

Introduction aux données de la recherche

Claire Tignolet – 03/10/25

claire.tignolet@chartes.psl.eu



Plan de la séance

- 1) De multiples enjeux
- 2) Définitions
- 3) Le cadre légal et réglementaire
- 4) La gestion des données en pratique

Objectif : Comprendre le minimum à FAIR

1. De multiples enjeux



Le potentiel économique des données

L'espace européen de la recherche (2012)



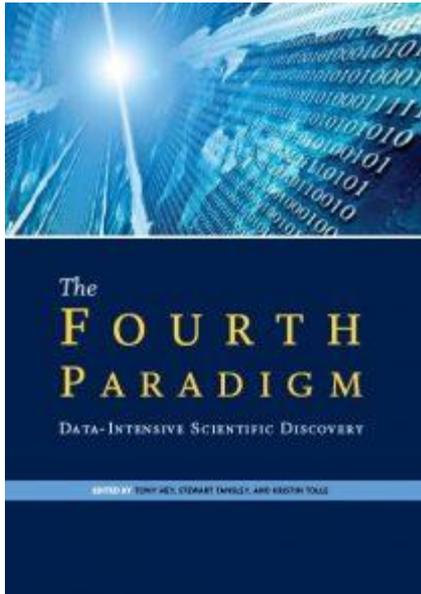
« optimiser la diffusion, l'accessibilité et le transfert des connaissances scientifiques »



Stratégie européenne pour les données (2019-24)

- A. Un cadre trans-sectoriel de **gouvernance** pour l'accès aux données et leur utilisation
- B. **Investissements** dans les données et renforcement des capacités et des **infrastructures** européennes pour l'hébergement, le traitement et l'utilisation des données, l'interopérabilité
- C. Donner à chacun les moyens d'agir, investir dans les **compétences** et dans les PME
- D. **Espaces européens communs des données** dans des secteurs stratégiques et des domaines d'intérêt public

De nouvelles perspectives de recherche



T. Hey, K. M. Tolle, and S. Tansley, [The fourth paradigm. Data-Intensive Scientific Discovery](#), Microsoft Research, 2009

CHRIS ANDERSON

SCIENCE JUN 23, 2008 12:00 PM

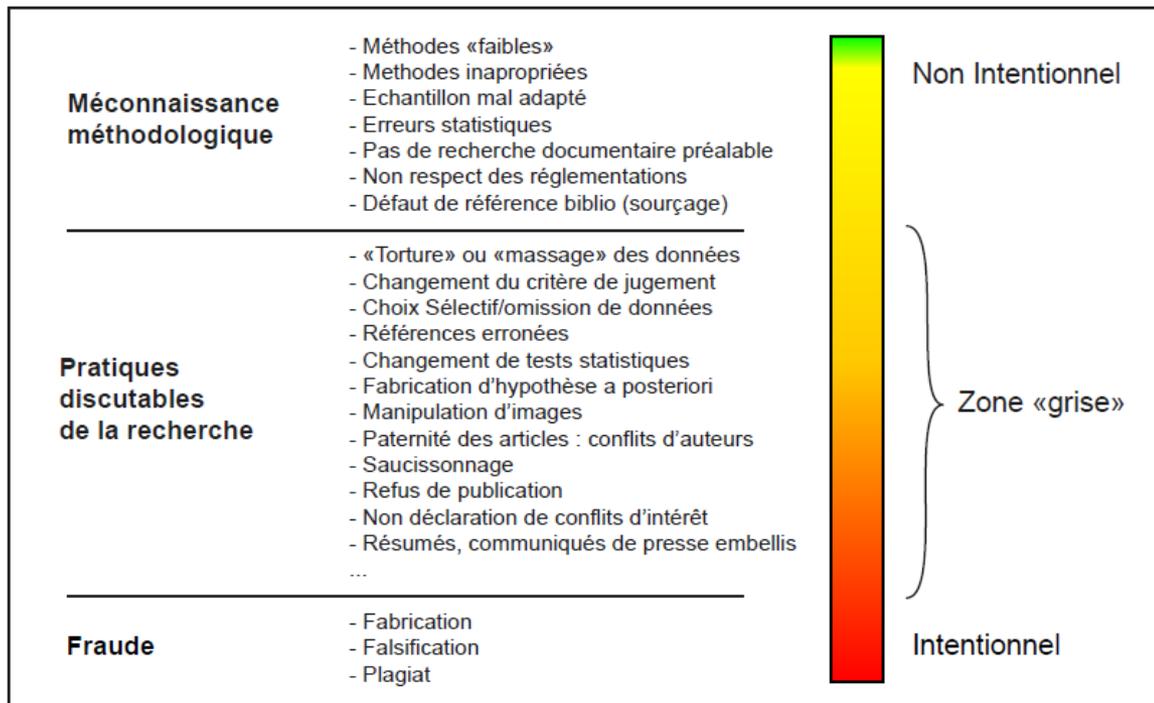
The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

Illustration: Marian Bantjes “All models are wrong, but some are useful.” So proclaimed statistician George Box 30 years ago, and he was right. But what choice did we have? Only models, from cosmological equations to theories of human behavior, seemed to be able to consistently, if imperfectly, explain the world around us. Until now. Today companies [...]

C. Anderson, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete”, [Wired](#), 2008

En français : [Le Débat](#), 2019/5

L'impératif d'intégrité scientifique



[7 cas de fraudes célèbres](#)
CNRS Le Journal (2014)

[Typologie des manquements](#), adaptée de H. Maisonneuve, AFIS
<https://www.afis.org/Biais-et-embellissements-polluent-la-science>

Données et rétractations

Retraction Watch

Tracking retractions as a window into the scientific process

Unreliable Data	The accuracy or validity of the data is questionable
Unreliable Image	The accuracy or validity of the image is questionable

Concerns/Issues About Data	Any question, controversy or dispute over the validity of the data
Concerns/Issues About Image	Any question, controversy or dispute over the validity of the image

Duplication of Data	Also known as “self-plagiarism”. Used when the all or part of the data from an item written by one or all authors of the original article, are repeated in the original article without appropriate citation.
---------------------	---

Falsification/Fabrication of Data	Intentional changes to data so that it is not representative of the actual finding
Falsification/Fabrication of Image	Intentional changes to an image so that it is not representative of the actual data

Plagiarism of Data	Used when the all or part of the data from an item not written by one or all authors of the original article, are repeated in the original article without appropriate citation.
--------------------	--

Manipulation of Images	The changing of the presentation of an image by reversal, rotation or similar action
------------------------	--

L'enjeu de reproductibilité



NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility

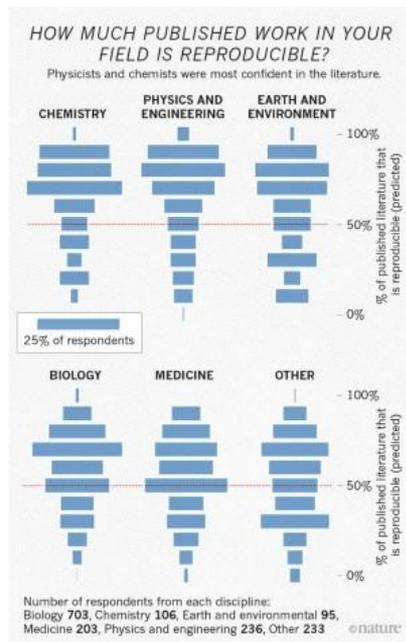
Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

[Enquête de mai 2016](#)

Plus de 70% des chercheurs ont échoué à reproduire l'expérience d'un collègue ; et plus de 50% n'ont pas réussi à reproduire leur propre expérience...



Towards Data Science DATA SCIENCE MACHINE LEARNING PROGRAMMING VISUALIZATION

Data Science's Reproducibility Crisis

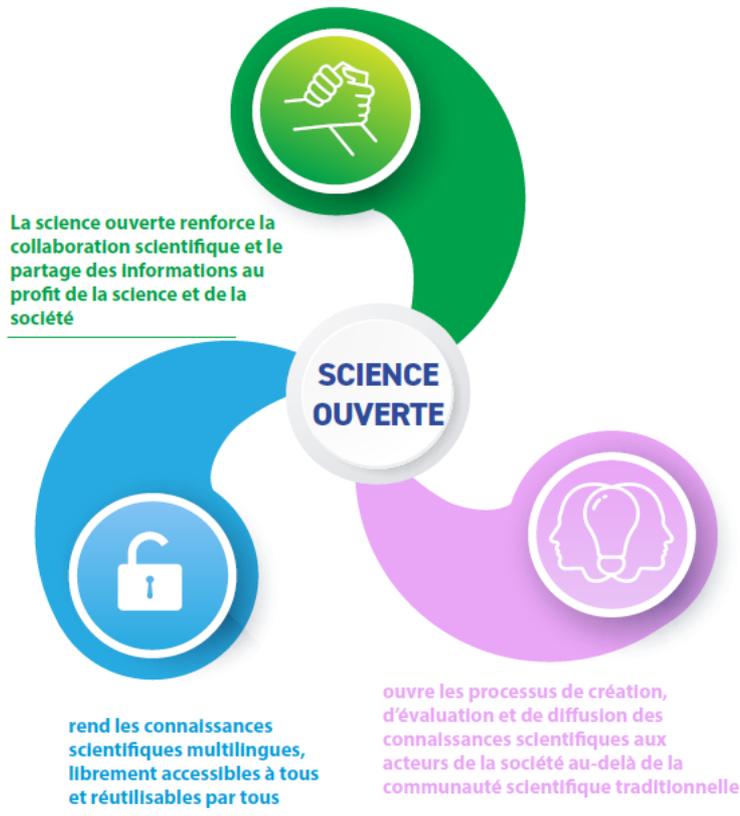


Zach Scott [Follow](#)
May 17, 2018 · 6 min read



A. Hoquet, « Reproduire un résultat scientifique, plus facile à dire qu'à faire », [The Conversation](#), mars 2020

La Science Ouverte



« La science ouverte s'entend comme un concept inclusif qui englobe différents mouvements et pratiques (...)

Elle inclut toutes les disciplines scientifiques et tous les aspects des pratiques savantes, y compris les sciences fondamentales et appliquées, les sciences naturelles et les sciences sociales et humaines, et repose sur les piliers essentiels suivants :

- les connaissances scientifiques ouvertes ;
- les infrastructures de la science ouverte ;
- la communication scientifique ;
- la participation ouverte des acteurs de la société ;
- et le dialogue ouvert avec les autres systèmes de connaissances »

Données et science ouverte



Open Science

« Les contributions au libre accès se composent de résultats originaux de recherches scientifiques, de **données brutes et de métadonnées**, de documents sources, de représentations numériques de documents picturaux et graphiques, de documents scientifiques multimédias. »

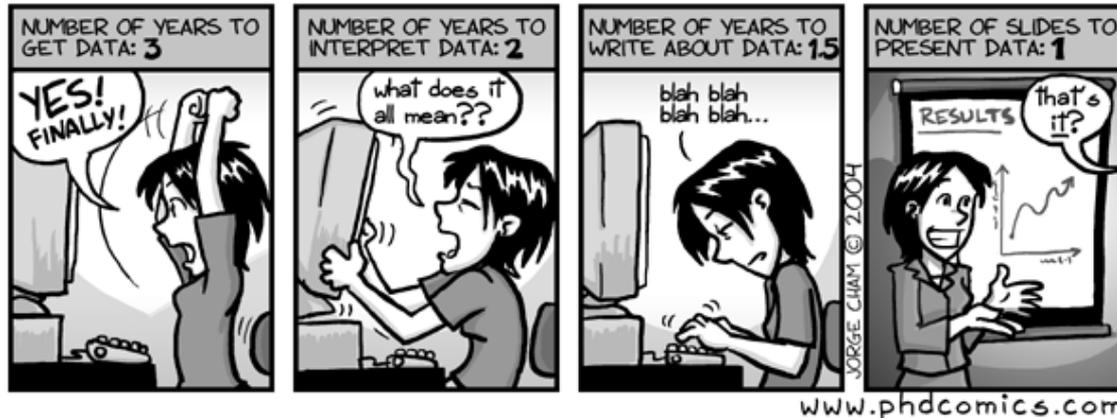
[Déclaration de Berlin sur le libre accès à la connaissance \(2003\)](#)

Des produits de recherche

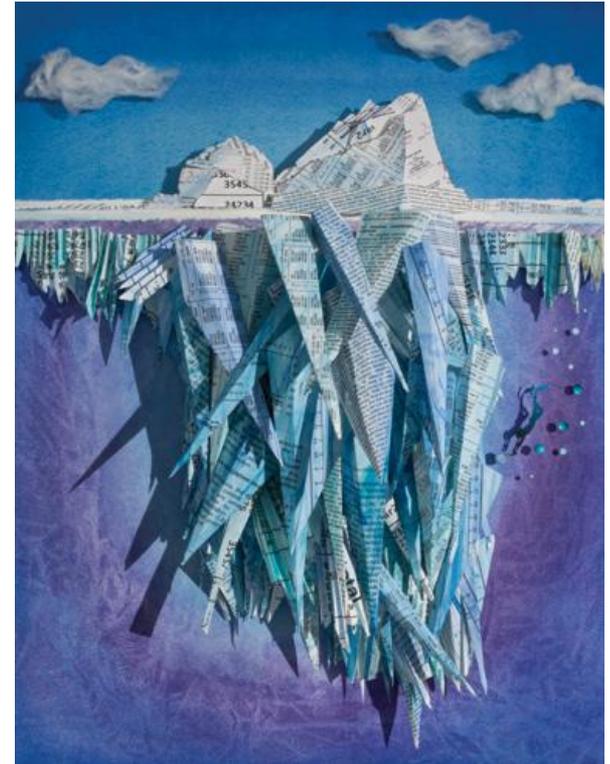
Données	<p>Tout type de données produites dans le cadre de la recherche, définies comme enregistrements factuels qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche.</p>	<p>Données tabulaires, images, sons, vidéos, données 3D, séquençage, données d'observation, données relationnelles, textes, annotations, etc.</p>	<p>Principe « aussi ouvert que possible, aussi fermé que nécessaire », principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable)</p>	<p>https://zenodo.org/ https://www.humanum.fr/consortiums</p>
<p>Comité pour la science ouverte, <i>Types de documents, productions et activités valorisées par la science ouverte et éligibles à une évaluation</i>, décembre 2019 (en ligne)</p>				
Plans de gestion de données	<p>Plans élaborés dans le cadre de projets de recherche, incluant les formats, la description des données et les règles de partage.</p>		<p>Renseignement et dépôt dans une plateforme spécialisée</p>	<p>https://dmp.opidor.fr/public_plans</p>
Logiciels	<p>Tout type de code et d'interface</p>	<p>Distinguer (1) les services accessibles en ligne (ex. application web), (2) les logiciels téléchargeables-réutilisables (mais dont le code n'est pas ouvert), (3) les logiciels libres et <i>open-source</i>. En qualifier la réutilisabilité : documentation, tutoriel communauté d'utilisateur, communauté de développement.</p>	<p>Dépôt dans une archive ouverte adéquate gérant les versions et moissonnée par Software Heritage</p>	<p>www.softwareheritage.org</p>
Protocoles de recherches	<p>Description de méthodes, de démarche de recherche</p>	<p>Design d'essais cliniques</p>	<p>Dépôt dans une archive ouverte adéquate</p>	<p>https://www.protocols.io/ https://cos.io/rr/ https://clinicaltrials.gov/</p>

Le défi de la valorisation

DATA: BY THE NUMBERS



J. Cham, « [Data : by the numbers](#) »,
PhD Comics, 31/05/2004

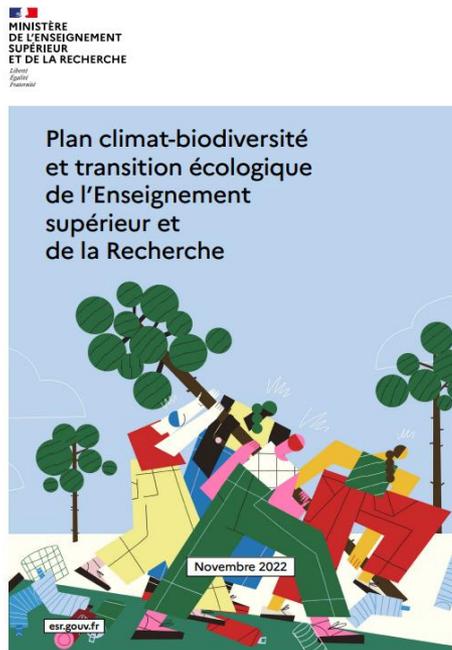


Tip of the iceberg,
The Scientist

On estime que :

