *:*

- *DoRANum. Les schémas de métadonnées.25 juin 2018. https://doranum.fr/metadonnees-standards-formats/schemas-metadonnees_10_13143_j8bc-0z74/*



Exemples de standards de Métadonnées :

- **Dublin Core (interdisciplinaire) :** description des ressources numériques. http://dublincore.org/
- **DataCite Metadata Schema :** métadonnées enregistrées dans le DataCite Metadata Store lors de la création d'un DOI pour un jeu de données. Permet l'identification précise et cohérente des données à des fins de citation et de réutilisation. https://schema.datacite.org/
- **DDI (Data Documentation Initiative) :** domaine des sciences sociales, comportementales et économiques. http://www.ddialliance.org/
- **CSMD-CCLRC Core Scientific Metadata Model :** domaines des sciences structurelles (chimie, science des matériaux, sciences de la terre, biochimie) http://icatproject-contrib.github.io/CSMD/



Exemples de standards de Métadonnées :

- **DwC (Darwin Core) :** domaine de la biodiversité. http://rs.tdwg.org/dwc/
- **EML (Ecological Metadata Language) :** très développé en écologie. En grande partie conçu pour décrire des ressources numériques. Il peut également être utilisé pour décrire des ressources non numériques telles que des cartes papier ou d'autres médias. https://knb.ecoinformatics.org/external//emlparser/docs/index.html
- **MIDAS-Heritage :** domaine de l'architecture. https://historicengland.org.uk/images-books/publications/midas-heritage/

Ressources pour trouver des standards de métadonnées :

- *DoRANum. Les standards de métadonnées : pourquoi, lequel ? 13 avril 2021. https://doranum.fr/metadonnees-standards-formats/standard-metadonnees_10_13143_y5py-w521/*
- *FAIRsharing.org. Répertoire de standards de métadonnées en Sciences de la Vie. https://fairsharing.org/search?fairsharingRegistry=Standard*
- *RDA, Research Data Alliance. Metadata Standards Catalog. https://rdamsc.bath.ac.uk/*
- *DCC, Digital Curation Centre. Disciplinary Metadata. https://www.dcc.ac.uk/guidance/standards/metadata*

*Consulter également les **informations fournies par les entrepôts de données sur les standards de métadonnées**.*

Exemple : Le Dublin Core

Élément simple	*Définitions*	*Exemple d'élément spécifique*
Title	nom de la ressource	
Subject	thème du contenu de la ressource	
Description	résumé, table des matières…	
Creator	auteur principal de la ressource	
Publisher	entité responsable de la diffusion de la ressource	
Contributor	co-auteurs associés à l'élaboration de la ressource	***Date Created***, *Date copyrighted, Date Valid, Date Available, Date Modified,* ***Date Accepted***, *Date Submitted, Date Issued*
Date	date de création ou mise à disposition	
Type	nature du contenu : image, son texte…	
Format	format ou taille de la ressource	
Identifier	référence univoque, DOI, URL, ISSN…	
Source	référence à une ressource à partir de laquelle la ressource actuelle a été dérivée ou créée	***Has Format*** *(Les relations de transformation de format sont celles où une ressource a été dérivée d'une autre à l'aide d'une technologie de reproduction ou de reformatage qui n'est pas fondamentalement une interprétation mais une représentation.)* ***isVersionOf*** *(Les relations de version sont celles où une ressource est un état ou une parution historique d'une autre ressource par le même créateur)*
Language	langue originale de la ressource	
Relation	référence à une ressource apparentée	
Coverage	périmètre spatial et temporel	
Rights	informations sur les droits associés à la ressource	

cnrs Inist 40

Le Dublin Core est un **standard international et multidisciplinaire**.

Il **comporte 15 éléments** (= minimum exigé) ayant trait au contenu, à la propriété intellectuelle, à la version.

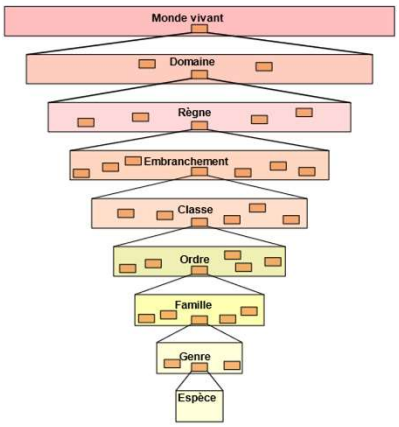
Ces 15 éléments restent cependant insuffisants pour représenter finement certaines données dans un univers dédié pour des utilisateurs et applications spécifiques. C'est pourquoi une quarantaine d'éléments plus spécifiques, qui constituent le Dublin Core étendu, viennent préciser les éléments simples.

Sources *:*

- *DoRANum. Métadonnées, standards, formats : fiche synthétique. 27 novembre 2017. https://doranum.fr/metadonnees-standards-formats/metadonnees-standards-formats-fiche-synthetique_10_13143_vbjs-6288/*
- *Wikipédia. Dublin Core. https://fr.wikipedia.org/wiki/Dublin_Core*

Enrichissement des métadonnées

- Utiliser des vocabulaires contrôlés disciplinaires utilisés couramment : (lexiques, thesaurus, ontologies…)
- Exemples :
  - Codex de médicaments
  - Classifications taxonomiques
  - Nomenclature internationale des formules chimiques
- Cela augmentera la capacité des données à être combinées avec d'autres données (interopérabilité)



cnrs Inist 41

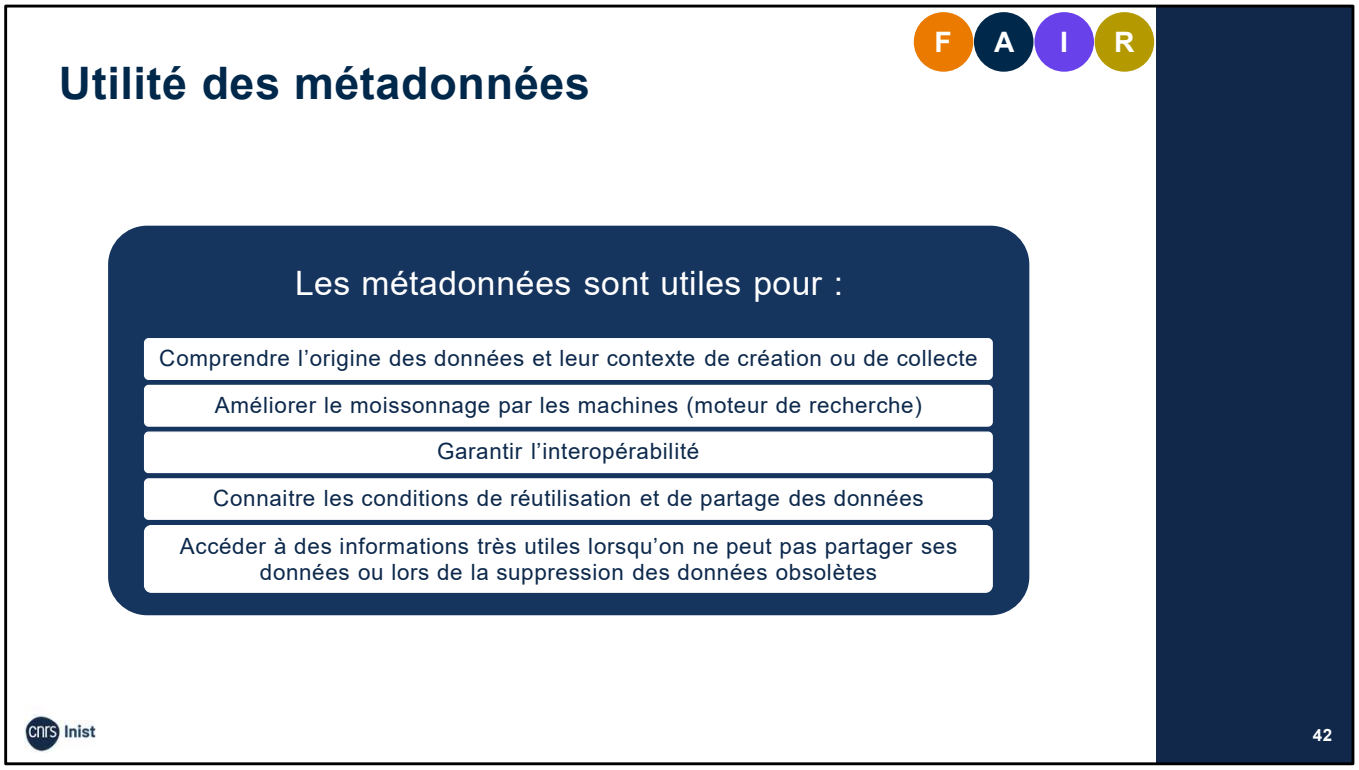
« Le **Thésaurus** est un ensemble organisé de termes contrôlés et normalisés représentant les concepts d'un domaine de connaissance. Les termes sont reliés entre eux par des relations de synonymie (terme équivalent), de hiérarchie (terme générique et terme spécifique) et d'association (terme associé) ; chaque terme appartient à une catégorie ou domaine. » *(Wikipedia. Thésaurus documentaire. https://fr.wikipedia.org/wiki/Th%C3%A9saurus_documentaire)*

« L'**ontologie** est un ensemble organisé de termes/concepts avec des relations sémantiques variées décrivant un domaine de connaissance. Exemple : FOAF (FriendOfAFriend). *(Ecole thématique E-Envir 2021. Interoperable & Reuse. Introduction aux concepts clefs et immersion. 2 au 5 novembre 2021. https://e-envir-21.sciencesconf.org/data/pages/6_E_ENVIR21_INTEROP_REUSE.pdf)*

« La **taxonomie** ou taxinomie est une branche des sciences naturelles qui a pour objet l'étude de la diversité du monde vivant. Cette activité consiste à décrire et circonscrire en termes d'espèces les organismes vivants et à les organiser en catégories hiérarchisées appelées taxons. *(Wikipedia. Taxonomie. https://fr.wikipedia.org/wiki/Taxonomie)*

« Par extension, [la taxonomie est une] classification, suite d'éléments formant des listes qui concernent un domaine, une science. » *(Dictionnaire Larousse. https://www.larousse.fr/dictionnaires/francais/taxinomie/76893)*

Source de l'image : Ariel Provost. File:Taxonomic hierarchy.svg (figure renversée : espèce en bas), CC BY-SA 4.0. https://commons.wikimedia.org/w/index.php?curid=114679881



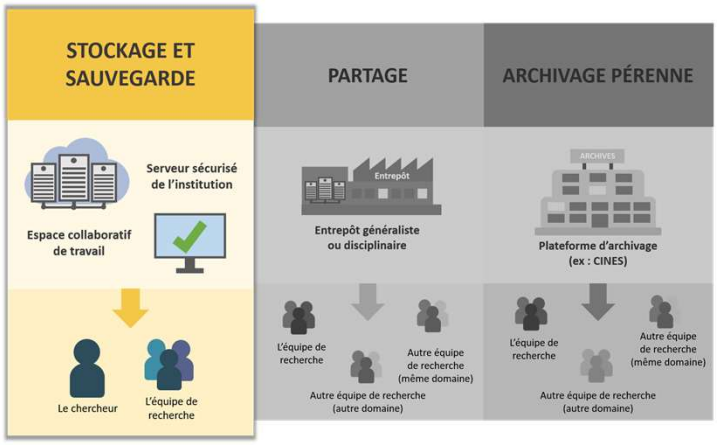
Quoiqu'il arrive aux données, les métadonnées resteront toujours accessibles. Il est important de mentionner que les données ont été supprimées et pourquoi.