- 10% seulement des données produites sont mises à disposition
- 90% restent stockées sur les disques durs sans valorisation

Enjeux environnementaux



[Plan Climat, biodiversité et transition écologique de l'ESR \(novembre 2022\)](#)



[Avis du COMETS "Intégrer les enjeux environnementaux à la conduite de la recherche - Une responsabilité éthique« \(décembre 2022\)](#)

Vers des pratiques numériques responsables sur les données de recherche

Pistes d'action dans :

- « Agir sur les données », [Ecoinfo](#)
- « Les impacts environnementaux et sociétaux des données : un défi pour l'avenir », [JRES 2022](#)
- Collectif [Labos 1.5](#)
- « Numérique responsable », [Collection AMUE #12](#)
- Dossier « Vers la sobriété numérique », [Arabesques, 2023](#)
- [OpenDataFrance – GreenData](#)

Bonnes pratiques

- Stockage hors ligne
- Sobriété dans les usages
- Gestion intelligente des flux de données
- Utilisation des formats et logiciels libres
- Bonne distance physique
- Minimisation des coûts de matériel
- Systèmes d'archivage passif
- Science ouverte et FAIR

Vers des pratiques numériques responsables sur les données de recherche

Pistes d'action dans :

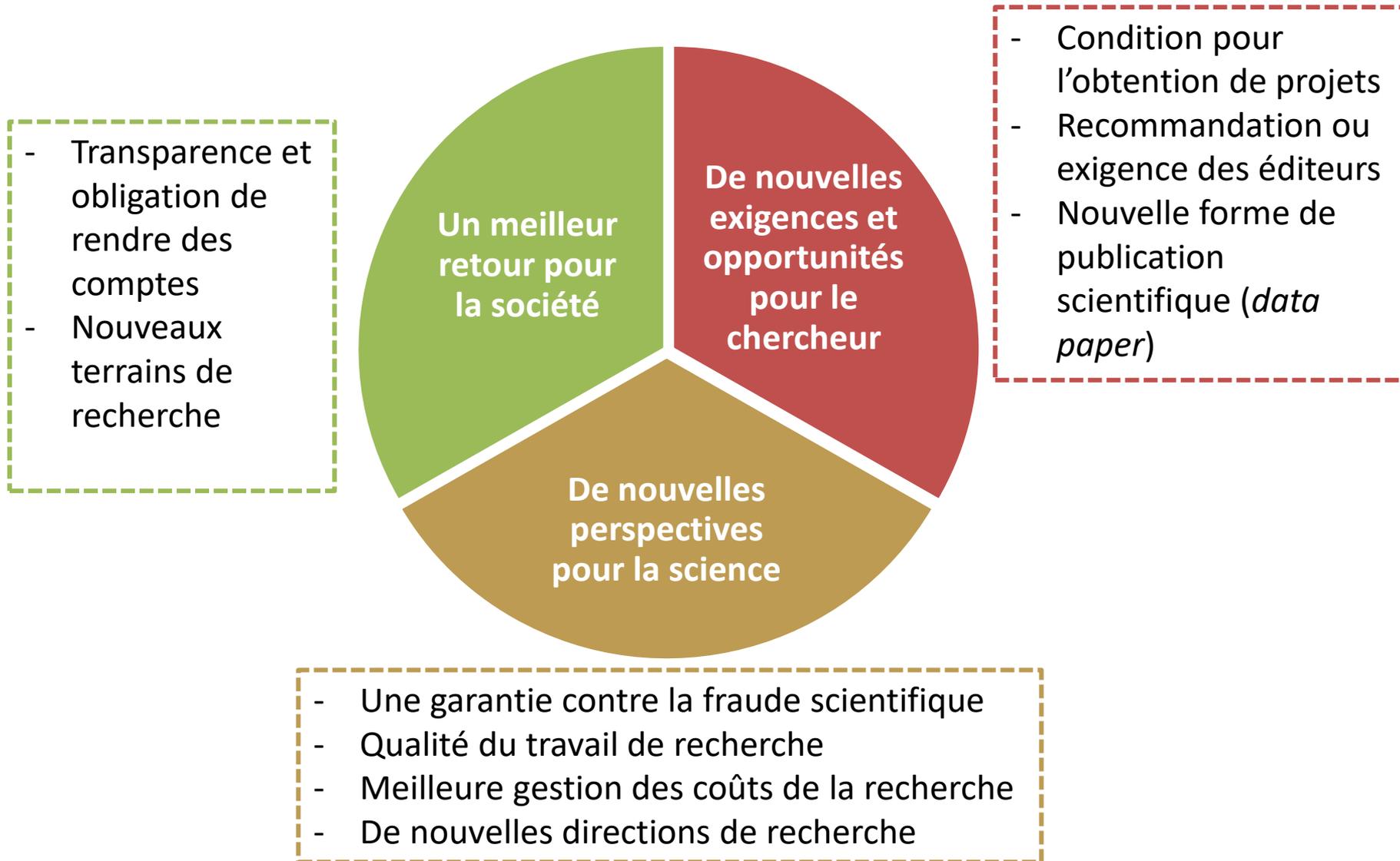
- « Agir sur les données », [Ecoinfo](#)
- « Les impacts environnementaux et sociétaux des données : un défi pour l'avenir », [JRES 2022](#)
- Collectif [Labos 1.5](#)
- « Numérique responsable », [Collection AMUE #12](#)
- Dossier « Vers la sobriété numérique », [Arabesques, 2023](#)
- [OpenDataFrance – GreenData](#)

Bonnes pratiques

- **Minimisation des coûts de matériel**
- Gestion intelligente des flux de données
- Bonne distance physique
- Systèmes d'archivage passif
- Stockage hors ligne

- Sobriété dans les usages
- Utilisation des formats et logiciels libres
- Science ouverte et FAIR

Des enjeux à différentes échelles



Du côté de l'Open data, les mêmes enjeux ?



Ressources OpenDataFrance

[Lien vers les ressources](#)

Description des ressources

LES ENJEUX POLITIQUES ET STRATÉGIQUES DE LA DONNÉE

Les enjeux politiques et stratégiques ∨

1 - Transparence

2 - Conformité réglementaire

3 - Transformation

4 - Confiance

5 - Souveraineté

6 - Gouvernance et coopération

7 - Valorisation

8 - Pilotage

9 - Transition

2. Définitions



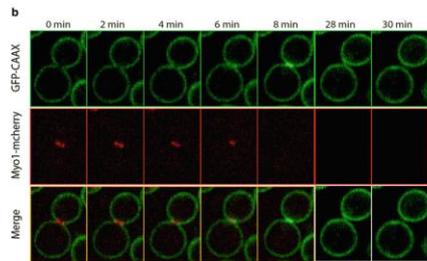


Fig. 1 Time-lapse of cells expressing GFP-CAAX and Cdc3-mCherry (a) or Myo1-mCherry (b). Maximal projections are shown for the mCherry channel but a single medial confocal slice is shown for GFP. The time is indicated in minutes

[Source](#)

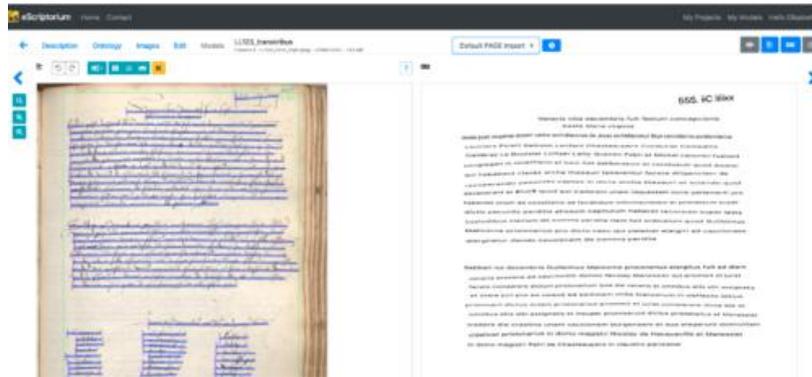
Table A1 - Part I. Summary of numbers of samples, sample origins and results

(sum of samples of national and co-ordinated programme)
 (pesticides covered by Directives 76/895, 86/362 and 90/642 and by national programmes)
 (surveillance sampling only, no follow-up enforcement sampling, including organic produce)

Reporting country: Sweden
 Year of sampling: 2007

	Number of samples		Sample origin						
	Total number of samples	Number of domestic samples	% domestic samples of total number of samples	Number of samples from other EU MS	% samples from other EU MS of the total number of samples	Number of samples on imports from TC	% samples from TC of the total number of samples	Number of samples with unknown origin	% samples from unknown origin of the total number of samples
Sum (certain products of plant origin, incl. fruit, vegetables)	1119	220	19,7	328	29,3	570	50,9	1	0,1
Cereals	264	166	62,9	20	7,6	70	26,5	8	3,0
Processed products (other than baby food)	98	12	12,2	46	46,9	20	20,4	20	20,4
Baby food	44	34	77,3	6	13,6	0	0,0	4	9,1

[Source](#)



[Source](#)

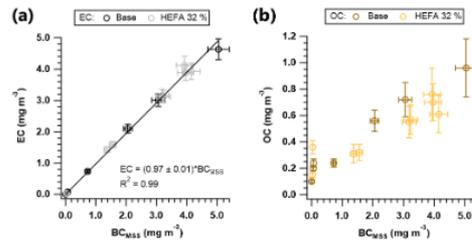


Figure 3. (a) Scatter plot and linear regression line of the mass concentrations of elemental carbon (EC) and black carbon measured with the micro soot sensor (BC_{MSS}). (b) Scatter plot of the mass concentrations of organic carbon (OC) and BC_{MSS}. Note that dark colours represent measurements with base fuel (Jet A-1), and light colours represent measurements with the HEFA blend (32 % vol).

```

1 // ConfirmedExoplanetsNASA
2
3 2020 2020 2020 2020 0002 0000 0009 0000
4 0032 0000 0189 0000 0002 0000 01bb 0000
5 0000 0000 0000 0000 0000 0000 0000 0000
6 0000 0000 0000 0000 0000 0000 0000 0000
7 0000 0000 4154 5452 0000 0000 0000 01bb
8 0000 00d8 0000 00e3 0000 0000 0000 0000
9 0000 0000 0000 0003 0000 00d8 0000 0048
10 0000 0f63 6f6d 2e61 7070 6c65 2e6d 6163
11 6c00 0000 0000 0120 0000 0048 0000 1563
12 6f6d 2e61 7070 6c65 2e71 7561 7261 6e74
13 696e 6500 0000 0168 0000 0053 0000 1763
14 6f6d 2e64 726f 7062 6f78 2e61 7474 7269
15 6275 7465 7300 0000 0100 0db6 dd74 ecfc
  
```

[Source](#)

[Source](#)

Données de la 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 des résultats de recherche. »



OCDE, *Recommandation du Conseil concernant l'accès aux données de la recherche financée sur fonds publics*, 2021

(1^e occurrence : *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, 2007).



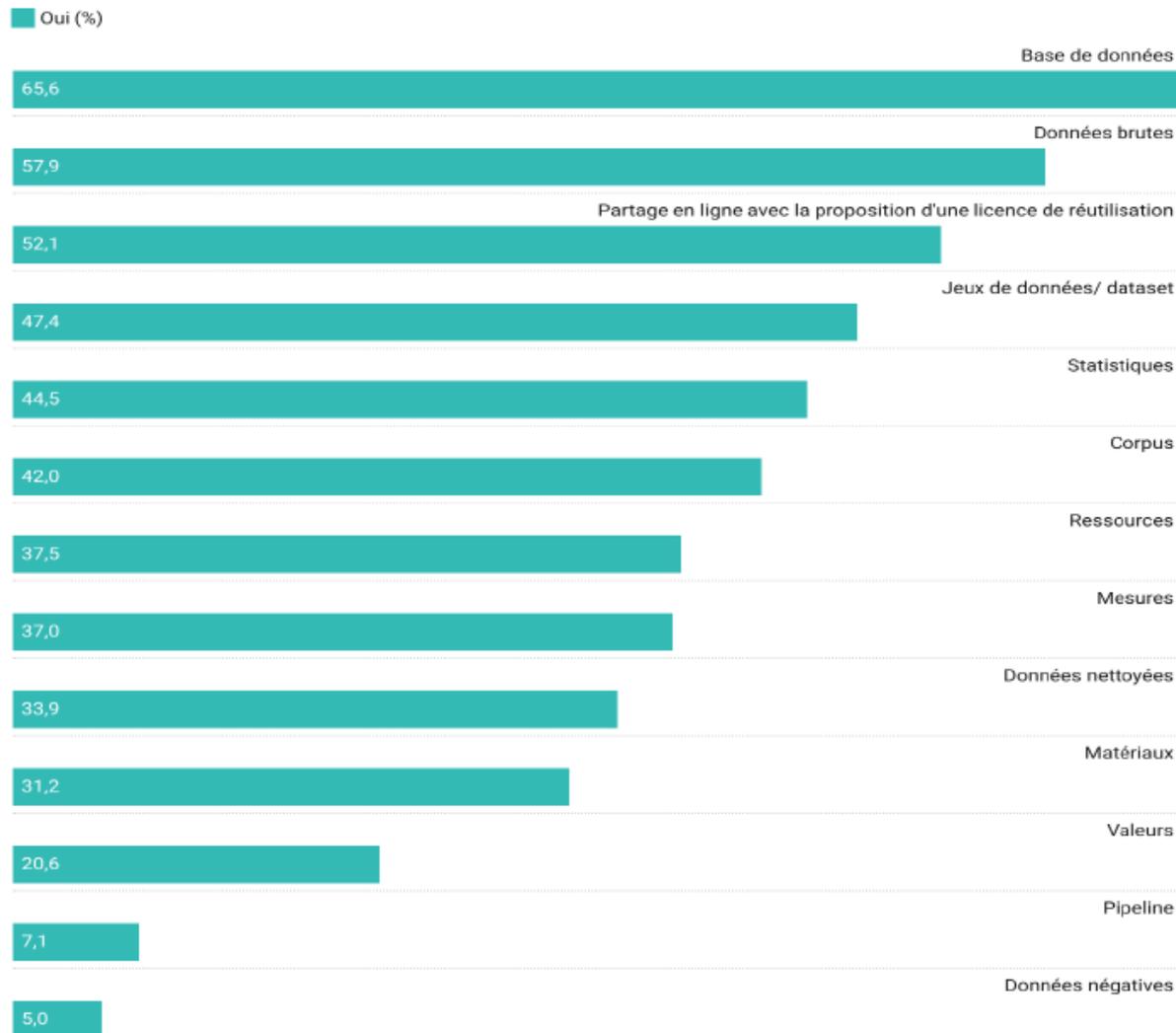
« Cette expression ne couvre pas les carnets de laboratoire, analyses préliminaires ou projets de documents scientifiques, programmes de travaux de recherche futurs, examens par les pairs, communications personnelles avec des collègues, ou objets matériels (échantillons de laboratoire, souches bactériennes ou animaux de laboratoire, par exemple). »



« **Autres objets numériques pertinents** au regard de la recherche financés sur fonds publics : métadonnées, algorithmes, flux de travail, modèles et logiciels (y compris le code connexe) issus de la recherche financée en tout ou partie sur fonds publics et utilisés dans le cadre de travaux de recherche et de développement. »

Le terme de « données » est largement employé aujourd'hui. Y a-t-il d'autres termes (ou des qualificatifs supplémentaires) que vous utilisez pour décrire les éléments que vous manipulez dans votre travail et qui vous servent à l'obtention des résultats de

Réponse à choix multiples.