Étape 3 :
Traitement et analyse
Stockage sécurisé durant le projet

Place dans le cycle de vie des données

Stockage et sauvegarde

Stockage et sauvegarde des données



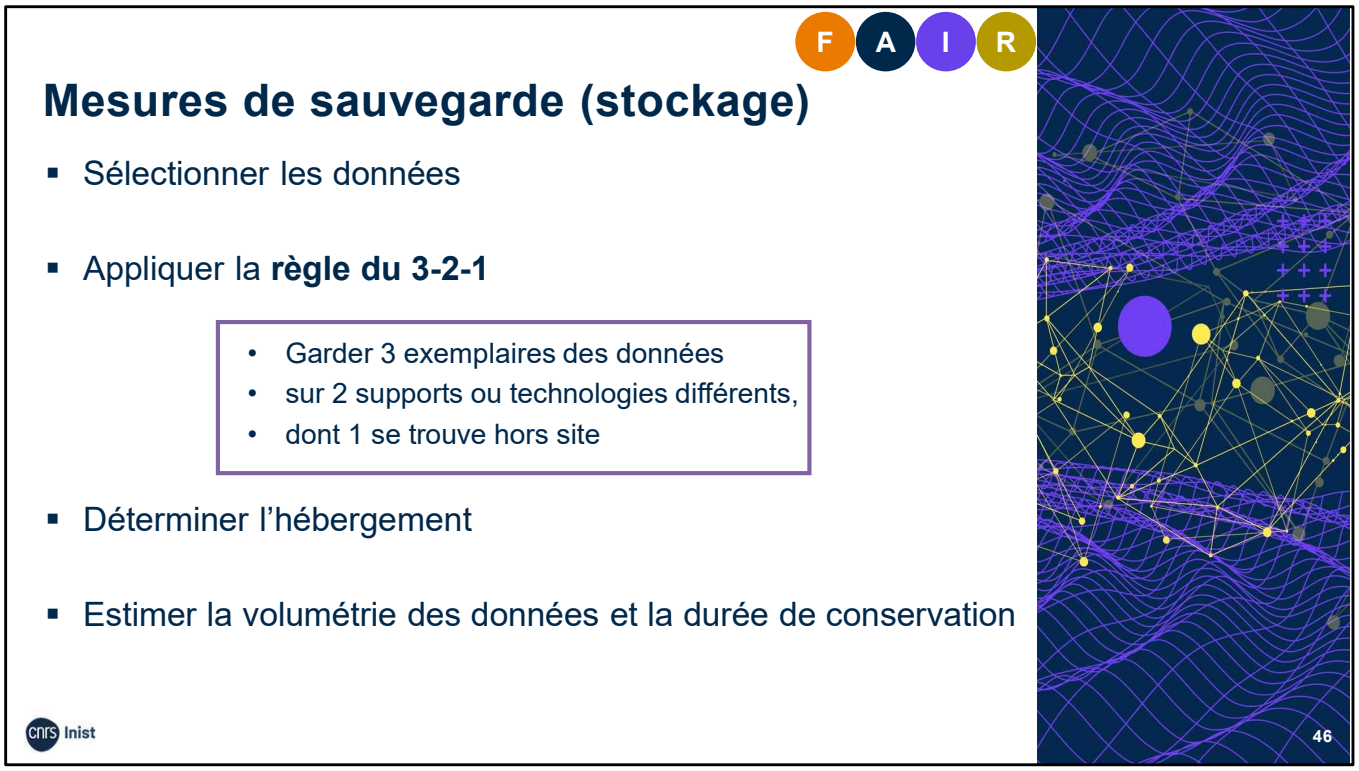
cnrs Inist

45

Le **stockage** et la **sauvegarde** sécurisés des données se font durant le projet. L'objectif est de garantir la sécurité des données et d'en faciliter l'accès pour l'ensemble des collaborateurs du projet.

Ressource *:*

- *DoRANum. Stockage, partage et archivage : quelles différences ? 1er juillet 2021. https://doranum.fr/stockage-archivage/stockage-partage-archivage-quelles-differences_10_13143_5dax-qp58/*
- *DoRANum. Stocker ses données de façon sécurisée. 26 juillet 2022. https://doranum.fr/stockage-archivage/stockage-donnees_10_13143_z0ge-nc29/*



Organiser et planifier la sauvegarde des données :

- À chaque point d'étape du projet, sélectionner les données à sauvegarder, à supprimer

Dans l'idéal, dupliquer et stocker les données à différents endroits sur différents supports en veillant à bien gérer les versions :

- Les différents états des données sont conservés en corrélation avec les différentes étapes de traitement
- Permet de revenir à une version antérieure si besoin.

Le mieux est de suivre la règle du 3-2-1, c'est-à-dire : garder 3 exemplaires des données, sur 2 supports ou technologies différents dont 1 se trouve hors site.

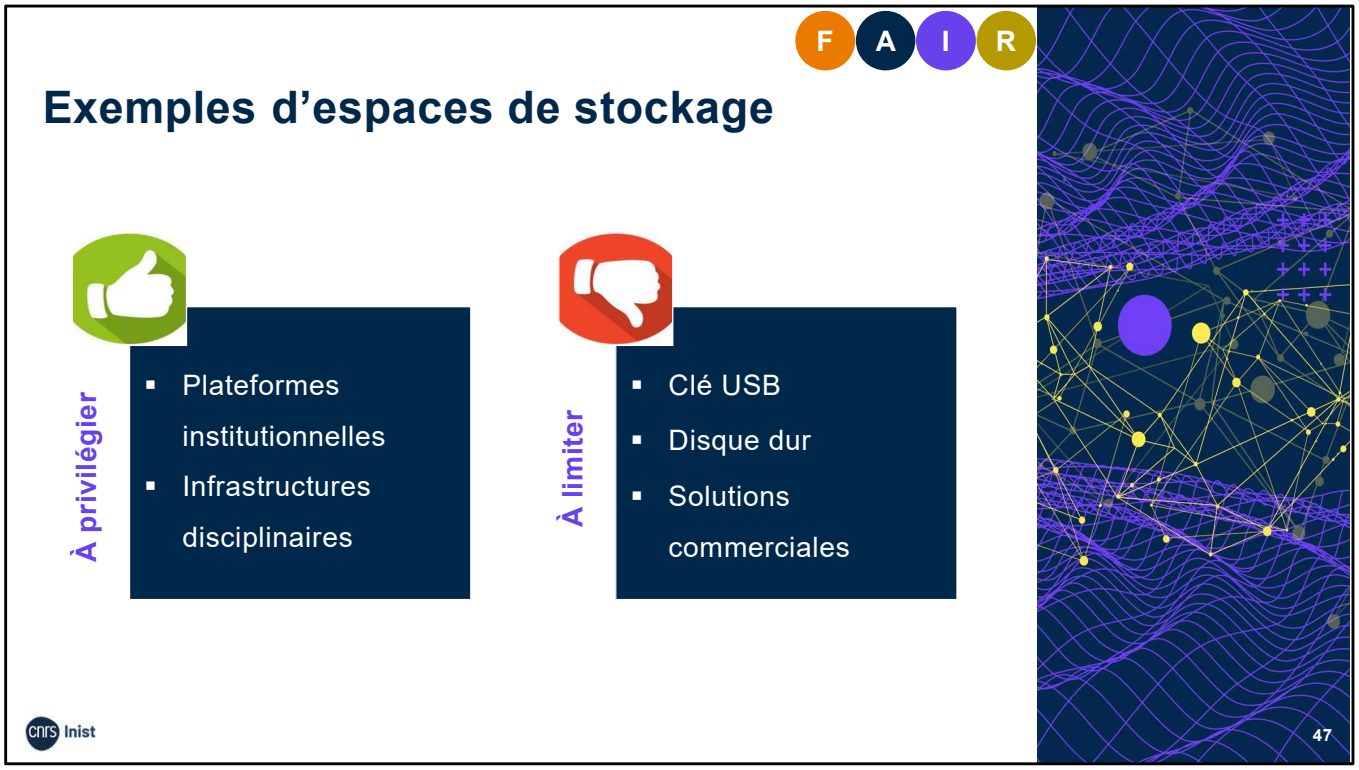
Déterminer l'hébergement : serveurs locaux (machines virtuelles), cloud institutionnel avec accès sécurisé…

Estimer la volumétrie des données et la durée de conservation

Contacter les services de votre institution pour l'hébergement, la volumétrie, la durée de conservation.