Synthèse de l'étude
[Décliner la science ouverte](#), COSO, 2022

Typologies des données de recherche

➤ Par mode d'obtention

C.L. Borgman, *Big data, little data, no data*, 2015 / *Qu'est-ce que le travail scientifique des données ?* [2020](#)

Données d'observation

- capturées en temps réel
- habituellement uniques, impossible à reproduire

Ex.: mesures sismiques, images d'une étoile en fin de vie, enquêtes sociologiques, fouilles archéologiques...

Données d'expérimentation

- obtenues à partir d'équipements de laboratoire
- souvent reproductibles, parfois coûteuses

Ex.: séquences peptidiques, résultats de réactions chimiques, observations sur des individus en situation de test...

Données de simulation

- générées par des modèles informatiques
- souvent reproductibles si le modèle est correctement documenté

Ex.: modélisation du changement climatique, « reproduction » du Big Bang, modèles économiques...

Records – Traces (C. Borgman)

- documents témoignant d'un phénomène ou d'une activité humaine
- uniques ou non

Ex.: fonds de photographies, documents d'archives, textes de loi, ouvrages littéraires...

+ Données compilées ou dérivées

- issues du traitement de données brutes
- souvent reproductibles mais coûteuses

Ex.: bases de données compilées, corpus textuel préparé pour le TDM...

Typologies des données de recherche

Une nouvelle proposition, issue du rapport *Données de recherche : caractérisation des fonctions clés liées à leur gestion, leur exploitation, leur diffusion et leur utilisation*, I. Blanc et A. Laurent [2025](#)

- Données d'expérimentation
- Données issues de capteurs ou d'instruments scientifiques
- Données d'enquête
- Données synthétiques (reproduisant les caractéristiques statistiques de données réelles)
- Données de simulation (générées par des modèles informatiques ou mathématiques, reproduisant artificiellement un phénomène)
- Données textuelles ou documentaires
- Données dérivées

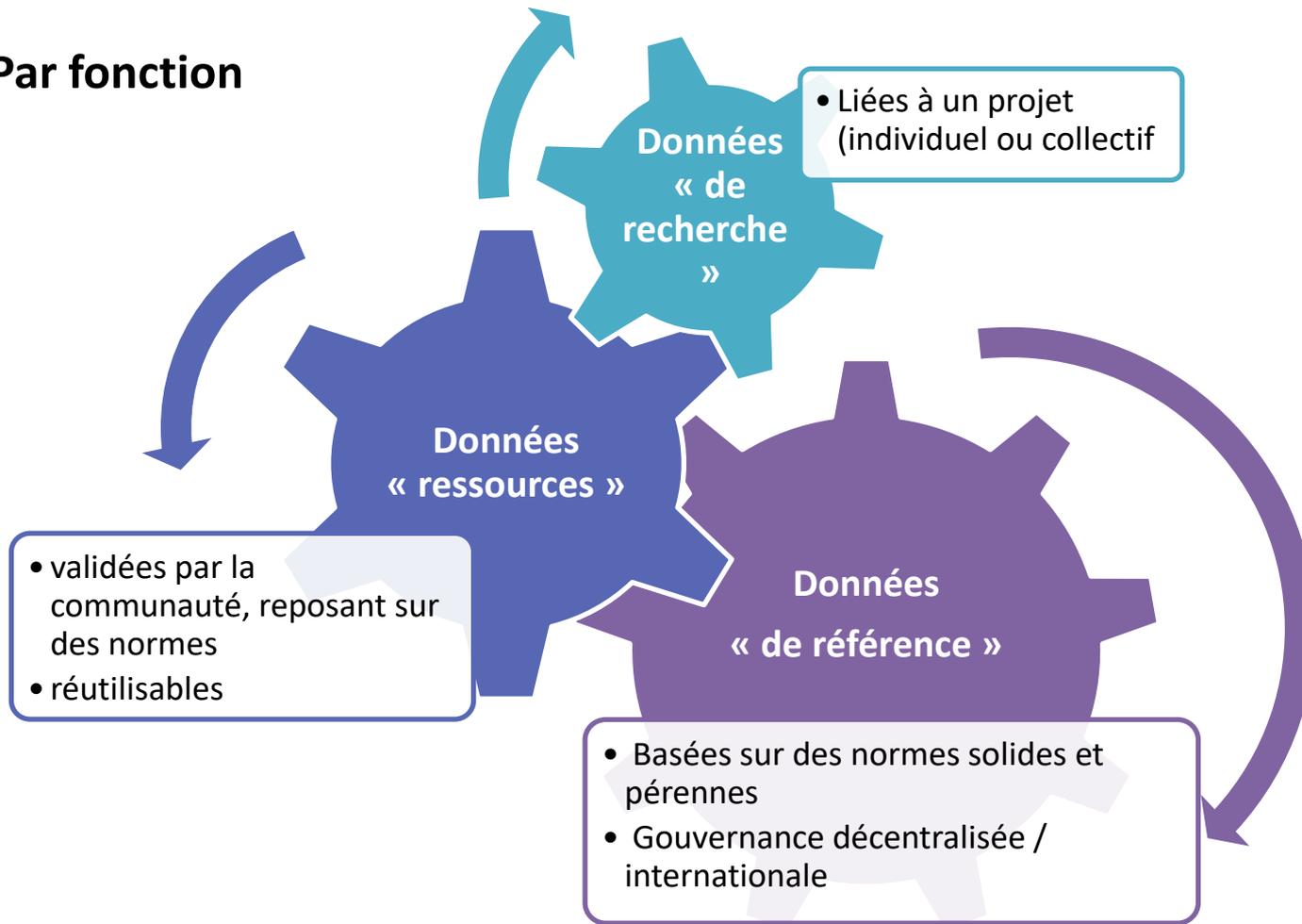
Typologies des données de recherche

Une nouvelle proposition, issue du rapport *Données de recherche : caractérisation des fonctions clés liées à leur gestion, leur exploitation, leur diffusion et leur utilisation*, I. Blanc et A. Laurent [2025](#)

- Données d'expérimentation
- Données issues de capteurs ou d'instruments scientifiques
- Données d'enquête
- Données synthétiques (reproduisant les caractéristiques statistiques de données réelles)
- Données de simulation (générées par des modèles informatiques ou mathématiques, reproduisant artificiellement un phénomène)
- Données textuelles ou documentaires
- Données dérivées

Typologies des données de recherche

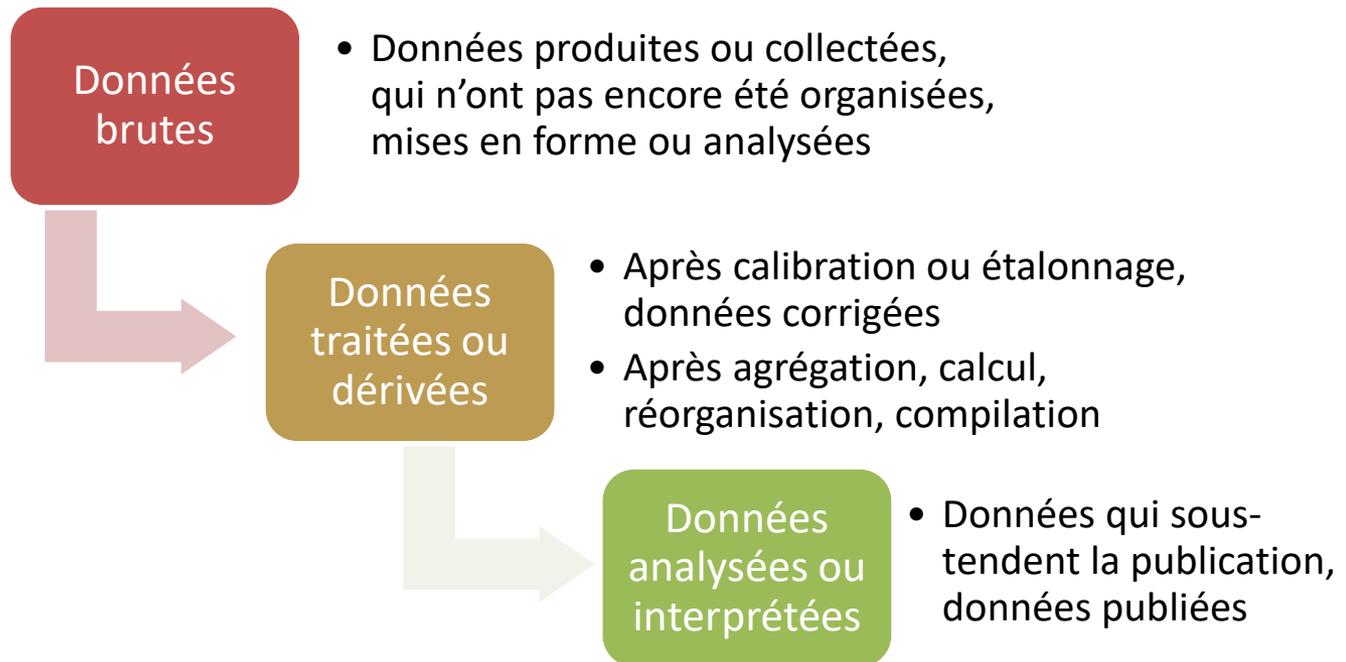
➤ Par fonction



C. L. Borgman, *Big data, little data, no data*, 2015
Qu'est-ce que le travail scientifique des données ? [2020](#)

Typologies des données de recherche

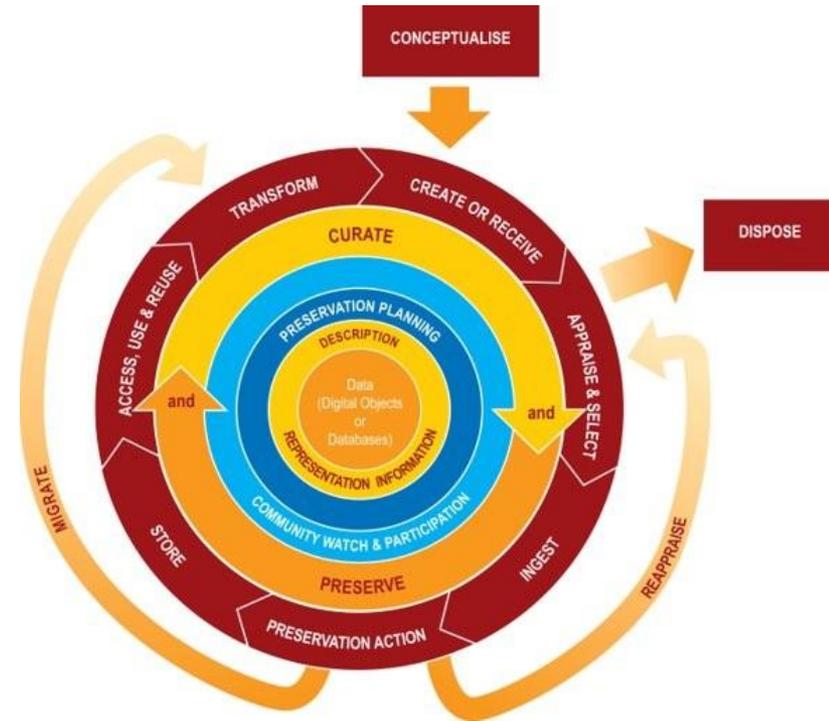
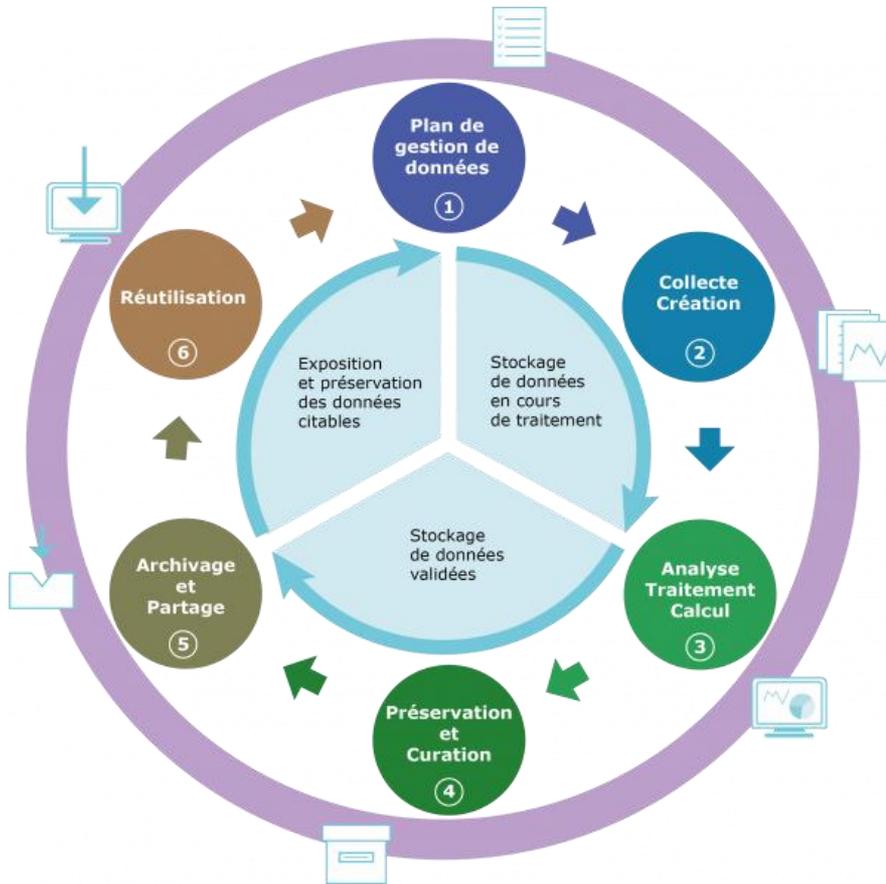
➤ Par niveau de traitement



Variations : Typologie des données du [CERN](#) pour le LHC ([data policy 2020](#))

1. Données à l'appui des résultats publiés (*published results*)
2. Ensembles conçus à des fins pédagogiques ou de vulgarisation (*outreach and education data*)
3. Données requises pour produire les résultats scientifiques (*reconstructed data*)
4. Données brutes (*raw data*)

Le cycle de vie des données



DCC Lifecycle

[Cycle de vie des données](#),

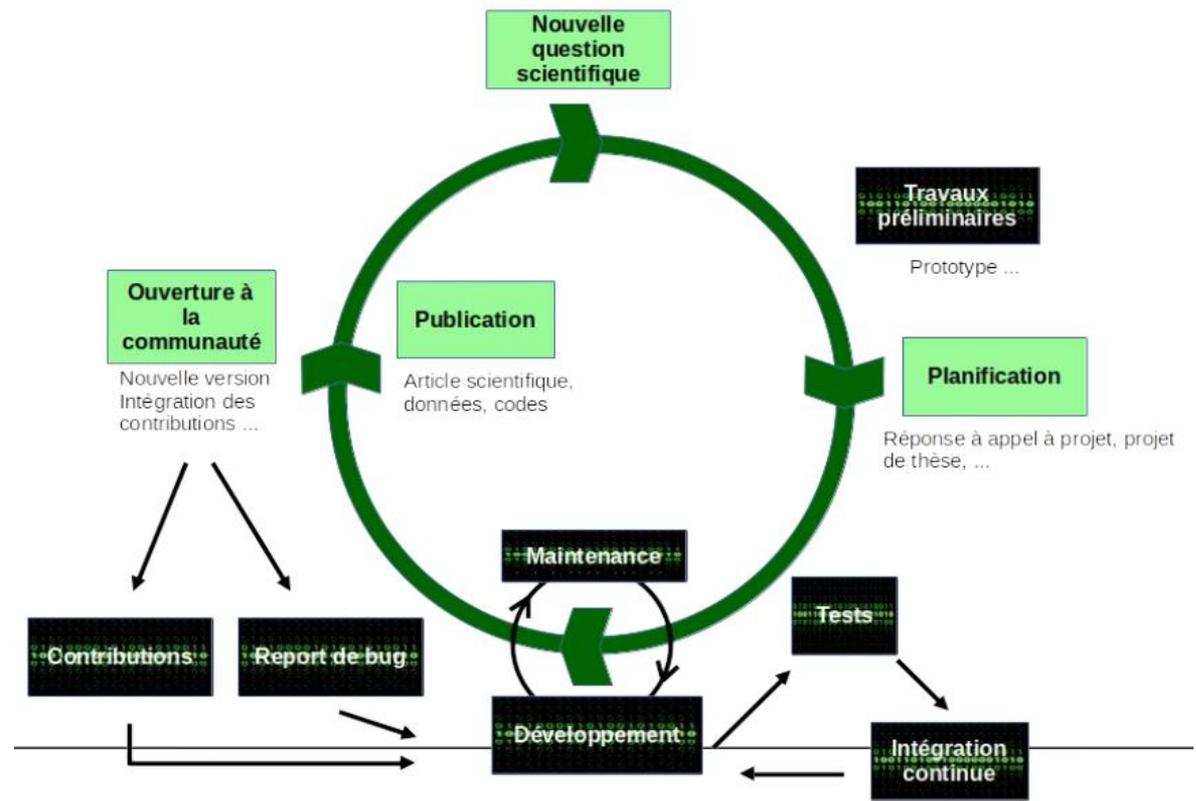
Science ouverte Grenoble Alpes, 2022, Inspiré de [Research Data Lifecycle](#), UK Data Archive

Autour des données : les codes et logiciels



[Passport Codes et logiciels, 2023](#)

Cycle de vie d'un code de recherche



Violaine Loubet ([Formation ADBU-MediaNormandie, 2023](#))



Deux exemples

- <https://www.nakala.fr/collection/10.34847/nkl.3f82fdy8>
- https://search-data.ubfc.fr/FR-18008901306731-2016-05-11_PROTEKER-a-near-shore-submarine-observatory-in.html

Quelles sont les principales caractéristiques de ces jeux de données ?

- Mode d'obtention ?
- Fonction ? (Recherche individuelle ? Données communautaires ? Données de référence ?)
- Niveau de traitement ?

3. Le cadre juridique et réglementaire



Pour une première approche :

- [Cours introductif sur les aspects juridiques et éthiques](#), DoRANum 2022

- [Connaître le droit des données de la recherche](#), A. Bassinet et F. Bouchet-Moneret, 2024

Ouvrage de référence :

A. Robin, *Droit des données de la recherche : science ouverte, innovation, politiques publiques*, Bruxelles, 2022

Principaux jalons

Open data

1978
Loi Cada
Accès aux documents administratifs

2003
Déclaration de Berlin
Libre accès à la connaissance

2003
Directive PSI
Réutilisation des documents publics

2003
Déclaration de l'OCDE
Accès aux données de la recherche financée par des fonds publics

2003
Revision de la directive PSI
Charte du G8 Pour l'ouverture des données publiques

2012
Obligation OA pour projets européens

2013
Loi Valter
sur la gratuité et les modalités de réutilisation

2016
Loi pour une République numérique

2015
Directive sur les données de forte valeur

2018
1^{er} Plan national pour la Science Ouverte
Libre accès aux publications et vers l'ouverture des données

2019
Règlement européen / Liste des données de forte valeur

2019
Obligation PGD pour projets ANR

2021
2^e Plan national pour la Science Ouverte :
Élargissement (codes source, pratiques, financement)

Science ouverte

Données de la recherche : des documents administratifs

- *Code régissant les relations entre le public et l'administration (CRPA)*
 - **Principe d'ouverture par défaut et de libre réutilisation**
- *Loi pour une République numérique (art. 30 II = Code de recherche L533-4)*
 - Précision sur le cadre d'application propre à la recherche
Périmètre : « activités de recherche **financées au moins pour moitié** par des dotations » (publiques, nationales, européennes) sous réserve de
 - Respecter le cas échéant les **droits spécifiques et réglementations particulières**
 - Avoir été **rendues publiques** par le chercheur, l'établissement ou l'organisme
 - Être **achevées**

La loi pour une République numérique (octobre 2016)

Article 30

II.- Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont **pas protégées par un droit spécifique ou une réglementation particulière** et qu'elles ont été **rendues publiques** par le chercheur, l'établissement ou l'organisme de recherche, **leur réutilisation est libre.** »

Les documents administratifs sont librement communicables à toute personne qui en fait la demande, sauf exceptions légales.

*Si possible, mise à disposition exhaustive et immédiate, pour tous
Sinon, mise à disposition ciblée, après examen et sur demande*

Communication obligatoire

- Données **géographiques informatisées** (directive INSPIRE): communication obligatoire d'office
- Données relatives à des émissions de substances dans l'**environnement** (convention d'Aarhus): communication obligatoire sur demande

Communication interdite

- Données présentant des **risques** pour la défense nationale
- Données à risque pour la **sécurité** de l'État, la sécurité publique, la sécurité de l'établissement
- Données protégées par des **secrets** (industriel et commercial, médical, financier...)

Communication sous conditions

- Données présentant des risques pour la protection du **potentiel scientifique** et technique de la nation
- Données protégées par le **droit d'auteur** ou autre droit de propriété intellectuelle
- Données **personnelles**
- Données **statistiques**
- Données liées à un contrat avec un tiers

➤ **Article D211-2 du Code de la recherche**

(qui reprend en partie l'ancien [Décret n° 2021-1572 du 3 décembre 2021](#) relatif au respect des exigences de l'intégrité scientifique dans les établissements de recherche)

Les établissements publics et fondations reconnues d'utilité publique mentionnés au troisième alinéa de l'article [L. 211-2](#) :

1° Veillent à ce que les travaux de recherche qu'ils conduisent ou auxquels ils participent respectent les exigences de l'intégrité scientifique ;

2° Assurent la formation des personnels et des étudiants au respect de ces exigences ;

3° Promeuvent la diffusion des publications en accès ouvert et la **mise à disposition des méthodes, protocoles, données et codes sources associés aux résultats de la recherche** ;

4° Définissent les **conditions de conservation, de communication et de réutilisation des résultats bruts** des travaux scientifiques menés en leur sein ;

5° Veillent à ce que tout signalement relatif à un éventuel manquement aux exigences de l'intégrité scientifique soit traité selon une procédure établie au regard des recommandations du Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur définies en application des dispositions de l'article [L. 114-3-1](#).

Les plans nationaux pour la Science Ouverte (PNSO)



I- 2018-2021

1. Généraliser l'accès ouvert aux publications
2. Structurer et ouvrir les données de recherche
3. S'inscrire dans une dynamique durable, européenne et internationale



II- 2021-2024 : Généraliser la science ouverte en France

1. Généraliser l'accès ouvert aux publications
2. Structurer, ouvrir et partager les données de recherche
3. Ouvrir et promouvoir les codes sources produits par la recherche
4. Transformer les pratiques pour faire de la science ouverte le principe par défaut

Les politiques d'établissement

Université Paris Saclay



Et d'organismes de recherche:

- [CNRS](#)
- [INRAE](#)
- [IRD](#)
- [INSERM](#)
- [CEA](#)

Université Bordeaux
Montaigne

Université Paris Cité

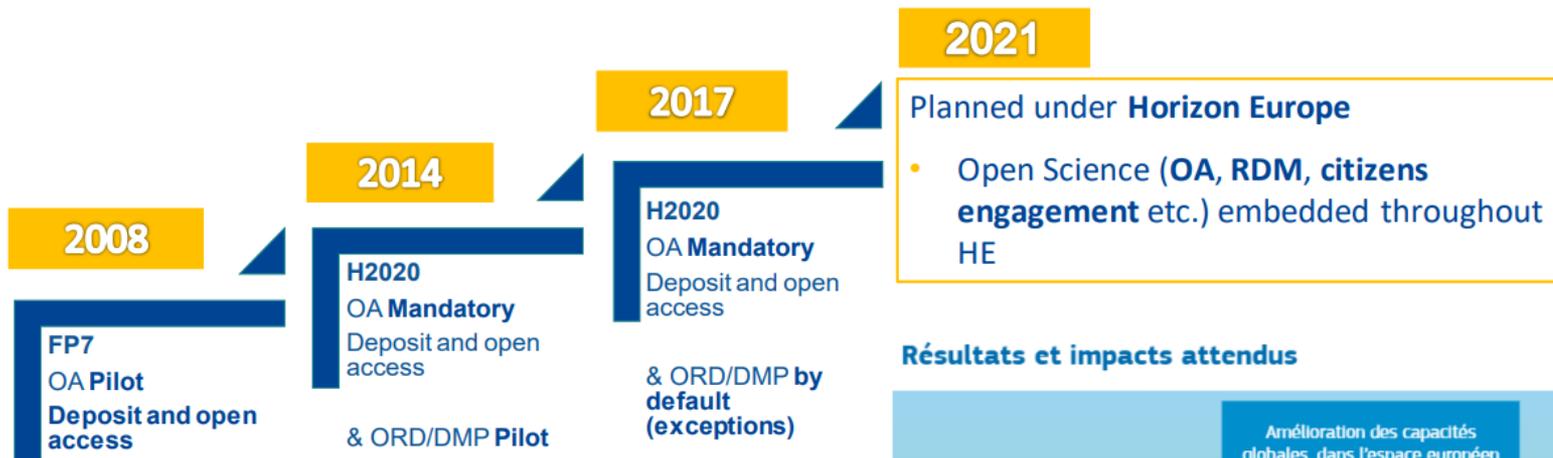
Université Lyon 2

Sorbonne Universit 

Universit  Sorbonne Paris Nord

Les attentes des programmes cadres européens

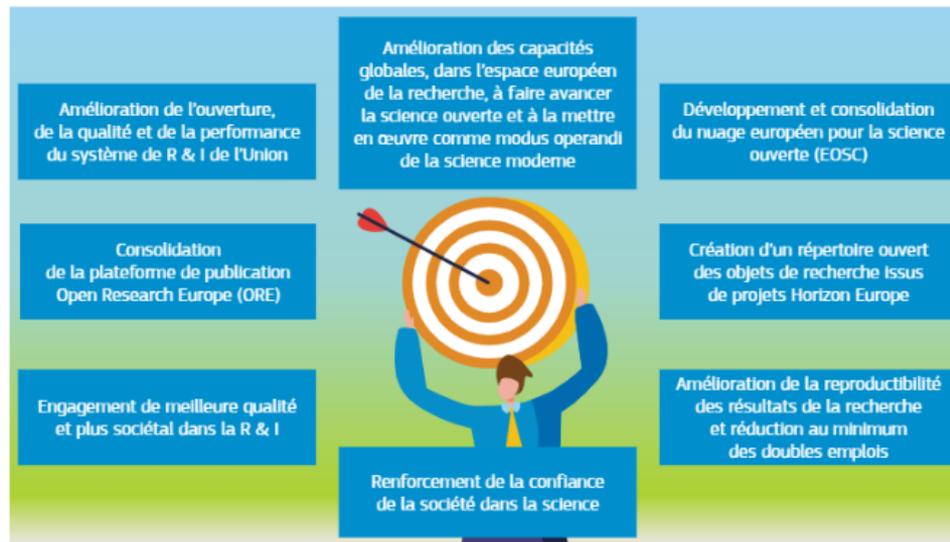
Evolution of open science policy across FPs



Exploitation and Open Science in Horizon Europe

Fiche : La science ouverte dans Horizon Europe

Résultats et impacts attendus





La Science ouverte dans Horizon Europe

Pratiques de SO obligatoires

- Publications en accès libre immédiat
- **Respect des principes FAIR en matière de gestion de données**
- **Rédaction d'un PGD**
- **Information sur l'accès aux données**
- **Dépôt des données de recherche, avec accès aussi ouvert que possible et aussi fermé que nécessaire**

Pratiques SO recommandées

- Implication de la société
- Partage et ouverture des résultats le plus tôt possible
- Utilisation d'infrastructures ouvertes
- Évaluation ouverte de la recherche
- ...

Accès immédiat à tous les résultats en cas d'urgence



[Horizon Europe -
Programme Guide v1.5 -
2022](#)

Voir RDM in Horizon Europe :
[Guide OpenAire](#)

the proposal stage. Proposers should be aware of both mandatory and recommended practices and integrate them into their proposals.

Open science practices include early and open sharing of research (for example through preregistration, registered reports, pre-prints, or crowd-sourcing); research output¹⁸ management; measures to ensure reproducibility of research outputs; providing open access to research outputs (such as publications, data, software, models, algorithms, and workflows); participation in open peer-review; and involving all relevant knowledge actors including citizens, civil society and end users in the co-creation of R&I agendas and contents (such as citizen science).

These practices are explained and relevant resources provided in a separate section further below (open science practices and resources).

Mandatory open science practices

- Some open science practices are **mandatory for all beneficiaries per the grant agreement**. They concern:
 - open access to scientific publications under the conditions required by the grant agreement;
 - responsible management of research data in line with the FAIR principles of 'Findability', 'Accessibility', 'Interoperability' and 'Reusability', notably through the generalised use of data management plans, and open access to research data under the principle 'as open as possible, as closed as necessary', under the conditions required by the grant agreement;
 - information about the research outputs/tools/instruments needed to validate the conclusions of scientific publications or to validate/re-use research data;
 - digital or physical access to the results needed to validate the conclusions of scientific publications, unless exceptions apply;
 - in cases of public emergency, if requested by the granting authority, immediate open access to all research outputs under open licenses or, if exceptions apply, access under fair and reasonable conditions to legal entities that need the research outputs to address the public emergency¹⁹.

These obligations are described in the Model Grant Agreement (Article 17) and detailed guidelines on complying with them are provided in the Annotated Grant Agreement (Article 17).

- Some open science practices are **mandatory per specific work programmes or call conditions**, which may provide for additional obligations to adhere to open science practices.

Recommended open science practices

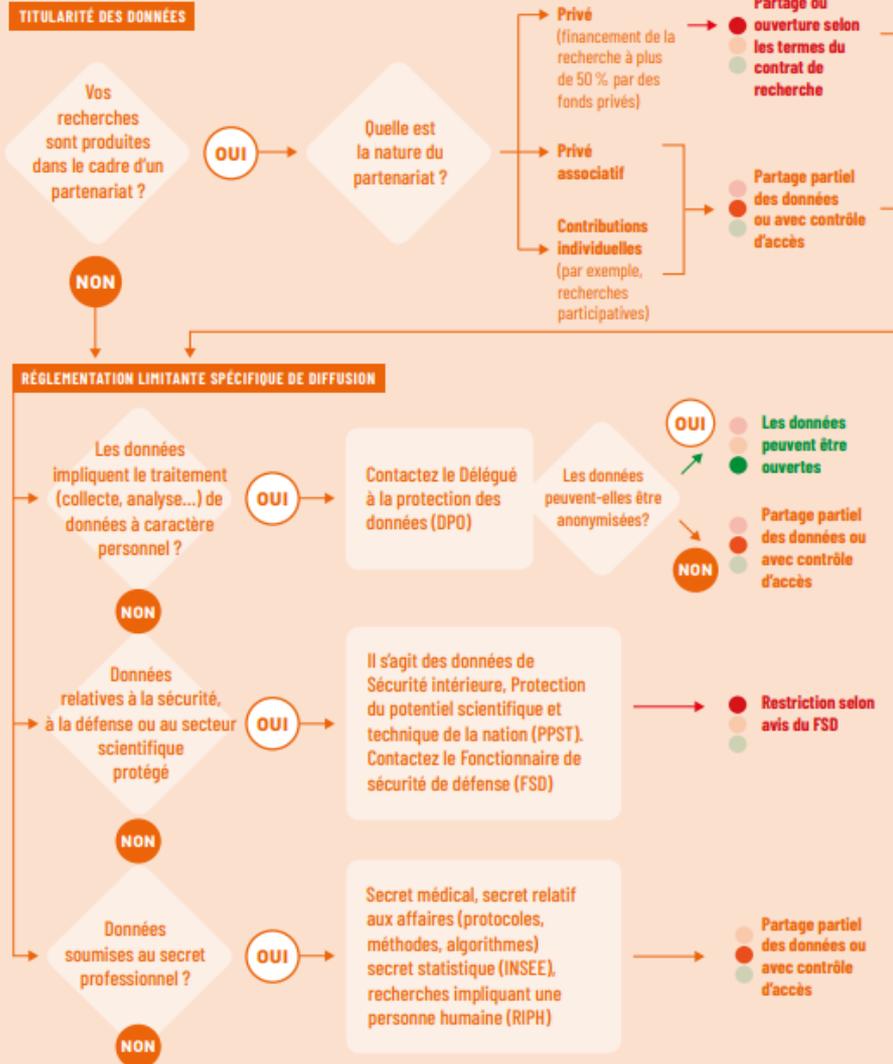
La politique des agences de financements en France



[Une politique commune depuis mars 2022](#)

- Favoriser le libre accès pour les publications scientifiques issues des projets financés
- **Promouvoir le partage et l'ouverture des données de la recherche**
- Partager et harmoniser nos pratiques et modalités d'évaluation des projets
- Informer et sensibiliser aux pratiques de science ouverte
- **Ouvrir les données relatives aux projets financés**
- **Déployer un modèle commun de plan de gestion des données structuré**
- Prendre part aux discussions sur la réforme de l'évaluation de la recherche
- Poursuivre la promotion de l'adoption de l'identifiant ORCID

Vous avez produit des données dans le cadre de vos activités de recherche, mais vous vous interrogez sur les modalités de partage et diffusion ?
Voici quelques questions pour vous guider :



Vous avez répondu NON à toutes ces questions ? Une fois achevées et validées scientifiquement, les données doivent être ouvertes. Le personnel d'appui de proximité peut vous accompagner dans la gestion et la préparation de la diffusion de vos données.