Source *:*

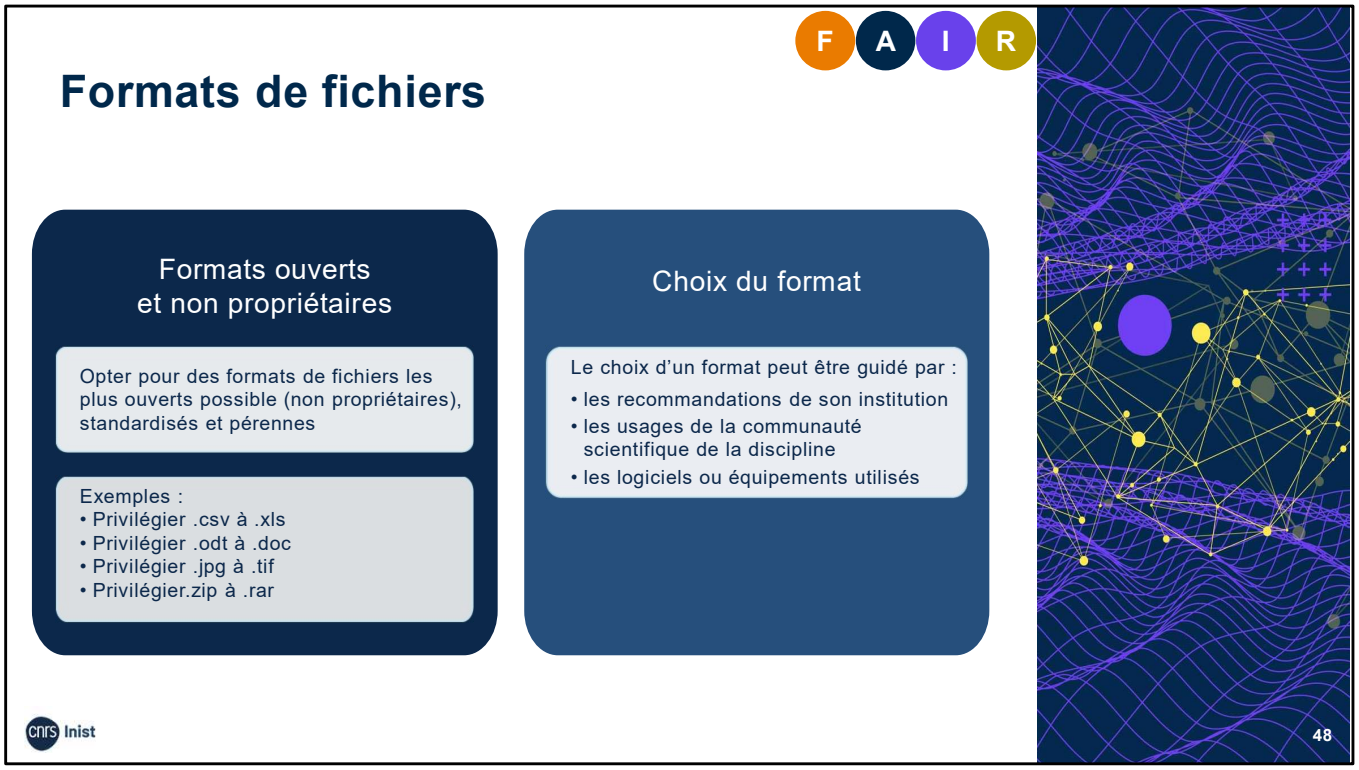
- *Université de Lille Sciences et Technologies. Sauvegardes ? https://indico.math.cnrs.fr/event/2317/contributions/1314/attachments/692/773/problematique_des_sauvegardes-journees_mathrice-20170928.pdf*



Exemples **Plateformes institutionnelles ou infrastructures disciplinaires** :

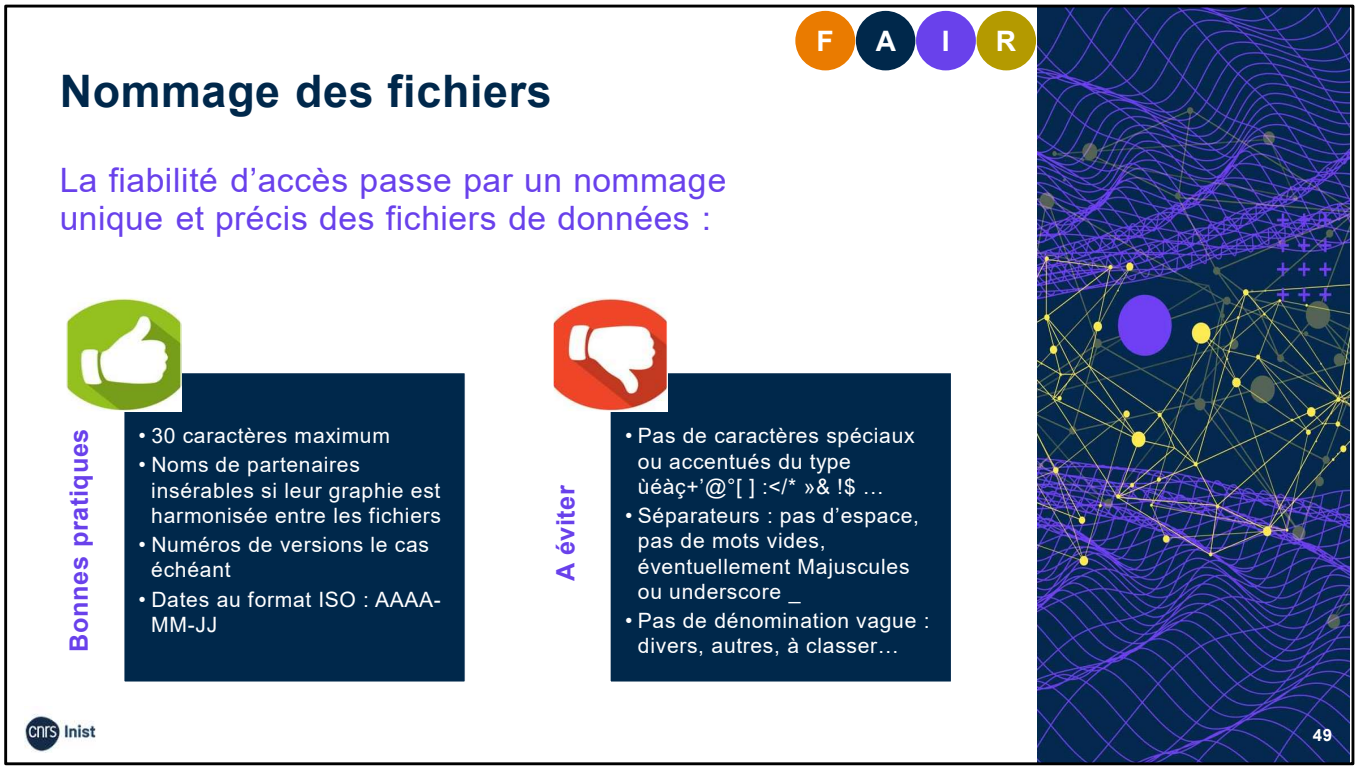
- Huma-Num Box : https://documentation.huma-num.fr/humanum-box/
- Stockage données recherche (PETA) à l'Université de Lorraine : https://numerique.univ-lorraine.fr/catalogue-des-services/stockage-donnees-recherche-peta
- Espaces de stockage du CC-IN2P3 : https://doc.cc.in2p3.fr/fr/Data-storage/storage-areas.html

Exemples de **solutions commerciales** : Google Drive, Dropbox...



Ressource *:*

- *DoRANum. Format ouvert ou fermé ? https://doranum.fr/stockage-archivage/quiz-format-ouvert-ou-ferme_10_13143_mcwq-qs64/*



Source *:*

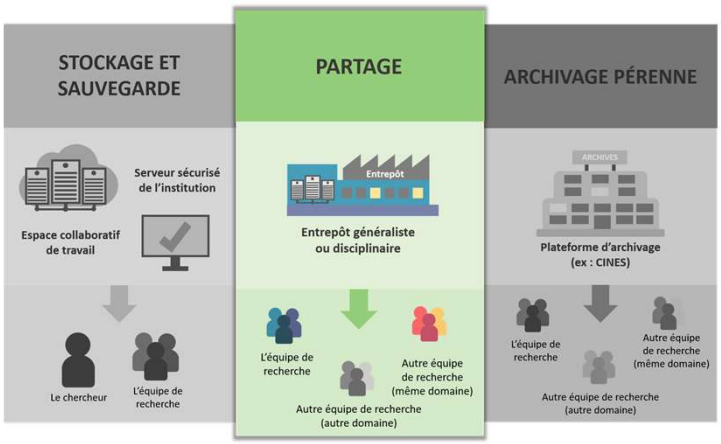
- *Arnould Pierre-Yves, Jacquemot-Perbal Marie-Christine. Guide de bonnes pratiques : Gestion et valorisation des données de recherche. 23 février 2016.* [*https://hal.science/hal-01275841/*](https://hal.science/hal-01275841/)