Et vous, dans quel cas vous situez-vous ?

[Logigramme](#) du Passeport Données de recherche, 2024

Voir aussi

[Qui a les droits ? Logigramme](#), Pôle IST de l'Ecole Ponts ParisTech, 2018

[Logigramme Ouverture des données](#), INRAE, 2022

[Arbre d'aide à la décision](#), Coop-IST, 2021

[Logigrammes sur les données personnelles](#), ARDoISE, 2025

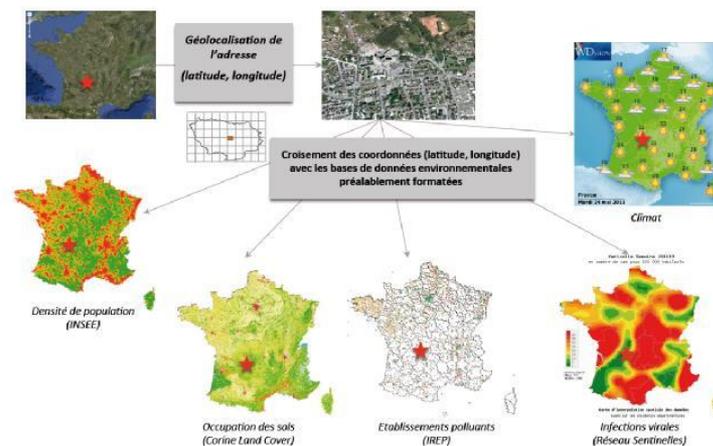
Cas particulier des données environnementales

Convention d'Aarhus (1998) :

- Assurer l'accès au public de l'information sur l'environnement
- Favoriser la participation du public à la prise de décision ayant des incidences sur l'environnement
- Étendre les conditions d'accès à la justice en matière d'environnement

Directive européenne Inspire de mars 2007 :

- ouverture systématique de 34 types de données géographiques dès lors qu'elles sont disponibles sous forme numérique
- Facilite leur partage et leur interopérabilité : <https://inspire-geoportal.ec.europa.eu/>



Source : SIG 2011

+ Protocole d'Aarhus (1998) sur les [polluants organiques persistants](#) : ouverture immédiate obligatoire pour les données concernant l'émission de substances polluantes dans l'environnement

Cas particulier du « secret des affaires » ou des « secrets industriels et commerciaux »

Données peuvent concerner :

- Secret des procédés
- Secret des informations économiques et financières
- Secret des stratégies commerciales ou industrielles

= Tous les éléments sensibles ayant notamment un impact sur l'environnement concurrentiel de l'établissement et de ses partenaires

- Communication interdite en dehors des intéressés
- Ou communicables après occultation des informations couvertes par ce secret



Voir : une [comparaison à l'échelle internationale](#) par l'OMPI ; les articles du [Code du commerce](#) en France

Cas particulier du « secret statistique »

Données soumises au secret statistique :

« toute information ayant trait à la vie personnelle et familiale, aux faits et aux comportements d'ordre privé recueillis au moyen d'une enquête statistique » ([loi n°51-711 du 7 juin 1951 modifiée](#))

Accessibles à des fins de recherche selon des procédures particulières

- Après avis du Comité du secret statistique
- [Portail Quetelet-Progedo](#) (accès à des données pseudonymisées)
- [Centre d'accès sécurisé aux données](#) (accès aux données dans un cadre de confidentialité renforcée)



[Guide du secret statistique](#)
(INSEE, janvier 2023)

Cas particulier des données personnelles ou dites sensibles



La [loi Informatique et Libertés](#) (1978, révisée 2019)

Le [RGPD](#) (Règlement général sur la protection des données) applicable dans les pays de l'Union européenne depuis mai 2018

Données à caractère personnel : toutes les informations qui permettent d'identifier directement ou indirectement la personne (Article 4 du RGPD) :

- Les données directement identifiantes : nom, prénom, adresse, photo, voix, etc.
- Les données indirectement identifiantes : un numéro de téléphone, le croisement d'informations

Données dites « sensibles » : données qui révèlent la prétendue origine raciale ou l'origine ethnique, les opinions politiques, les convictions philosophiques ou religieuses, l'appartenance syndicale, l'orientation sexuelle, les données de santé, les données biométriques qui permettent d'identifier une personne, les données génétiques

- Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte, [Guide pour la recherche, INSHS, v2, 2021](#)

5 principes clés dans le traitement de ces données

1. La finalité : définir les objectifs du traitement
2. La minimisation : s'en tenir au strict nécessaire
3. La conservation : limiter la conservation des données
4. Les droits : respecter les droits des personnes
5. La sécurité : sécuriser les données



Ouverture et/ou communication : sur demande uniquement si au moins une de ces 3 conditions :

- Avoir recueilli le consentement des personnes
- Avoir convenablement anonymisé les données
- Obéir à un autre texte législatif ou réglementaire

RGPD et recherche ? Dérogations et sécurisation du cadre

- Finalité : peut rester assez indéterminée ; un traitement ultérieur selon d'autres finalités est possible
- Conservation : au-delà de la durée initiale du projet, à des fins archivistiques
- Accès autorisé à des tiers dans un cadre spécifique
- Accès possible aux données dites sensibles et aux données de santé

Cas particulier des recherches impliquant la personne humaine

« tout essai ou expérimentation organisé et pratiqué sur l'être humain, en vue du développement des connaissances biologiques ou médicales »

➤ Code de la Santé Publique (CSP) + loi Jardé (2012, modifiée 2016)

- recherche interventionnelle comportant une intervention sur la personne non justifiée par sa prise en charge habituelle
- recherche interventionnelle ne comportant que des risques et contraintes minimales
- recherche non interventionnelle



Comités de protection des personnes (CPP) : chargés d'émettre un avis sur

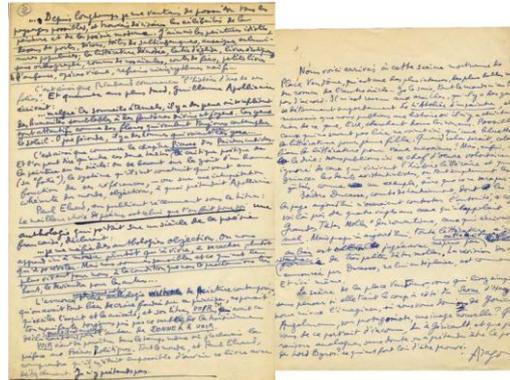
- les conditions dans lesquelles le promoteur de la recherche assure la protection des personnes et notamment des participants ;
- le bien-fondé et la pertinence du projet de recherche ;
- sa qualité méthodologique.

Et les données de santé ? Un cas particulier de données sensibles => RGPD

Cas des données protégées par le droit d'auteur de tiers



Images (y compris photos)



Œuvres littéraires (y compris inédites)



Logos, marques, designs

La diffusion publique de données protégées par le droit d'auteur de tiers est autorisée si :

- L'œuvre est dans le domaine public
- L'auteur a donné son autorisation expresse (attention : qui ne dit mot REFUSE)
- L'auteur a par ailleurs déjà autorisé ce type d'usage par le biais d'une licence

4. La gestion des données de recherche



Tom Gauld, *Le département des théories fumeuses*, 2020
([post du 22/04/2018](#))

Bonnes pratiques en matière de recherche

« Gestion et pratiques en matière de données »



ALLEA, [Code de conduite européen pour l'intégrité en recherche](#), révisé 2023

- Les chercheurs, les institutions et les organismes de recherche **assurent une gestion, une curation et une conservation** adéquates de toutes les **données, métadonnées, protocoles, codes, logiciels et autres matériaux de recherche** pendant une période raisonnable et clairement définie.
- Les chercheurs, les institutions et les organismes de recherche s'assurent que l'accès aux données est **aussi ouvert que possible, mais aussi fermé que nécessaire** et, le cas échéant, **conforme aux principes FAIR** (autrement dit, facile à trouver, accessible, interopérable et réutilisable) en ce qui concerne la gestion des données.
- Les chercheurs informent les participants à la recherche de la façon dont leurs données seront utilisées, réutilisées, consultées, conservées et supprimées, en **conformité avec le RGPD**.
- Les chercheurs, les institutions et les organismes de recherche font preuve de **transparence** concernant la manière de consulter ou d'utiliser leurs données, métadonnées, protocoles, codes, logiciels et autres matériaux de recherche.
- Les chercheurs, les institutions et les organismes de recherche reconnaissent que les données, métadonnées, protocoles, codes, logiciels et autres matériaux de recherche sont des **produits légitimes de la recherche**, qui peuvent être cités.
- Les chercheurs, les institutions et les organismes de recherche veillent à ce que les **contrats ou les accords** relatifs aux résultats de la recherche comprennent des **dispositions équitables et justes** en ce qui concerne la gestion de leur utilisation, de leur propriété, et/ou de leur protection en vertu des droits de propriété intellectuelle.

LE CYCLE DE VIE DES DONNÉES

Un outil pour améliorer la gestion / la mise en qualité / l'ouverture des données

PAR OÙI COMMENCE-T-ON ?

PGD

PEUT-ON RÉUTILISER DES DONNÉES ?

Produire des données Faciles à trouver, Accessibles, Interopérables et Réutilisables (FAIR)

Respecter le droit des personnes pour les Données à Caractère Personnel (DCP)

Ouvrir ses données autant que possible, les fermer autant que nécessaire



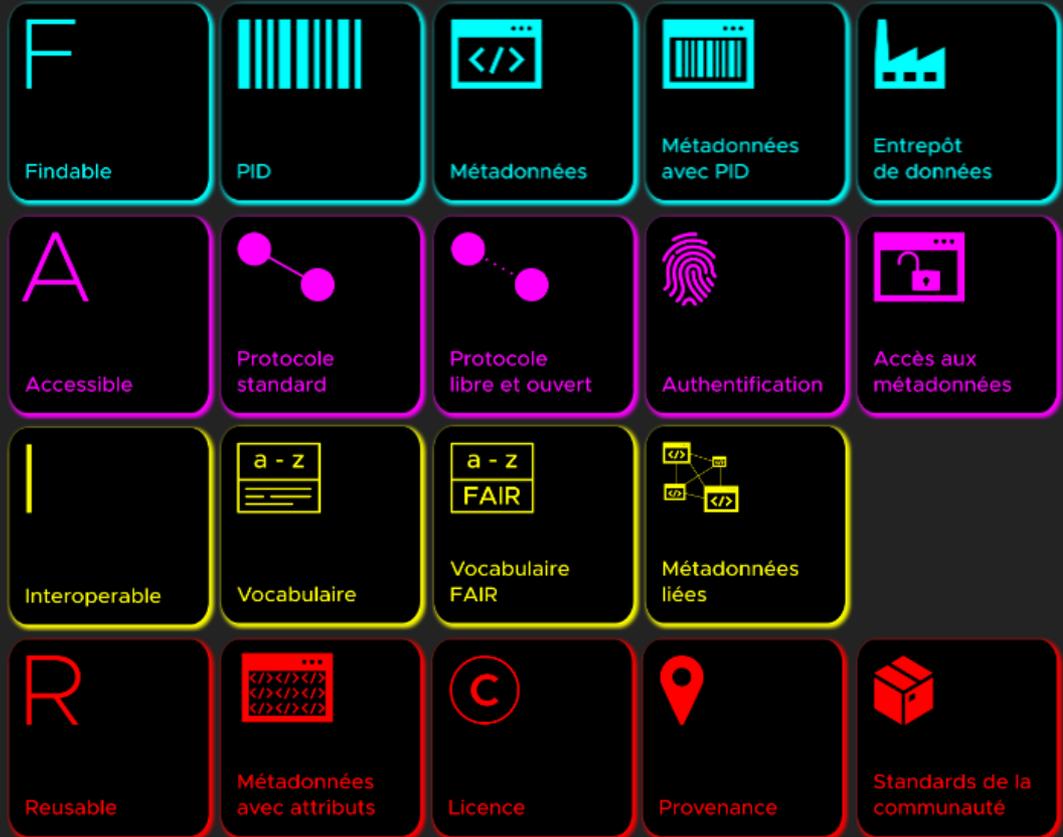
Le cycle de vie des données, INRAE, v 2 2024

Les principes FAIR

Présentation interactive sur [DoRANum](#)

Les principes FAIR

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



Voir aussi « [The FAIR Guiding Principles for scientific data management and stewardship](#) », Scientific data, 2016

Guide pratique

En français : <https://dorum.fr/appli-fair-aware-pleine-page-vf/>

En anglais : <https://www.fairsfair.eu/fair-aware>



FACILE A TROUVER

1. Savez-vous qu'un jeu de données déposé dans un entrepôt de données doit se voir attribuer un identifiant international, unique, pérenne et qui dirige vers une page de présentation des données ? ⓘ
2. Savez-vous que lorsque vous déposez un jeu de données dans un entrepôt, vous devez fournir certains détails (appelés métadonnées de référencement) afin de rendre vos données faciles à trouver, compréhensibles et réutilisables par d'autres personnes ? ⓘ
3. Savez-vous qu'un entrepôt qui donne accès à vos données doit fournir les métadonnées qui décrivent celles-ci dans un format apte à être lu par des machines aussi bien que par des humains ? ⓘ

ACCESSIBLE

4. Savez-vous que l'accès à votre jeu de données peut nécessiter d'être restreint (soumis à autorisation) et que vos métadonnées devraient comprendre des informations sur les licences applicables en cas de ré-utilisation de données ? ⓘ
5. Savez-vous que vos métadonnées doivent rester disponibles sur le long terme, même dans le cas où les données ne sont plus accessibles ? ⓘ



INTEROPERABLE

6. Savez-vous que les métadonnées qui décrivent vos jeux de données doivent se baser sur des vocabulaires contrôlés (standardisés, structurés sous la forme d'ontologies), autant que possible ? ⓘ

REUTILISABLE

7. Savez-vous que vos métadonnées doivent inclure des informations relatives à la provenance des données collectées et/ou générées ? ⓘ
8. Savez-vous que les métadonnées décrivant vos données doivent suivre des standards reconnus par la communauté scientifique dans votre domaine de recherche ? ⓘ
9. Savez-vous que vos données doivent être déposées de préférence dans un format de fichier ouvert (pour faciliter leur réutilisation) qui soit pris en charge par votre entrepôt, pour une conservation à long terme. ⓘ
10. Savez-vous que la conservation de vos jeux de données à long terme exige l'intervention de professionnels de la gestion et de la conservation des données ? ⓘ



Des données faciles à trouver ?