Étape 4 :
Partage, diffusion des données
Dépôt dans un entrepôt

Place dans le cycle de vie des données

cnrs Inist

Partage, diffusion des données



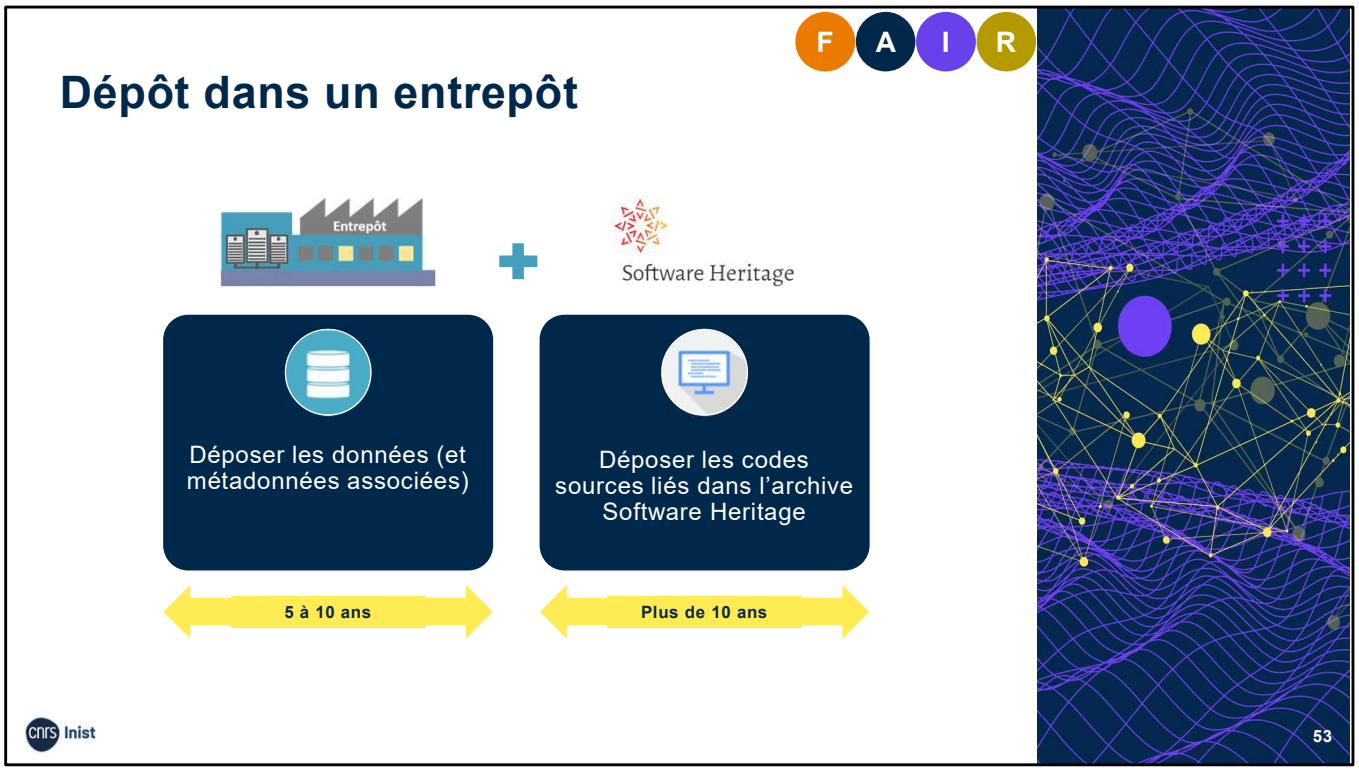
cnrs Inist

52

Le **partage** consiste à déposer les données dans un entrepôt de données afin de les rendre accessibles facilement et de permettre leur réutilisation par des chercheurs du même domaine ou d'un autre domaine, selon les principes FAIR et sur le court et le moyen terme. Le partage se fait souvent à l'issue du projet.

Ressource *:*

- *DoRANum. Stockage, partage et archivage : quelles différences ? 1er juillet 2021. https://doranum.fr/stockage-archivage/stockage-partage-archivage-quelles-differences_10_13143_5dax-qp58/*



Pour les logiciels, **déposer les codes sources dans HAL** (https://hal.archives-ouvertes.fr/), **en lien avec Software Heritage**, archive de logiciels (https://www.softwareheritage.org/?lang=fr).

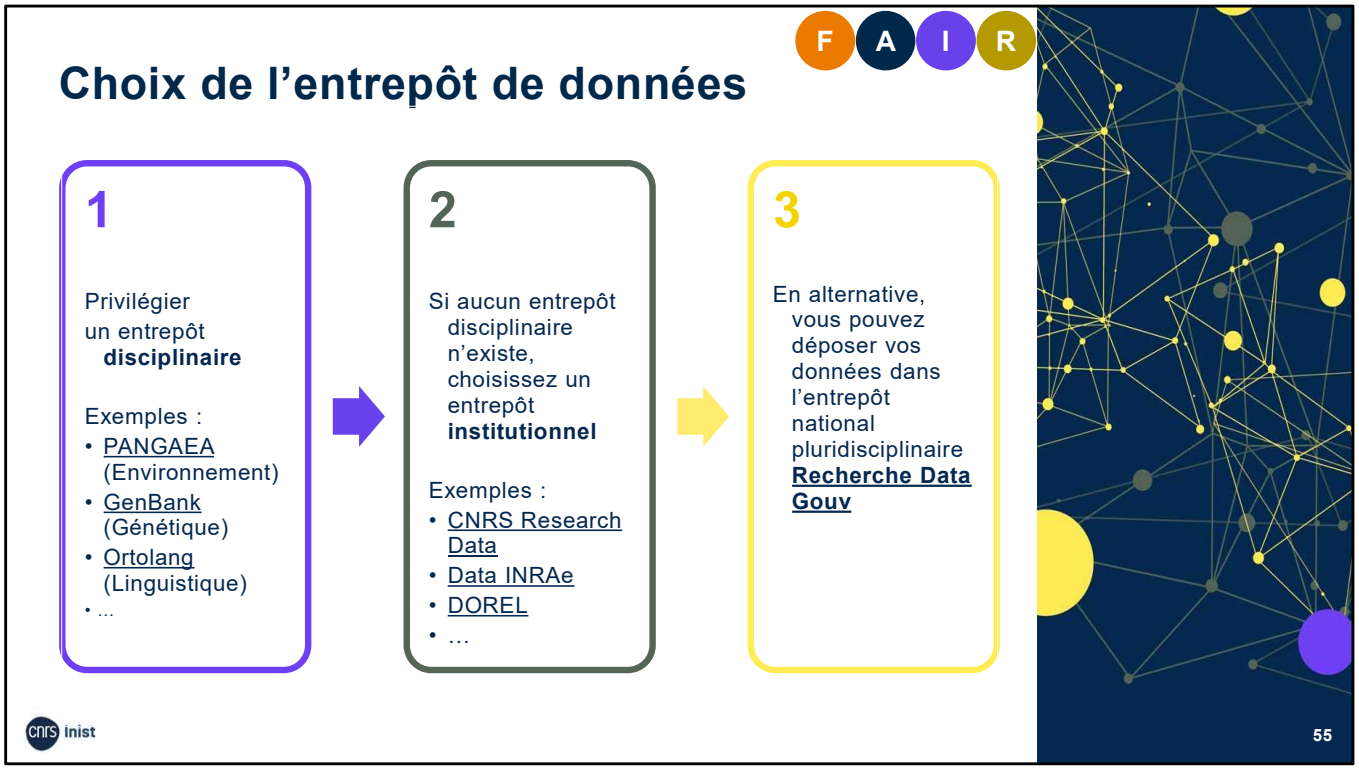
Tout ces éléments contribuent aux principes FAIR :

- Sélectionner les données à partager
- Estimer la volumétrie des données et éventuellement le budget.
- Vérifier la compatibilité et l'interopérabilité des formats de données (I/R)
- Migrer si possible vers un format ouvert (I/R)
- Préparer si nécessaire les codes sources qui permettront de lire et traiter les données (A/I/R)
- Compléter et enrichir les métadonnées (en fonction de l'entrepôt choisi) (F/A/I/R) :
    - Si ce n'est pas déjà fait, choisir un standard de métadonnées
    - S'il n'en existe pas d'adapté, créer un schéma de métadonnées
    - Compléter les champs pour chaque jeu de données, suivant le standard adopté.

Ressource *:*

DoRANum. Vérifier ses données de recherche. 6 septembre 2022.

https://doranum.fr/depot-entrepots/verifier-donnees-recherche_10_13143_5rs6-4r06/



Exemples d'entrepôts disciplinaires :

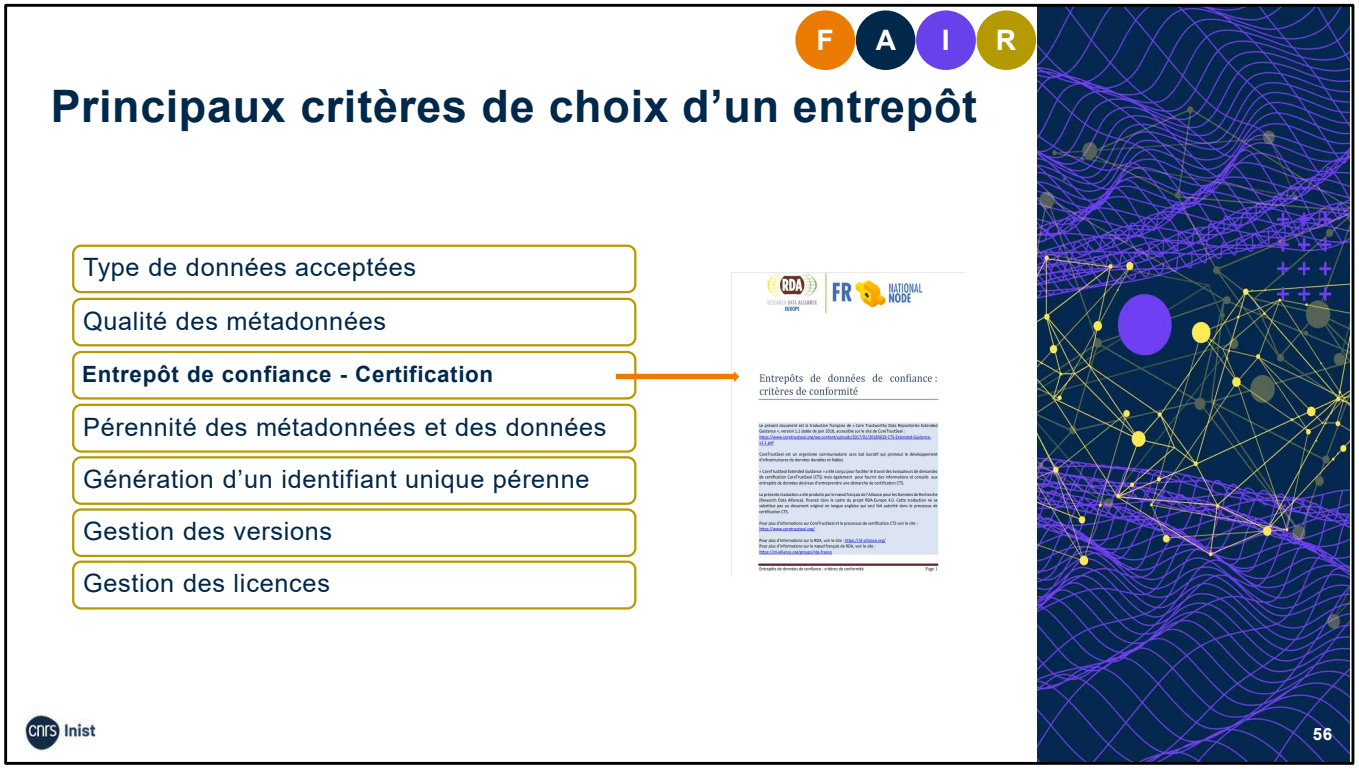
- **PANGAEA** (Environnement) : https://www.pangaea.de/
- **GenBank** (Génétique) : https://www.ncbi.nlm.nih.gov/genbank/
- **Ortolang** (Linguistique) : https://www.ortolang.fr/

Exemples d'entrepôts institutionnels :

- **CNRS Research Data** : https://entrepot.recherche.data.gouv.fr/dataverse/cnrs
- **Data INRAE** : https://entrepot.recherche.data.gouv.fr/dataverse/inrae
- **DOREL** : https://dorel.univ-lorraine.fr/

Entrepôt pluridisciplinaire :

Recherche Data Gouv : https://entrepot.recherche.data.gouv.fr/



Un entrepôt digne de confiance devrait permettre et assurer :

- le repérage et l'identification des données
- la recherche, la citation et le téléchargement des données
- la gestion des versions des jeux de données
- le référencement d'informations pertinentes complémentaires, telles que d'autres jeux de données et publications
- l'accès ouvert à des informations mises à jour, y compris sur des données non publiées, protégées, rétractées ou supprimées : métadonnées archivées sur le long terme, même si les données correspondantes ne sont plus disponibles
- la récupération des métadonnées par les machines
- l'accès aux données dans des conditions bien définies (licences)
- l'authenticité et l'intégrité des données
- la confidentialité et le respect des droits des personnes et créateurs de données
- la pérennité des données et métadonnées

Sources *:*

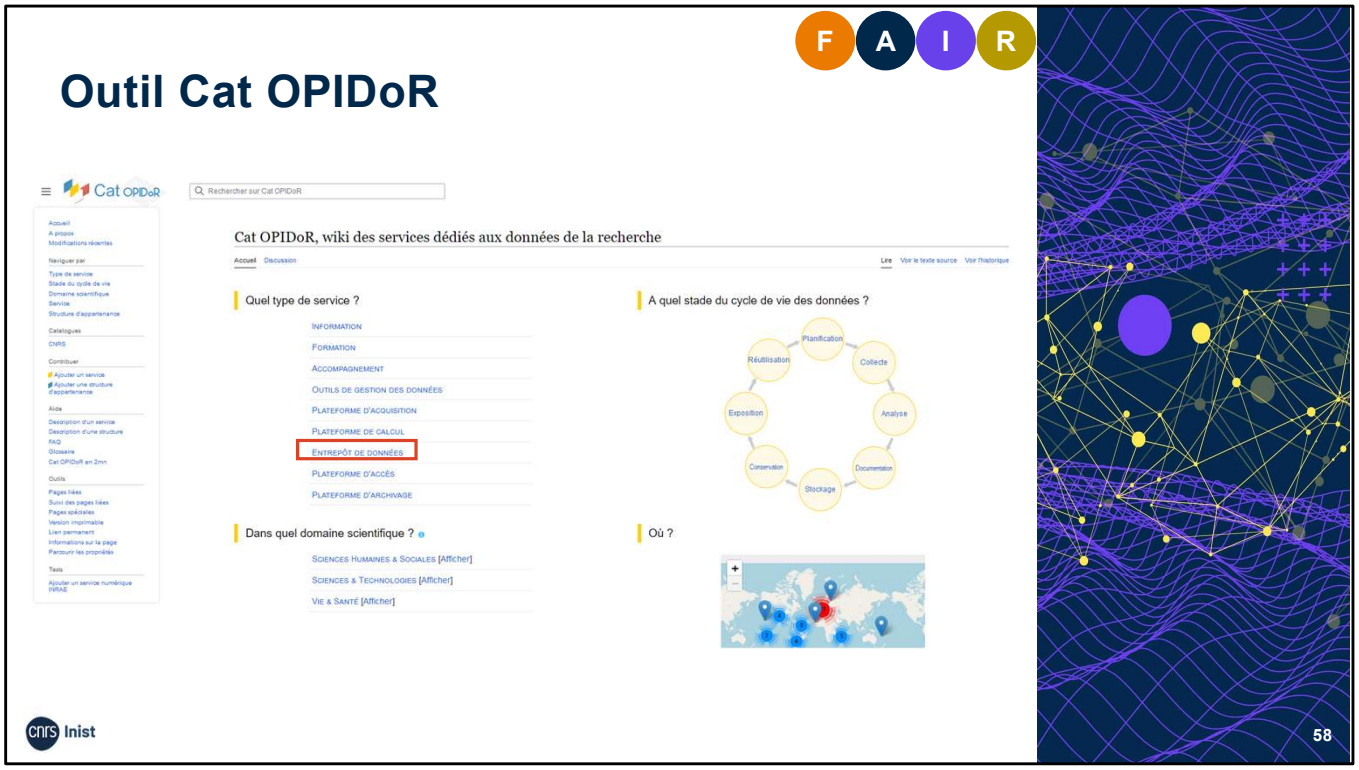
RDA. Entrepôts de données de confiance : critères de conformité. https://www.rd-alliance.org/system/files/documents/CoretrustsealFR.pdf

Science Europe. Practical guide to the international alignment of research data management – Extended Edition. 27 janvier 2021. https://doi.org/10.5281/zenodo.4915862

Ressource *:*

DoRANum. Les critères pour choisir un entrepôt de données. 30 août 2023. https://doranum.fr/depot-entrepots/criteres-pour-choisir-entrepot_de_donnees_10_13143_zqpb-9449/

- **CatOPIDoR** : Catalogue pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche. Wiki des services dédiés aux données de recherche. Il recense des entrepôts français. https://cat.opidor.fr/index.php/Entrep%C3%B4t_de_donn%C3%A9es
- **Re3data** : annuaire recensant des entrepôts au niveau international et permettant de filtrer selon plusieurs critères (discipline, attribution d'un identifiant pérenne, types de données ou formats acceptés, préservation à long terme des données, choix de la licence, certification, etc.). https://www.re3data.org/
- **OAD** : Open Access Directory – Data repositories : liste d'entrepôts et de bases de données pour les données ouvertes. https://oad.simmons.edu/oadwiki/Data_repositories
- **OpenDOAR** : annuaire mondial d'entrepôts ou d'archives ouvertes en libre accès. La recherche et la navigation peuvent s'effectuer sur différents critères tels que le nom, la région du monde ou le pays et le logiciel. https://v2.sherpa.ac.uk/opendoar/
- **FAIRsharing** : ressource permettant d'identifier et citer des standards, des bases de données ou des entrepôts existant pour leurs données et leur discipline. https://fairsharing.org/



Inist-CNRS. Cat OPIDoR. https://cat.opidor.fr/

Recense et décrit les services français dédiés aux données scientifiques.

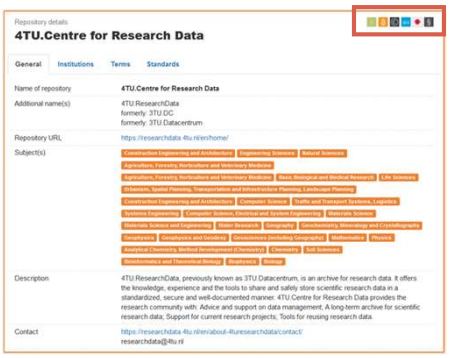
Proposé sous forme d'un wiki, cet outil collaboratif ouvert à tous permet de repérer et ajouter des services utiles dans le cadre d'un projet de recherche

Cat OPIDoR présente par domaine scientifique :

- des sites d'information,
- de formation,
- des outils de gestion,
- des plateformes,
- Des entrepôts de données

pour accompagner les chercheurs sur l'ensemble des étapes clés de la gestion, collecte, stockage, conservation et ouverture des données.

Recherche d'entrepôts



Dans **re3data**, exemple de recherche à partir des critères suivants :

- Informations complémentaires fournies
- Libre accès aux données
- Conditions d'utilisation et licence fournis
- Génération d'un DOI
- Certification
- Politique de l'entrepôt fournie

cnrs Inist

59

Possibilité de consulter l'annuaire par sujet, type de contenu, pays. La recherche graphique par sujet est matérialisée par une succession de cercles concentriques, du plus général vers le plus spécifique.

Plusieurs entrepôts répondent à ces 6 critères :

- 4TU.Centre for Research Data
- CLARIN repository at the University of Tübingen
- NASA Socioeconomic Data and Applications Center
- PANGAEA

Identifiants pérennes

Une équipe dédiée de l'Inist-CNRS se tient à disposition pour conseiller dans l'attribution de DOI aux données de recherche et fournit :

- un accès à un espace test pour enregistrer temporairement des DOI et vérifier la compatibilité de ce service avec ses propres workflows
- un préfixe unique de DOI
- un accès à la plateforme Metadata Store de DataCite (MDS) pour commencer à créer les DOI
- une assistance à la création et à la conversion de métadonnées…
- un accompagnement dans l'utilisation des différents services proposés par DataCite.

Ressources *:*

- *DoRANum. Zoom sur le DOI. 21 mai 2021. https://doranum.fr/identifiants-perennes-pid/zoom-doi_10_13143_j5xt-6j41/*
- *Inist-CNRS. PID OPIDoR. https://opidor.fr/identifier/*

Créé par Software Heritage, SWHID (Software Hash Identifier, anciennement Software Heritage IDentifier) est un système d'identifiant pérenne sous forme de code alphanumérique qui identifie de manière unique les codes sources des logiciels.