FACILE A TROUVER

1. Savez-vous qu'un jeu de données déposé dans un entrepôt de données doit se voir attribuer un identifiant international, unique, pérenne et qui dirige vers une page de présentation des données ? ⓘ
2. Savez-vous que lorsque vous déposez un jeu de données dans un entrepôt, vous devez fournir certains détails (appelés métadonnées de référencement) afin de rendre vos données faciles à trouver, compréhensibles et réutilisables par d'autres personnes ? ⓘ
3. Savez-vous qu'un entrepôt qui donne accès à vos données doit fournir les métadonnées qui décrivent celles-ci dans un format apte à être lu par des machines aussi bien que par des humains ? ⓘ

- **L'idée** : permettre à toute personne intéressée d'avoir connaissance de l'existence de vos données en interrogeant n'importe quel moteur de recherche, sans devoir explorer précisément la plateforme sur laquelle elles sont conservées
- **La solution** : décrire les caractéristiques de mes données de manière précise et compréhensible, à la fois pour un ordinateur et pour un humain, de façon à ce qu'elles soient renvoyées comme résultat d'une requête

La mise en pratique :

- Décrire vos données avec un maximum d'informations, notamment de mots-clés, par des **métadonnées**
- (Faire) attribuer un identifiant unique et pérenne à son jeu de données
 - **DOI** ou assimilé
- Déposer mes données dans un entrepôt indexé par les moteurs de recherche
 - Chercher un **entrepôt compatible avec le protocole OAI-PMH**

Des données accessibles ?



ACCESSIBLE

4. Savez-vous que l'accès à votre jeu de données peut nécessiter d'être restreint (soumis à autorisation) et que vos métadonnées devraient comprendre des informations sur les licences applicables en cas de ré-utilisation de données ? ⓘ

5. Savez-vous que vos métadonnées doivent rester disponibles sur le long terme, même dans le cas où les données ne sont plus accessibles ? ⓘ

- **L'idée** : une fois identifiées par un individu ou un moteur de recherche, vos données doivent pouvoir être obtenues (éventuellement sous certaines conditions)
- **La solution** : bien préciser les conditions selon lesquelles vous souhaitez rendre vos données accessibles ou non, et les automatiser autant que possible

La mise en pratique :

- Déposer vos données dans un **entrepôt interrogeable** par internet
- Choisir un entrepôt qui intègre un **contrôle d'accès** (si nécessaire)
- Détailler les **conditions d'accès** à chacun de vos jeux de données
- Si des **programmes ou logiciels** sont indispensables à l'utilisation de vos données, les **fournir** ou expliquer comment récupérer ces logiciels
- Même si vos données ne sont pas accessibles, faire en sorte que leur description précise et les **informations bibliographiques** les concernant le soient
 - Ces informations sont à préciser dans les **métadonnées** qui décrivent vos données, à créer directement dans un **formulaire au moment du dépôt** dans l'entrepôt, ou en amont du dépôt **au sein d'un fichier à part**

Des données interopérables ?

INTEROPERABLE

6. Savez-vous que les métadonnées qui décrivent vos jeux de données doivent se baser sur des vocabulaires contrôlés (standardisés, structurés sous la forme d'ontologies), autant que possible ? ⓘ

- **L'idée** : faire en sorte que vos données puissent être utilisées conjointement à d'autres jeux de données, ou par d'autres interfaces techniques que celles que vous avez vous-même utilisées
- **La solution** : créer et décrire vos données selon des pratiques très répandues ou très bien documentées

La mise en pratique :

- Utiliser des **formats de fichiers ouverts**, non-propriétaires, si possible normés:
- Décrire vos données en **termes non-ambigus** et employés par la communauté la plus vaste possible (ex.: un identifiant-chercheur plutôt qu'un nom-prénom)
- Associer à vos données des mots-clés qui emploient un **vocabulaire de description courant** dans ma discipline
- Si j'ai besoin d'employer des termes très spécifiques pour décrire vos données, proposer des **équivalents** dans des vocabulaires plus répandus ou mieux connus

Des données réutilisables ?



REUTILISABLE

7. Savez-vous que vos métadonnées doivent inclure des informations relatives à la provenance des données collectées et/ou générées ? ⓘ
8. Savez-vous que les métadonnées décrivant vos données doivent suivre des standards reconnus par la communauté scientifique dans votre domaine de recherche ? ⓘ
9. Savez-vous que vos données doivent être déposées de préférence dans un format de fichier ouvert (pour faciliter leur réutilisation) qui soit pris en charge par votre entrepôt, pour une conservation à long terme. ⓘ
10. Savez-vous que la conservation de vos jeux de données à long terme exige l'intervention de professionnels de la gestion et de la conservation des données ? ⓘ

- **L'idée** : indiquer très précisément comment et par qui vos données pourront être réutilisées
- **La solution** : fournir ces informations de manière à ce qu'elles soient compréhensibles à la fois par un individu et par une machine

La mise en pratique :

- Choisir une **licence de diffusion** parmi celles recommandées par votre institution, votre communauté disciplinaire ou le gouvernement
- Intégrer ces informations de licence dans les **métadonnées** qui décrivent vos données (à préciser au moment du dépôt dans l'entrepôt)
- Préciser la façon dont vous souhaitez être crédité(e) par les réutilisateurs de vos données, en fournissant par exemple un **modèle de citation**

Les principes CARE

#BeFAIRandCARE



FAIR Guiding Principles for scientific data management and stewardship can be accessed here

<https://www.gida-global.org/care>

Voir aussi « [The CARE data principles for Indigenous data governance](#) », Data Science Journal, 2020

Le référentiel Green Data

10 BONNES PRATIQUES
POUR LIMITER L'IMPACT ENVIRONNEMENTAL DE L'OUVERTURE DES DONNÉES

CPENDATA FRANCE

Faire moins, faire autrement voire ne pas faire

Le numérique, dont l'ouverture des données, sa mise à disposition et ses services associés, sont sources d'une empreinte environnementale non négligeable. Afin de réduire cette externalité négative, ces bonnes pratiques visent à la sobriété des données.

Autrement dit, envisager une forme de soutenabilité, limiter le remplacement et l'empilement de technologies, ou encore l'effet rebond. C'est pourquoi, il convient avant tout et à tout moment, de questionner le besoin pour évaluer collectivement la pertinence à produire, publier et maintenir à chaud certaines données.

www.opendatafrance.net
v1 - septembre 2022

- Engager et conduire une démarche "donnée ouverte et responsable" pour mieux coordonner et diffuser ces pratiques en interne
- Prioriser la publication de données selon leur intérêt (démocratique, environnemental, économique, social...) et faire des choix pragmatiques
- Standardiser les jeux de données pour garantir leur interopérabilité et utilité
- Documenter précisément les métadonnées pour faciliter la gestion et la réutilisation des données
- Proposer une granularité et emprise temporelle et géographique adaptée pour limiter les flux et les téléchargements de données non nécessaires
- Réduire le volume unitaire des données pour limiter le stockage des données et les flux d'échange
- Proposer un accès aux données par API pour faciliter la récupération de données dynamiques et réduire les données transmises
- Faciliter la découvrabilité des jeux de données et limiter leur duplication
- Mettre en place une politique d'archivage pour garantir la pérennité des données
- Contrôler l'hébergement des données grâce au Code de conduite européen pour les centres de données

<https://opendatafrance.gitbook.io/greendata-pour-un-impact-maitrise-des-donnees/greendata/livrables-greendata>

1.
Documenter

2. Analyser,
traiter

3. Stocker,
sauvegarder,
archiver

4. Diffuser,
partager,
publier

1. Documenter



©Pixabay

À l'échelle des données et des jeux de données

- Métadonnées
- Fichiers *read me*

À l'échelle du projet

- Cahiers de laboratoire
- Plan de gestion de données

Les métadonnées

Descriptives

Description du contenu intellectuel

- Titre
- Résumé
- Mots-clés
- Mode d'obtention

Caractéristiques techniques

- Format
- Logiciel utilisé
- Version
- Taille
- Etc.

Droits liés

- Producteur
- Droits d'accès
- Droits de réutilisation
- Etc.

Relations entre données

- A pour partie...
- Est une partie de...
- Est une version de...

administratives

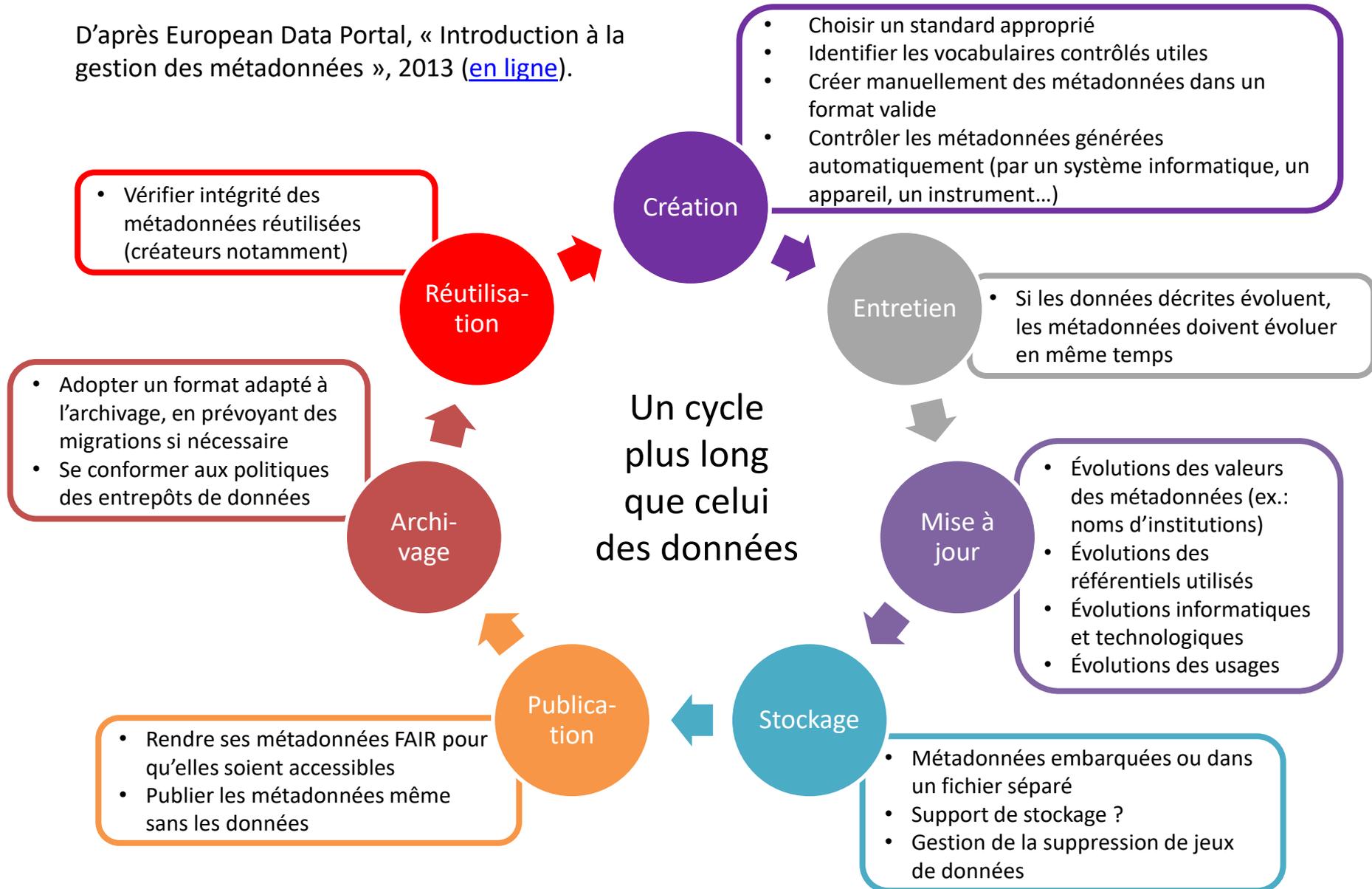
structurelles

- ✓ Pour soi
- ✓ Pour les autres
- ✓ Pour être lues par des machines

- Embarquées
- Externes

Le cycle de vie des métadonnées

D'après European Data Portal, « Introduction à la gestion des métadonnées », 2013 ([en ligne](#)).



Renseigner les métadonnées

Leur structuration varie en fonction...

- Du type de ressources
- Du domaine scientifique
- De la communauté
- De l'entrepôt

Voir [Digital Curation Center](#)

[Les métadonnées Datacite](#)

- [DataCite Metadata Properties](#)
 - [Overview](#)
 - [1. Identifier](#)
 - [2. Creator](#)
 - [3. Title](#)
 - [4. Publisher](#)
 - [5. PublicationYear](#)
 - [6. Subject](#)
 - [7. Contributor](#)
 - [8. Date](#)
 - [9. Language](#)
 - [10. ResourceType](#)
 - [11. AlternateIdentifier](#)
 - [12. RelatedIdentifier](#)
 - [13. Size](#)
 - [14. Format](#)
 - [15. Version](#)
 - [16. Rights](#)
 - [17. Description](#)
 - [18. GeoLocation](#)
 - [19. FundingReference](#)
 - [20. RelatedItem](#)

General Policies v1.0



Metadata types and sources: All metadata is stored internally in JSON-format according to a defined [JSON schema](#). Metadata is exported in several standard formats such as MARCXML, Dublin Core, and DataCite Metadata Schema (according to the [OpenAIRE Guidelines](#)).

<https://about.zenodo.org/policies/>

Les fichiers ReadMe

Contient des informations sur les règles de nommage et d'organisation :

- le contenu d'un ensemble de fichiers
- le contenu d'un fichier (entêtes des colonnes...)
- les logiciels ou codes informatiques nécessaires pour les lire
- précautions à prendre pour la réutilisation
- la personne à contacter pour plus d'informations

Modèles :

- [Recherche data gouv \(2023\)](#)

- Cornell University : W. Kozlowski, *Guidelines for writing "readme" style metadata*, 2014 ([en ligne](#))

- 4TU Centre for Research Data : *Guidelines for creating a README file*, 2017 ([en ligne](#))

Un aperçu :

```
<Les textes d'aide sont écrits entre chevrons et sont destinés à être supprimés avant toute sauvegarde>
<Un README : Pourquoi ?

***La documentation d'un jeu de données doit être suffisante pour permettre à n'importe quel réutilisateur de
comprendre et d'évaluer sa qualité. Le README fournit des informations complémentaires et accessibles lorsqu'elles
ne sont pas déjà mises à disposition dans les métadonnées du jeu de données, dans les métadonnées des fichiers,
et/ou dans des fichiers associés, ou des fichiers accessibles à long terme sur des services d'hébergement (entrepôt
de fichiers ou publication). Dans ce dernier cas, nous vous prions d'inclure les URLs des documents en question ou
leurs références***>

<Privilégier les formats text document (.txt), ou markdown (.md)>

Modèle de Fichier RDG README --- Général --- Version: 0.1 (2022-11-22)

Ce fichier README a été généré le [YYYY-MM-DD] par [NAME].

Dernière mise-à-jour le : [YYY-MM-DD].

# INFORMATIONS GENERALES

## Titre du jeu de données :

## DOI: # INFORMATIONS SPECIFIQUES AUX DONNEES POUR : [NOM DU FICHIER]

## Adresse de contact : <Le cas échéant, reproduire cette section pour chaque dossier ou fichier.
Les éléments se répétant peuvent être expliqués dans une section initiale commune.>

<Ci-après suit une liste d'é pertinence de certains dépen >Pour les données tabulaires, fournir un dictionnaire des données/manual de codage contenant les informations
suivantes :>
<***Supprimer toute section ## Liste des variables/entêtes de colonne :

Pour chaque nom de variable ou entête de colonne, indiquer :

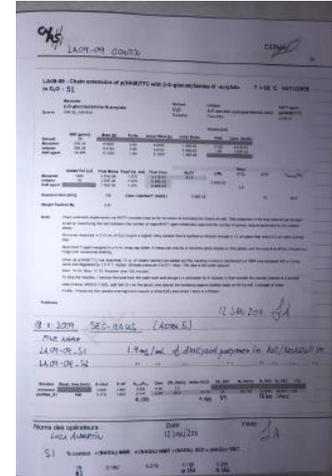
-- le nom complet de la variable sous forme "lisible par les humains" ;
-- la description de la variable ;
-- unité de mesure, si applicable ;
-- séparateur décimal *i.e.* virgule ou point, si applicable ;
-- valeurs autorisées : liste ou plage de valeurs, ou domaine ;
-- format, si applicable, e.g. date>

## Code des valeurs manquantes :
<Définir les codes ou symboles utilisés pour les valeurs manquantes.>

## Informations additionnelles :
<Toute information que vous jugez utile pour mieux comprendre le fichier>
```

Les cahiers de laboratoire

- Outil d'enregistrement
- Outil de traçabilité
- Outil scientifique
- Outil juridique et contractuel



De la version nationale papier (2007)...