Ressources *:*

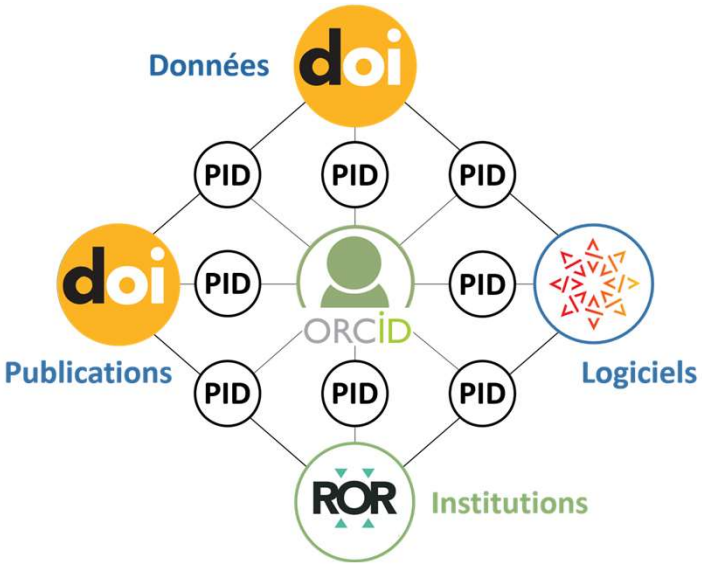
- *DoRANum. Zoom sur SWHID. 2 juillet 2021. https://doranum.fr/identifiants-perennes-pid/zoom-swhid_10_13143_3qqg-yx41/*
- *Software Heritage. https://www.softwareheritage.org/?lang=fr*
- *HAL Documentation. Déposer le code source d'un logiciel. https://doc.archives-ouvertes.fr/deposer/deposer-le-code-source/*

Ressources *:*

DoRANum. Zoom sur ORCID. 5 décembre 2022. https://doranum.fr/identifiants-perennes-pid/zoom-orcid_10_13143_c6rx-9w77/

DoRANum. Zoom sur ROR. 8 juin 2022. https://doranum.fr/identifiants-perennes-pid/zoom-sur-ror_10_13143_h5zy-bn73/

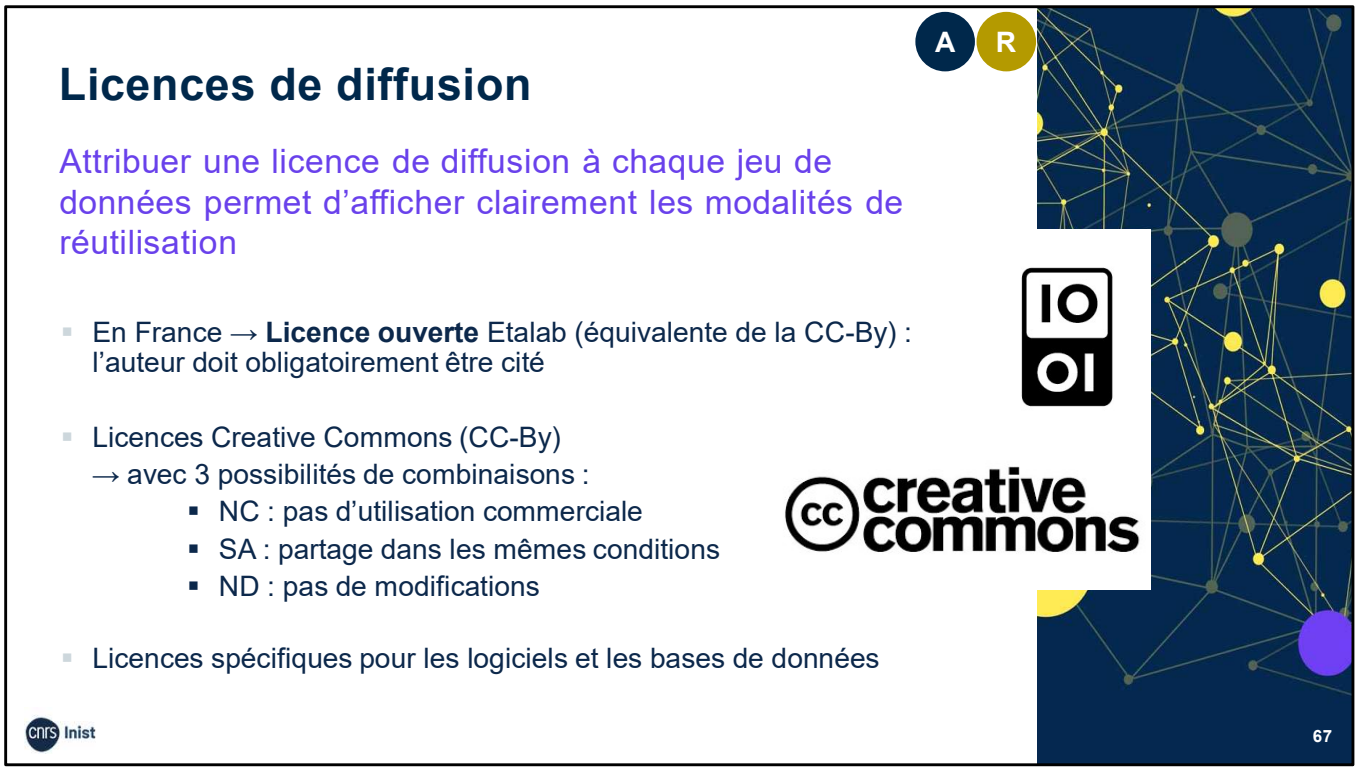
Identifiants pérennes – Pour résumer



Modalités de protection

Accès :

- Accès limité dans le temps par un embargo
  - Pour disposer du temps nécessaire au dépôt de brevets…
  - En fonction de la discipline
  - Peut être déterminé par les éditeurs
- Accès limité à certaines personnes
  - Ex : limitation aux membres du consortium ou à une communauté scientifique



En France, la licence la plus adaptée pour les données publiques est la **Licence ouverte** Etalab (équivalente de la CC-By) : l'auteur doit obligatoirement être cité Sinon, il est possible d'utiliser les licences Creative Commons (CC-By), avec 3 possibilités de combinaisons :

- NC : pas d'utilisation commerciale
- SA : partage dans les mêmes conditions
- ND : pas de modifications

Il existe des licences spécifiques pour les logiciels et les bases de données

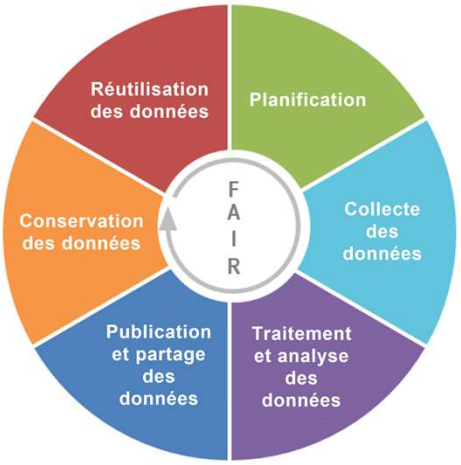
Ressources *:*

- *DoRANum. Guide des licences ouvertes.* https://doranum.fr/aspects-juridiques-ethiques/guide-des-licences-ouvertes_10_13143_tv6f-sv31/
- *DoRANum. Les licences de réutilisation dans le cadre de l'open data et de la loi pour une République numérique.* https://doranum.fr/aspects-juridiques-ethiques/les-licences-de-reutilisation-dans-le-cadre-de-lopen-data-2_10_13143_ssh2-zd93/
- *Data.gouv.fr. Licences de réutilisation.* https://www.data.gouv.fr/fr/licences
- *Licentia by Inria. License your Data.* http://licentia.inria.fr/.
- *License Selector.* http://ufal.github.io/public-license-selector/
- *Choose an open source license :* https://choosealicense.com/

Étape 5 :
Archivage pérenne

Place dans le cycle de vie des données

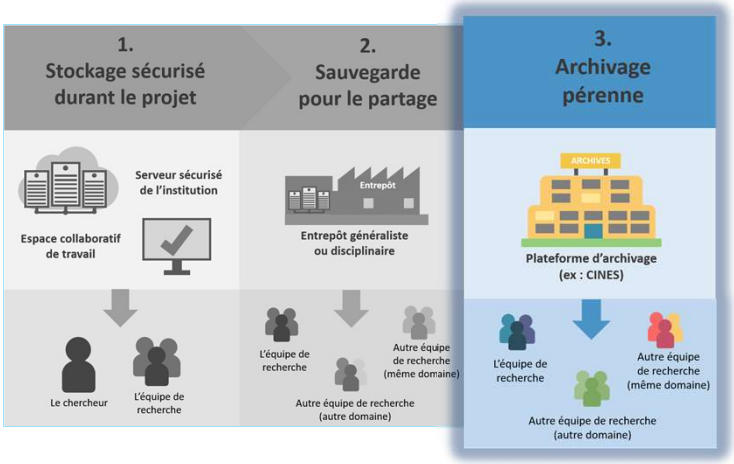
Étape 3
Archivage pérenne



cnrs Inist

69

Archivage pérenne



cnrs Inist

70

L'**archivage pérenne** a pour objectif de conserver les données, d'en garantir l'accès et d'en préserver l'intelligibilité sur le long terme, c'est-à-dire plus de 30 ans. En réalité, l'archivage pérenne concerne peu de données. Seulement celles qui présentent une grande valeur scientifique reconnue par la communauté dont elles proviennent. Soit parce qu'elles sont très couteuses, soit parce qu'elles sont uniques, non reproductibles.

***Ressource** :*

- *DoRANum. Stockage, partage et archivage : quelles différences ? 1er juillet 2021. https://doranum.fr/stockage-archivage/stockage-partage-archivage-quelles-differences_10_13143_5dax-qp58/*

Définition et objectifs

L'archivage pérenne a pour objectif de :

- Conserver uniquement les **données à forte valeur scientifique**, **impossibles ou coûteuses** à (re)produire
- Conserver les données dans leur **aspect physique** comme dans leur **aspect intellectuel**
- En préserver l'**intelligibilité** sur le très long terme
- De manière à ce qu'elles soient en permanence **accessibles et compréhensibles**

Archivage pérenne coûteux et soumis à la décision de l'institution

cnrs Inist

71

Valeur scientifique des données :

Sont-elles uniques, non reproductibles (ou à des coûts trop élevés) ?

Ont-elles une valeur historique (représentent-elles un point de repère dans les découvertes scientifiques ?)

…

Attention, l'archivage pérenne est une opération très couteuse. Elle est soumise à une décision de votre institution/laboratoire. Dans le cadre d'un projet financé, ce coût peut être éligible.

Source *:*

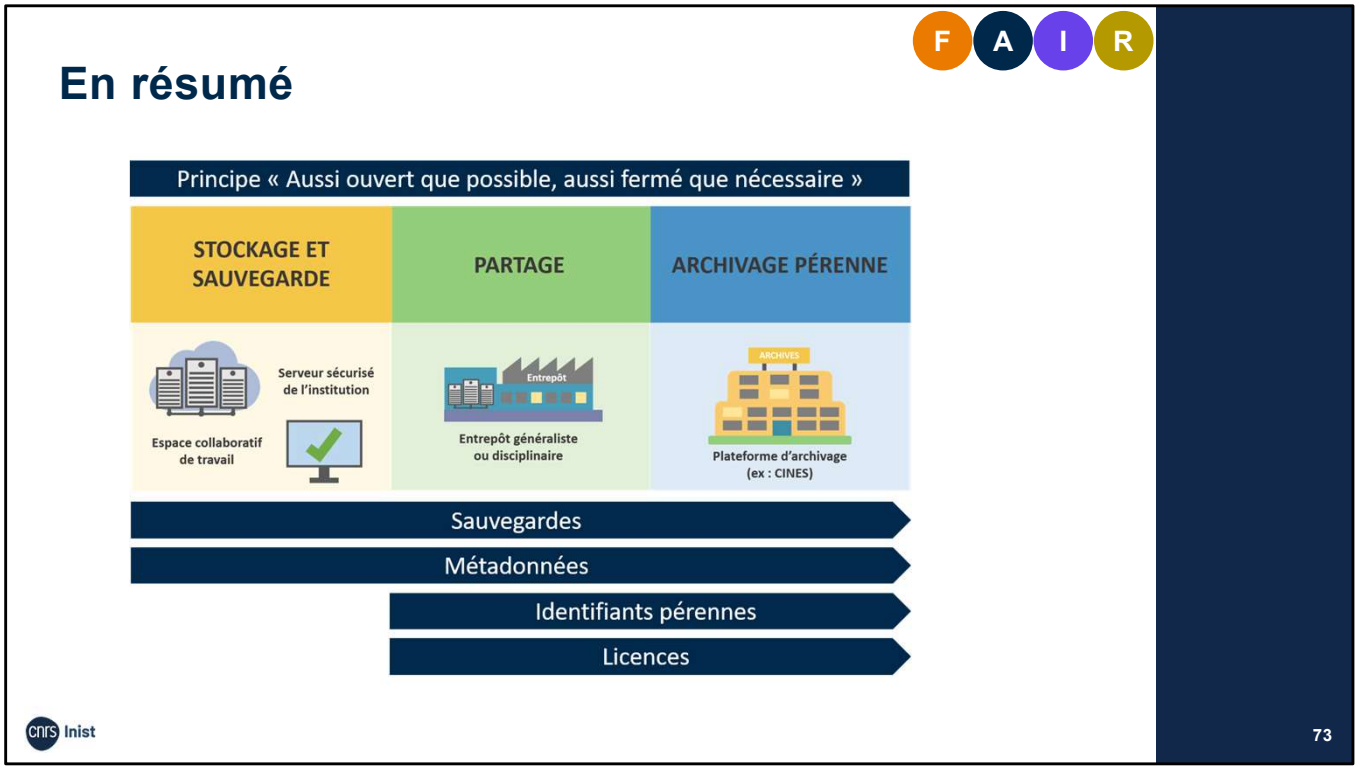
NERC Data Value Checklist. [*https://www.ukri.org/publications/nerc-data-value-checklist/*](https://www.ukri.org/publications/nerc-data-value-checklist/)

Outil FACILE : pour tester la validité et/ou l'éligibilité de ses formats de fichiers selon les règles du CINES.

Ressources *:*

CINES. Nos solutions d'archivage. https://www.cines.fr/archivage/nos-solutions-darchivage/

Programme VITAM. https://www.programmevitam.fr/



- Le **stockage** et la **sauvegarde** sécurisés des données se font durant le projet. L'objectif est de garantir la sécurité des données et d'en **faciliter l'accès** pour l'ensemble des collaborateurs du projet.
- Le **partage** consiste à déposer les données dans un entrepôt de données afin de les rendre **accessibles facilement** et de permettre leur **réutilisation** par des chercheurs du même domaine ou d'un autre domaine, selon les principes FAIR et sur le court et le moyen terme. Le partage se fait souvent à l'issue du projet.
- L'**archivage pérenne** a pour objectif de conserver les données, d'en garantir l'**accès** et d'en préserver l'intelligibilité sur le long terme.

Les **bonnes pratiques** de **sauvegarde**, d'ajout de **métadonnées**, attribution d'**identifiants pérennes** et de **licences** participent tous au respect des **principes FAIR**.

Ressource *:*

DoRANum. Stockage, partage et archivage : quelles différences ? 1er juillet 2021. *https://doranum.fr/stockage-archivage/stockage-partage-archivage-quelles-differences_10_13143_5dax-qp58/*

Étape 6 :
Réutilisation et valorisation des données

Place dans le cycle de vie des données

cnrs Inist

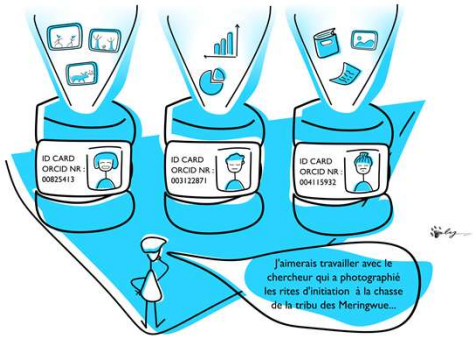
- Plus de visibilité et meilleure valorisation de ses travaux : si vos publications sont reliées à vos données et que vous donnez accès à celles-ci, cela renforce la confiance en vos publications et leur apporte davantage de crédit. Cela permet aussi d'éviter de dupliquer des données ou des résultats déjà publiés.
- Des données mal documentées ou mal conservées sont des données perdues ! Les principes FAIR facilitent la bonne gestion des données, pendant et après le travail de recherche. L'objectif d'un travail de recherche est d'être lu, vérifié, cité, réutilisé. Ceci est impossible si personne n'a accès aux informations, si le format choisi pour le stockage des données n'est plus lisible, si l'information est perdue, ou si l'adresse url d'accès n'est plus disponible !
- Science citoyenne : vos publications, vos données gagnent à être mise à la disposition de tous, mais pour ce faire, il faut qu'elles soient compréhensibles et manipulables, et accompagnées de métadonnées.

***Source** :*

Bibliothèque de l'Observatoire de Paris. GT Science ouverte. Des principes ? Pour quoi FAIR ? 13 avril 2021. https://www.observatoiredeparis.psl.eu/IMG/pdf/pour-quoi-fair-3.pdf

Du côté du réutilisateur

- Trouver des jeux de données déjà existants (dans les entrepôts et data papers)
- Découverte de potentiels partenaires
- Reproductibilité et réplicabilité des résultats



cnrs Inist

77

La **reproductibilité** est la capacité, par une équipe différente, de reproduire une expérience, sans se fier au dispositif expérimental et aux codes logiciels développés par l'équipe d'origine.

La **réplicabilité** est la capacité, par une équipe différente, de reproduire une expérience en ré-utilisant le même dispositif expérimental décrit (y compris les codes logiciels).

***Ressources** pour rechercher des jeux de données :*

- *DataCite Commons. https://commons.datacite.org/*
- *OpenAIRE EXPLORE. https://explore.openaire.eu/search/find/datasets*
- *Google Dataset Search. https://datasetsearch.research.google.com/*

Voici un exemple de valorisation des données : le data paper.

Structure d'un data paper :

Partie descriptive :

- Éléments communs aux articles classiques : titre, résumé, mots-clés…
- Éléments spécifiques aux données : types de données, formats, processus et méthodes de production, métadonnées, réutilisation…

Accès aux données : déposées dans un entrepôt. L'identifiant des données (exemple DOI) permet d'établir le lien du data paper vers les données.

Ressources *:*

- *Exemples de data paper :*

Richardson Andrew D., Hufkens Koen, Milliman Tom et al. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery.

https://www.nature.com/articles/sdata201828

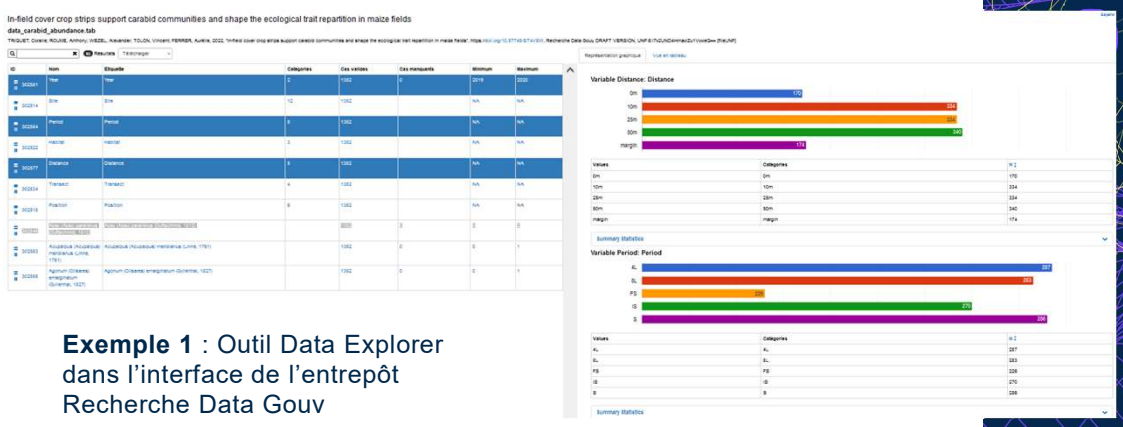
Richardson Andrew D., Hufkens Koen, Milliman Tom et al. Tracking vegetation phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset.

https://www.nature.com/articles/s41597-019-0229-9

- *Exemple de data journal : Journal of Physical and Chemical Reference Data.* *https://pubs.aip.org/aip/jpr*
- *Exemples de revues publiant des data papers : Cirad. CoopIST. Publicer un Data paper.* *https://coop-ist.cirad.fr/gerer-des-donnees/publier-un-data-paper/4-choisir-la-revue*
- *DoRANum. Data papers et data journals.* *https://doranum.fr/data-paper-data-journal/*