... à des solutions électroniques ([Labguru](#) INSERM, [ElabFTW](#) CNRS)

Voir aussi les travaux du [GT CLE](#) du COSO et le [comparatif](#) Dataacc)

Le plan de gestion de données (PGD) Data management plan (DMP)

Un document de pilotage

Il décrit la façon dont les données produites seront obtenues, documentées, utilisées, gérées et disséminées, pendant la phase de recherche et une fois le projet terminé.

Un document évolutif

- À compléter au fur et à mesure du projet
- À envoyer aux financeurs à différents stades du projet



The data lifecycle - Mushonz

Exigences minimales pour les plans de gestion de données



- Description des données et collecte ou réutilisation des données existantes
- Documentation et qualité des données
- Exigences légales et éthiques, codes de conduite
- Stockage et sauvegarde pendant le processus de recherche
- Partage des données et conservation à long terme
- Responsabilités et ressources en matière de gestion des données

Guide pratique pour une harmonisation internationale
de la gestion des données de recherche
Science Europe (2018, v2 [2021](#)) Voir la traduction en [français](#)

Quelques modèles

Sylvie Cocard (INRAE), [Rédiger un plan de gestion de données](#), 2022

Verburg, M., & Grootveld, M. (2022, février 22). **FAIR-Aware Additional guidance to the Science Europe DMP assessment rubric.**
<https://doi.org/10.5281/zenodo.6088215>



- Informations sur le PGD
- Informations sur le projet ou la structure
- Présentation générale des données
- Organisation et documentation des données
- Droits de propriété intellectuelle
- Confidentialité
- Stockage et sécurité des données
- Partage des données à l'issue du projet
- Archivage et conservation des données après la fin du projet

[Trame projet v2](#)
43 questions
[Trame structure](#)
39 questions



- Description des données et collecte ou réutilisation de 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

[ANR Modèle de PGD](#)
15 questions
[Guide en français](#)



- Résumé descriptif des données
- Données FAIR
- Allocation de ressources
- Sécurité des données
- Aspects éthiques
- Autres



- Data Summary
- FAIR data
- Other research outputs
- Allocation of resources
- Data security
- Ethics
- Other issues

[H2020 FAIR DMP \(v.f.\)](#)
9 questions
[Horizon Europe Data](#)

this template is recommended
for Horizon Europe beneficiaries



- Description des données et collecte ou réutilisation de données existantes
- Documentation et qualité des données
- Exigences légales et éthiques, codes de conduite
- Traitement et analyse des données
- Stockage et sauvegarde pendant le processus de recherche
- Partage des données et conservation à long terme

[Science Europe](#)
[« modèle structuré »](#)
(+ [version anglaise](#)) 21

Les outils de rédaction de PGD

Exemples d'outils d'aide à la rédaction de DMP

Ces outils se distinguent par leurs fonctionnalités spécifiques, leurs modèles de DMP, leur convivialité, leur couverture des exigences de financement, leur personnalisation, leur approche visuelle ou leur intégration avec d'autres fonctionnalités de gestion de projet.

Le choix de l'outil dépendra des préférences individuelles, des besoins spécifiques du projet de recherche et des exigences des organismes de financement ou des institutions de recherche auxquels les chercheurs sont soumis.



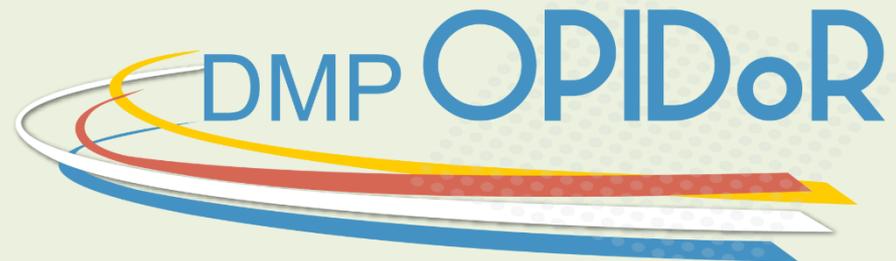
[Extrait de *La question des données en bio-informatique*, cours Doranum-Callisto](#)



Des modèles et des exemples de votre institution ?

- 1) Rendez-vous sur DMP Opidor > Modèles de PGD. Cherchez s'il existe un modèle de DMP propre à votre institution.
- 2) Rendez-vous sur DMP Opidor > PGD Publics. Existe-t-il des exemples issus de votre institution ? De votre laboratoire ?

<https://dmp.opidor.fr/>



En l'absence de modèle spécifique, utilisez les recommandations des établissements et organismes dans DMP Opidor

<https://opidor.fr/dmp-opidor-des-recommandations-plus-faciles-a-trouver-et-plus-lisibles/>



Cliquez ici pour sélectionner les recommandations appliquées à votre plan

Vous trouverez ci-dessous la liste des organismes proposant des recommandations et des conseils pour vous guider dans la rédaction de votre plan tout en respectant leur politique de gestion de données.

Vous avez la possibilité de sélectionner 6 organismes au maximum. Puis cliquez pour enregistrer votre sélection.

CY Cergy Paris Université

CY Cergy Paris Université (FR)

CY Cergy Paris University (EN)

AgroParisTech - Institut des sciences et industries du vivant et de l'environnement

AgroParisTech

CNRS

Service Protection des Données - CNRS (FR)

CNRS général (FR)

CNRS general (EN)

Data Protection Office - CNRS (EN)

IMT Mines Alès - Ecole Mines-Télécom



Conseils pour la rédaction de PGD

1. Commencer tôt
2. Réfléchir à la réutilisation
3. Vérifier les politiques en vigueur
4. Se faire aider
5. Penser large
6. S'inspirer du travail d'autrui
7. Être précis là où c'est nécessaire
8. Être concret
9. Oser avouer qu'on ne sait pas (encore)
10. Mettre à jour

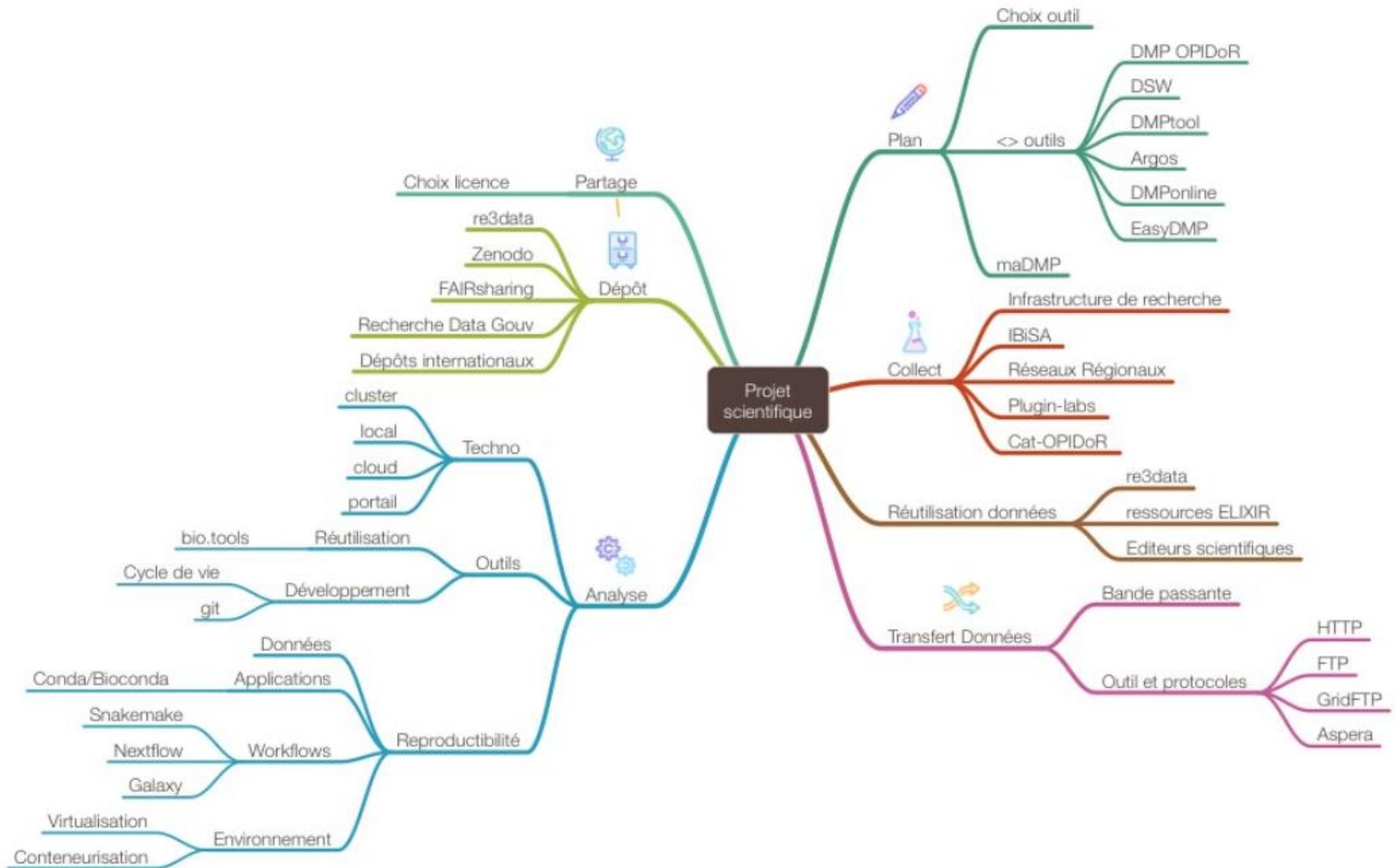


[10 Tips for Writing a Data Management Plan](#), 2018

2. Analyser, traiter

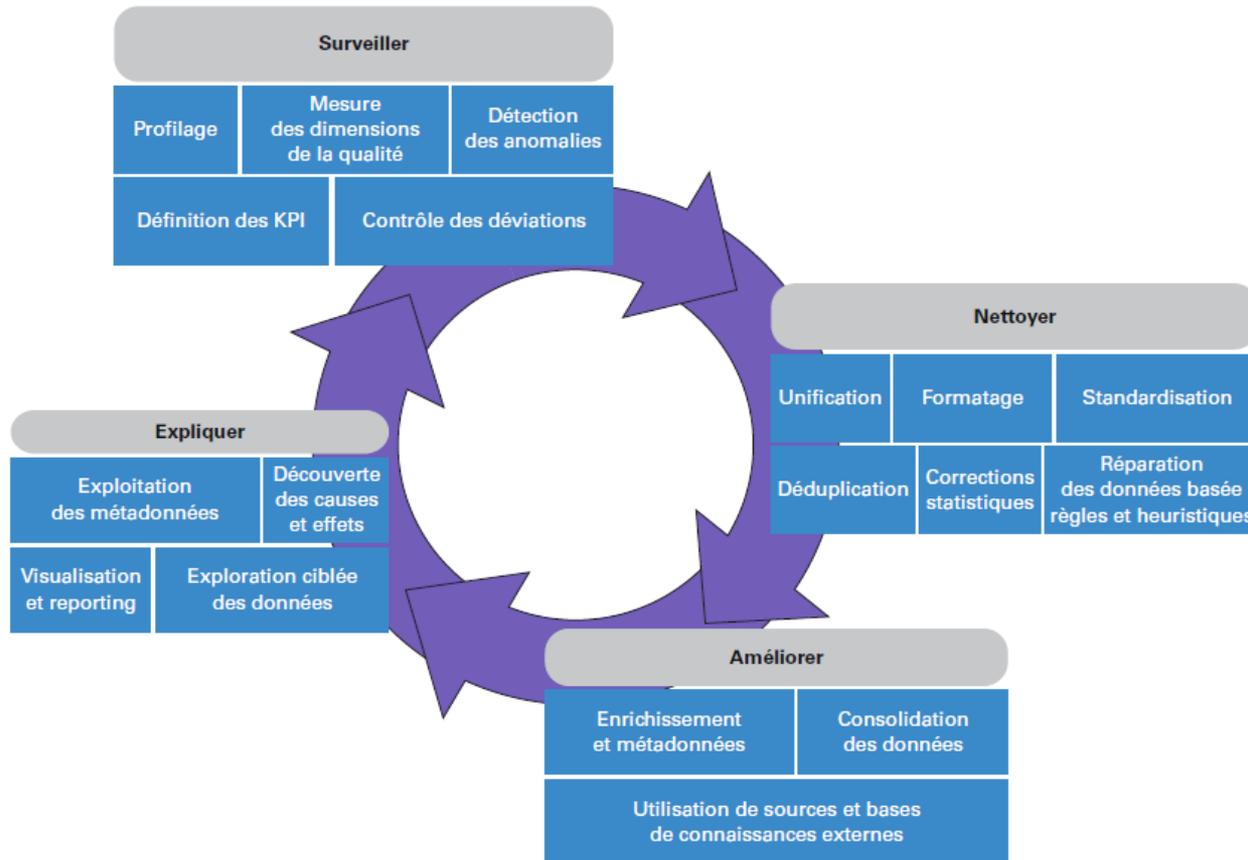


Wikicommons



Extrait de *La question des données en bio-informatique*, cours Doranum-Callisto

La qualité des données



Laure Berti-Equille, [Qualité des données, Fiche H3700V2](#), Techniques de l'ingénieur

Voir aussi Luc Decker, [La démarche Qualité des données](#), IRD

Best Practices

The image displays two Excel spreadsheets side-by-side, illustrating data entry best practices. Both spreadsheets use a single table structure with consistent column headers and data types.

Top Spreadsheet (data.xls):

1	Site	Date	Plot	Species	Weight	Acult	Rodent Trapping
2	DeepWell	2/13/2010		1 DIPO	12.1	j	3/15/2010
3	Deep Well	Feb-10		2 Pero	13.22	j	
4	nioSalado	2/13/2010	1a	pero	16	N	
5	nioStadu	*	1*	CleGap	18.92	gut away	
6				Mean1	15.06		

Bottom Spreadsheet (SEV_SmallMammalData_v.5.25.2010.xls):

1	Date	Site	Plot	Species	Weight	Adult	Comments
2	2/5/2010	Deep Well		1 DIPO	13.2	y	
3	2/4/2010	Deep Well		1 CLEGAP	11.6	j	
4	2/5/2010	Rio Salado		1 DIPO	14.2	y	
5	2/5/2010	Rio Salado		2 PERO	10.1	y	
6	3/15/2010	Deep Well		1 DIPO	15.2	y	plot burned
7	3/15/2010	Deep Well		2 DIPO	21.7	y	pregnant
8	3/15/2010	Rio Salado		1 CLEGAP	16.2	j	

- Columns of data are consistent: only numbers, dates, or text
- Consistent Names, Codes, Formats (date) used in each column
- Data are all in one table, which is much easier for a statistical program to work with than multiple small tables which each require human intervention



Data Entry and Manipulation

Source : [DataONE Education Module: Data Entry and Manipulation](#).

Voir aussi

- [Ten Simple Rules for large-scale data processing](#), PLoS Computational biology, 2022

3. Stocker, sauvegarder, archiver



Kues1 on Freepik

STOCKAGE ET SAUVEGARDE



Serveur sécurisé de l'institution

Espace collaboratif de travail



Le chercheur



L'équipe de recherche

PARTAGE



Entrepôt généraliste ou disciplinaire



L'équipe de recherche



Autre équipe de recherche (même domaine)

Autre équipe de recherche (autre domaine)

ARCHIVAGE PÉRENNE



Plateforme d'archivage (ex : CINES)



L'équipe de recherche



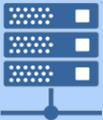
Autre équipe de recherche (même domaine)



Autre équipe de recherche (autre domaine)

Introduction aux données de recherche, aux principes FAIR et aux PGD – [INIST 2019](#)
[Doranum](#) : stockage, partage, archivage, quelle différence ?