Explorer par la visualisation des données

En complément, penser à exposer ses données **sous forme visuelle** (cartographies, graphiques, etc.) via une plateforme



Exemple 1 : Outil Data Explorer dans l'interface de l'entrepôt Recherche Data Gouv

cnrs Inist

79

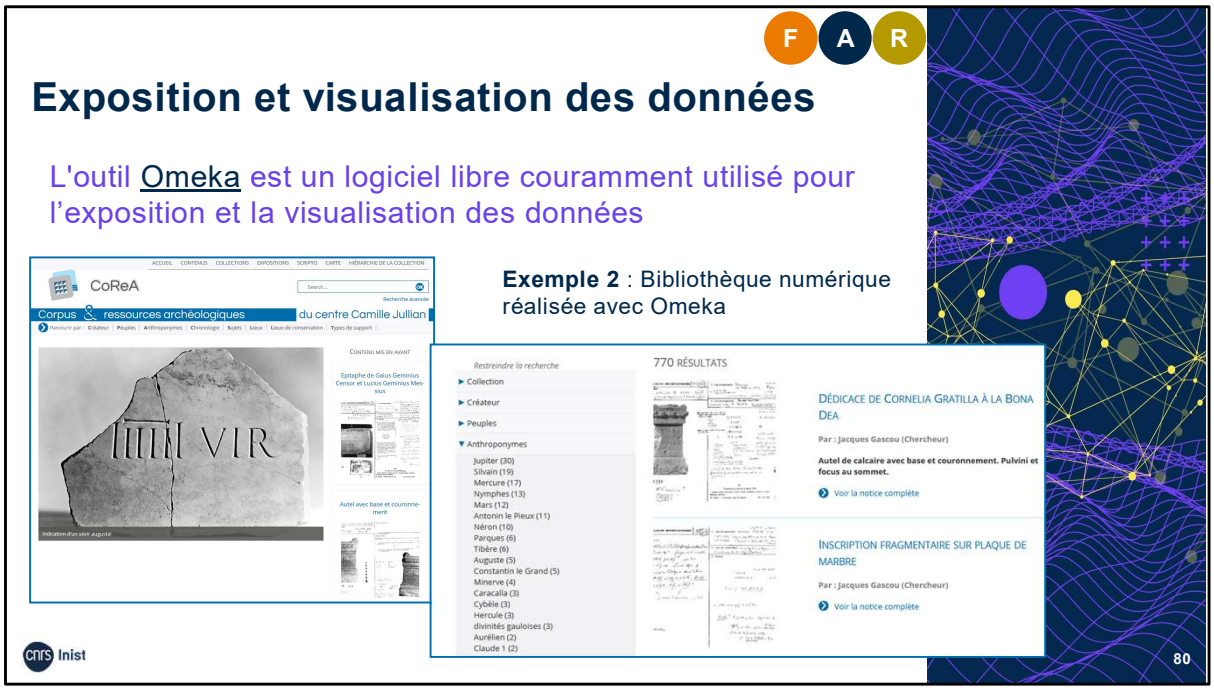
La visualisation devient une nécessité en présence de grandes quantités de données complexes et difficilement appréhendables.

Dans l'interface de l'**entrepôt Recherche Data Gouv**, l'outil Data Explorer permet d'explorer les données tabulées. Il propose une représentation graphique ou en tableau pour des analyses croisées des données du fichier ou d'un sous-ensemble de ces données, obtenus en sélectionnant les variables à prendre en compte. Il permet également de télécharger les données ou le sous-ensemble étudié.

Suivant le type de données, l'outil Data Explorer est visible dans « Modalités d'accès au fichier ». Exemple : https://entrepot.recherche.data.gouv.fr/file.xhtml?persistentId=doi:10.57745/DRUUWF&version=1.0

Ressource *:*

DoRANum. La visualisation des données : quelques exemples. 29 septembre 2023. https://doranum.fr/acces-visualisation/la-visualisation-des-donnees-quelques-exemples_10_13143_csje-a859/



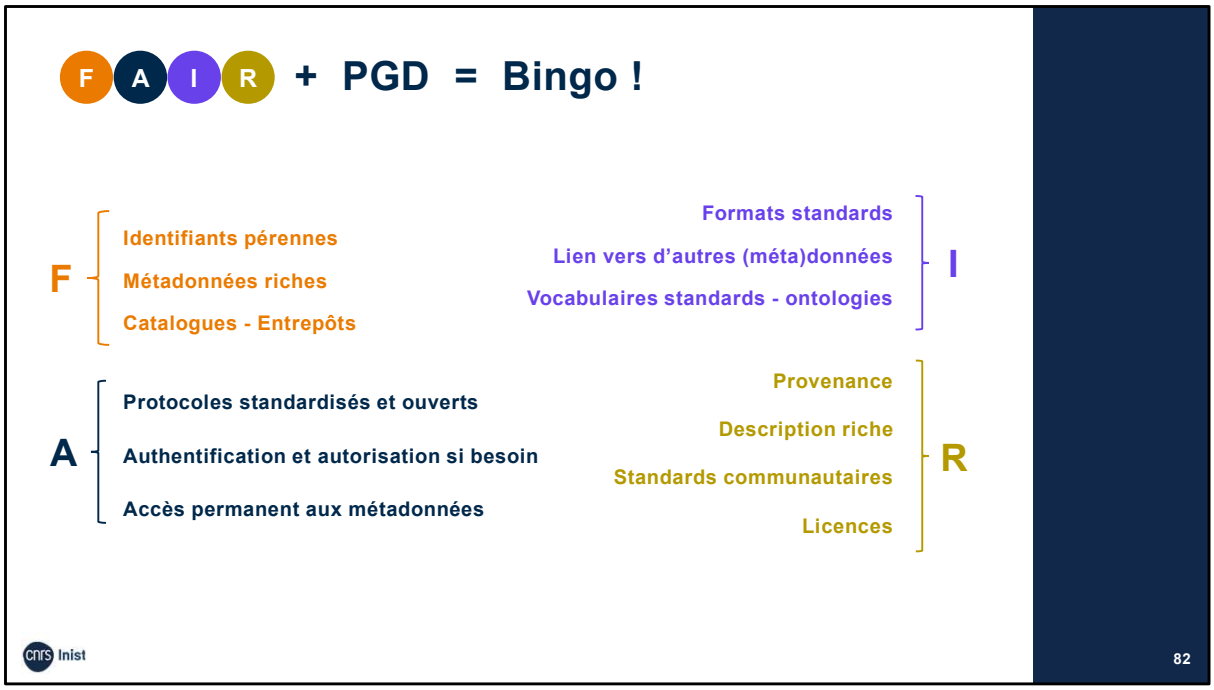
Exemple : Bibliothèque numérique CoReA (Corpus et Ressources Archéologiques) pour la documentation archéologique du Centre Camille Julian, réalisée avec Omeka. Permet de naviguer dans des corpus et ressources en archéologie. http://ccj-corea.cnrs.fr/

Omeka est un outil qui s'inscrit pleinement dans la science ouverte et qui permet la création de bases de données au plus près des principes FAIR.
À partir de données de recherche brutes, l'outil permet de créer des bases de données éditorialisées, autrement dit structurées, accessibles, et visibles sur le web.
L'outil offre une grande modularité des fonctionnalités grâce à de nombreux plugins, et traite les divers objets multimédia (textes, images, sons, vidéos).
l'outil offre plusieurs avantages techniques :
– l'interface est simple et intuitive ;
– les métadonnées sont moissonnables, permettant notamment le référencement dans d'autres bases ;
– une base Omeka peut être connectée à d'autres services grâce à une API REST.

Ressources *:*
DoRANum. Accès et visualisation. https://doranum.fr/acces-visualisation/

04

À retenir



En résumé,

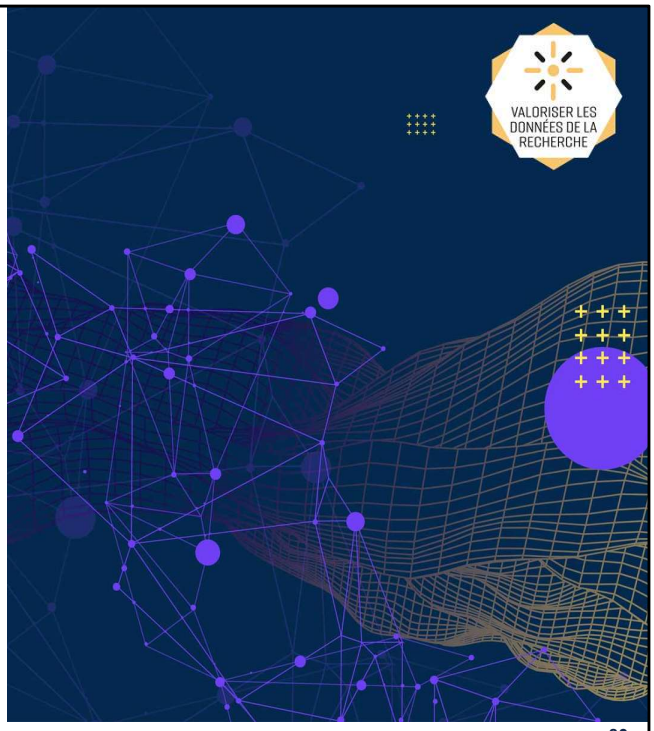
- Vous rendrez vos données **Faciles à trouver** en les déposant dans un entrepôt de confiance, vérifiant qu'un identifiant pérenne leur est attribué et en les décrivant précisément avec des métadonnées.
- Vous rendrez vos données **Accessibles** en déterminant qui peut y accéder et comment. Ne pas oublier que si vos données ne sont pas ouvertes, les métadonnées doivent l'être.
- Vous rendrez vos données **Interopérables** en utilisant des standards de votre discipline et utilisant des formats ouverts.
- Vous rendrez vos données **Réutilisables** en documentant vos données de manière à ce que les autres utilisateurs puissent les comprendre et en leur attribuant une licence.

Si vous rédigez votre **Plan de Gestion de Données** en pensant aux bonnes pratiques à mettre en œuvre en suivant les principes FAIR, vous atteindrez une gestion optimale de vos données.

Merci de votre attention

contact@doranum.fr

Pour en savoir plus, rendez-vous sur https://doranum.fr



cnrs Inist

83