Comparatif des supports de stockage

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 Ordinateur professionnel	★★☆☆ Sujet au piratage informatique, aux détériorations et pannes	★☆☆☆ Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...)	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter les données confidentielles et sensibles
 Support externe	★☆☆☆ - Sujet au vol, à la perte du support - Durée de vie limitée (dégradation du matériel)	★★★★★ Facilement transportable, il permet de transférer les données vers un autre ordinateur	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles
 Serveur institutionnel	★★★★★ Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)	★★☆☆ La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures	★★★☆☆ Coût assez important mais pas forcément répercuté sur l'usager	- Pour un stockage plus pérenne - Adapté pour le stockage de données sensibles et des versions « stables » de vos données - Toutes les institutions ne proposent pas ce service
 Serveur Cloud	★★☆☆ On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent	★★★★★ Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage	★★★☆☆ Payant à partir d'une certaine limite de stockage	- Pour un partage avec des personnes externes à l'institution - Ne pas y mettre de données sensibles ou confidentielles - Pas de contrôle sur la procédure de sauvegarde des données

Comparatif des différents supports de stockage, [Doranum](#)

Les coûts du stockage des données

- **Coût financier**

- Le stockage a un coût qu'il ne faut pas négliger
- Par sa nature, contrairement au calcul par exemple, le stockage n'est pas vraiment mutualisable
- On peut optimiser en partageant une plateforme et les coûts humains associés
- Plus un stockage est performant (on accède rapidement aux données), plus il est cher
- Exemple de coût :
 - disque dur externe 2 To : 35 à 70 € environ
 - 1 To sur la plateforme Summer UGA : de 25 € à 80 € annuel en fonction du type de stockage
 - 1 To sur la plateforme Microsoft Azur : de 50 à 180 € environ par mois en fonction du type de stockage

- **Coût environnemental**

- Le stockage, et encore plus l'archivage, a un coût environnemental non négligeable
- Exemple : pour la plateforme Bettik à l'UGA (stockage performant pour le calcul) : 1 Go.an émet 12 g CO₂e – en 2019 : 1.4 Po.an → 17 t CO₂e

Séminaire Stockage des données de la recherche, UGA, 2021 (sur [Dorandum](#))

Voir aussi

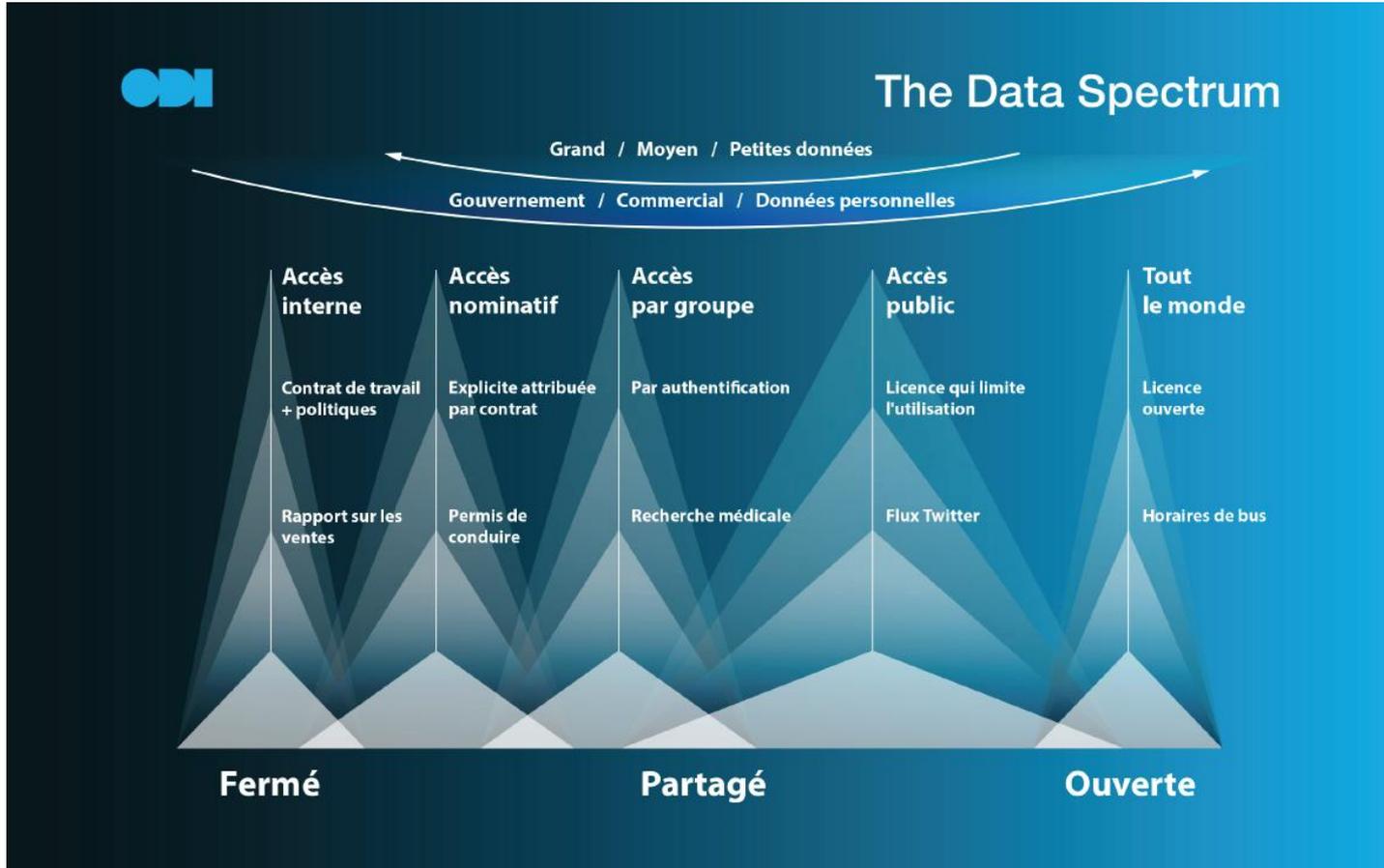
- Estimating Costs : <https://www.openaire.eu/estimating-costs-rdm-tool>
- EPFL Library, Cost Calculator for Data Management : <https://costcalc.epfl.ch/>

Les durées de conservation des archives > données

C. Laplante, « Les données de recherche : comment évaluer pour mieux conserver? », [Convergence](#), 2021
 Voir aussi les guides sur Doranum : [documents à archiver](#) ; [documents à trier](#) ; [documents à détruire](#)

Typologies documentaires	DUA Les durées sont exprimées en années	Sort final	Observations
II.1.2.4.3. Validation des données			
Fichiers de résultats, mails ou courriers de validation.			
II.1.2.4.3.1. Traitement des données par les organismes/ administrations fournisseurs de données	10 ans		
II.1.2.4.3.2. Données brutes et saisie de celles-ci dans la base de données	DUA identique aux données travaillées correspondantes	C	<i>La trace d'e ces corrections (à l'exclusion des erreurs de saisie) doit être conservée au même titre que les versions initiale et finale de la base de données afin de conserver l'historique de la construction de la base de données finale à l'origine des résultats de l'étude.</i>
II.1.2.4.4. Données travaillées (bases regroupant les données d'enquête)			<i>Pour chaque étude, conserver les données épidémiologiques ou d'enquête pouvant justifier les résultats de l'étude. Parmi ces données, ne conserver que les données les plus complètes, soit au stade primaire, soit les données travaillées nominatives ou anonymisées.</i>
Bases de données directement ou indirectement nominatives : version initiale et version finale.	Durée de conservation autorisée par la CNIL	C	
Bases de données anonymisées : version initiale et version finale.	Selon les besoins de l'organisme	C	<i>Les versions intermédiaires des bases de données peuvent être supprimées. Seules la première version et la version finale sont conservées.</i>
II.1.2.5. Patrimoine de données brutes rassemblées dans de grandes bases de données et plateformes de recherche			
II.1.2.5.1. Données techniques			
Documentation sur la reprise de données, description du contenu et de la structure des données, description technique et fonctionnelle de la base de données, dossier d'exploitation et d'évolution de la base.	Durée de la recherche + 10 ans	T	<i>Conserver les données brutes et les versions initiales et finales des bases de données.</i>
II.1.2.5.2. Données d'enregistrement			
Registre d'enregistrement des échantillons, souches et prélèvements.	Même durée que le dossier d'analyse / de recherche / d'étude associée	D	
II.1.2.5.3. Analyse des données travaillées			
Programmes d'exploitation/de calcul/d'analyses des bases de données, fichiers de résultats, listing, tables de données, fichiers log.	2 ans	D	
II.1.2.6. Typologies relatives à la fabrication de substances en interne			

4. Diffuser, partager, publier des données



[The Data Spectrum](#), ODI

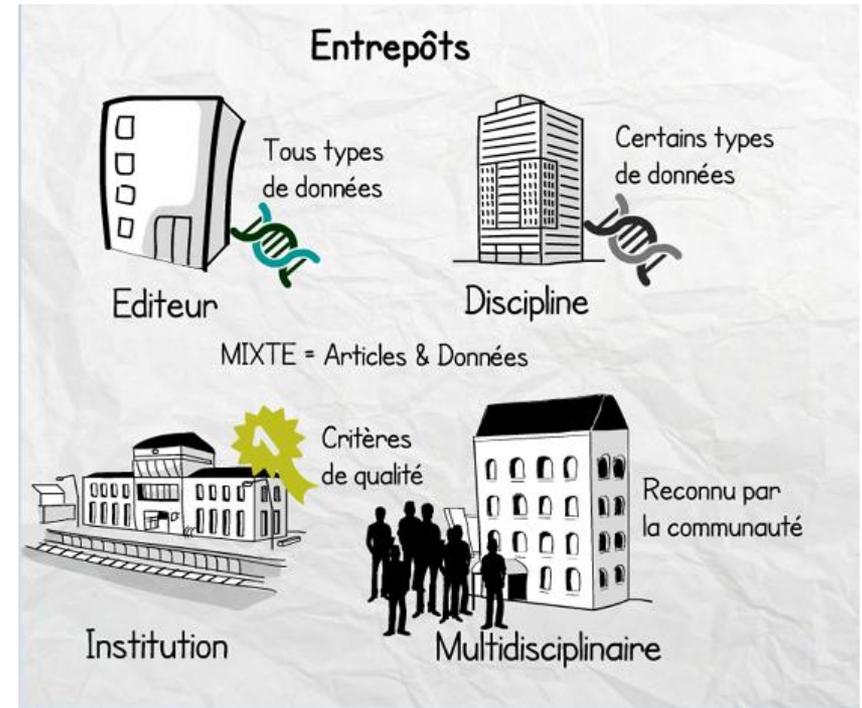
Quelles données partager ?

La pyramide de publications des données



Les entrepôts de données de recherche

« Les entrepôts de données sont des **plateformes web** sur lesquelles les chercheurs peuvent **déposer et rechercher des données scientifiques**. Ils proposent des **services spécifiques à l'activité de recherche** (description des jeux de données, choix des conditions d'accès, attribution de licence...) et offrent une garantie de **sécurité**. La plupart présente un système de recherche détaillé avec de nombreux filtres. »



Les entrepôts de données de recherche, [Cours Doranum sur Callisto](#), 2021

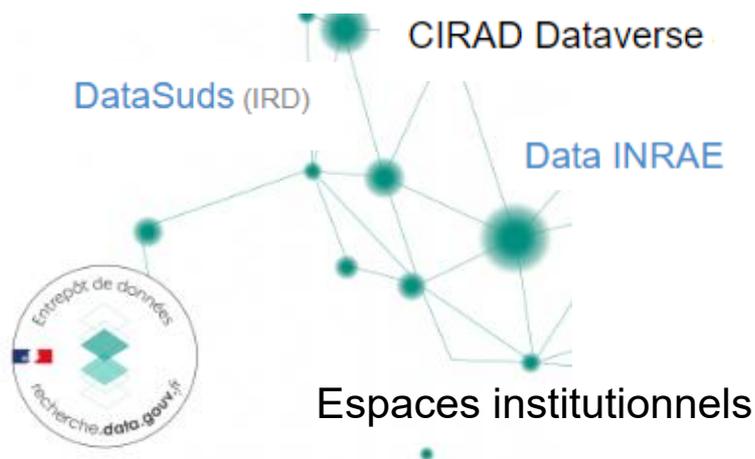
multidisciplinaires

disciplinaires



International

France



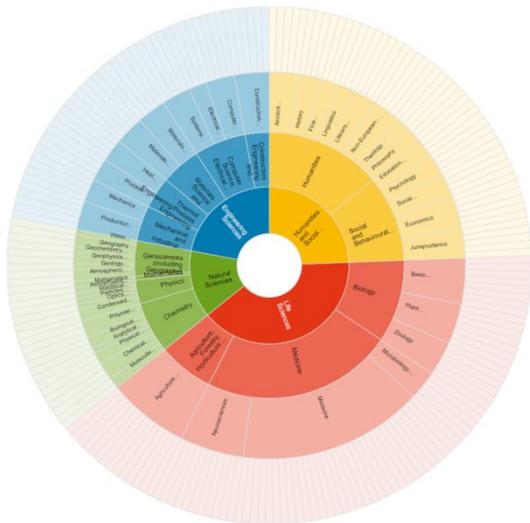
data.sciencespo



Annuaire et listes d'entrepôts



[re3data](https://re3data.org)



[OAD Data repositories](#)

Contents [hide]

- 1 Archaeology
- 2 Astronomy
- 3 Biology
- 4 Chemistry
- 5 Computer Science
- 6 Energy
- 7 Engineering
- 8 Environmental sciences
- 9 Geology
- 10 Geosciences and geospatial data
- 11 Linguistics
- 12 Marine sciences
- 13 Medicine



[Cat Opidor > entrepôts de données](#)

Vous pouvez utiliser les filtres afin d'affiner votre recherche ? :

Type d'entrepôt			
<input type="checkbox"/> Disciplinaire	<input type="checkbox"/> Institutionnel	<input type="checkbox"/> Lié à un projet ou un équipement etc	<input type="checkbox"/> Multidisciplinaire
<input type="checkbox"/> National			
Domaine scientifique			
<input type="checkbox"/> Sciences & Technologies	<input type="checkbox"/> Sciences Humaines & Sociales	<input type="checkbox"/> Vie & Santé	
Sciences Humaines & Sociales			
Sciences & Technologies			
Vie & Santé			
Thématique/Mots clés			
Selectionner une valeur de filtre			
Label ou certification ?			
<input checked="" type="radio"/> OU <input type="radio"/> ET			
<input type="checkbox"/> Centre de référence thématique	<input type="checkbox"/> CoreTrustSeal	<input type="checkbox"/> CoRTecS	<input type="checkbox"/> Co
<input type="checkbox"/> HDS	<input type="checkbox"/> ISO 27001	<input type="checkbox"/> ISO 27701	<input type="checkbox"/> NC
<input type="checkbox"/> SNDS	<input type="checkbox"/> SNO	<input type="checkbox"/> Spatio	<input type="checkbox"/> WT
Attribution d'un identifiant pérenne			

➤ Voir aussi <https://commons.datacite.org/repositories>

Choisir un entrepôt pour ses données

- Est-il dans mon périmètre disciplinaire ?
Recommandé par mes collègues ? Les financeurs ? Mon employeur ?
- Accepte-t-il mes fichiers (format, taille...) ?
- Accepte-t-il mes métadonnées ? Fournit-il un identifiant pérenne ?
- Permet-il de modifier mes données et accéder aux anciennes versions ?
- Propose-t-il des licences de diffusion ?
Permet-il de restreindre l'accès ?
- Quelle procédure de validation du dépôt ?
Quels coûts éventuels ?
- Est-il certifié / reconnu ? [Core Trust Seal](#) ?
[Principes TRUST](#) ?
- Voir S.-A. Sansone et al., *Data Repository Selection: Criteria That Matter*, Zenodo ([en ligne](#)).

Autant que possible, privilégier un entrepôt (par ordre de préférence) :

1. Spécialisé dans la gestion de son type spécifique de données
2. Spécialisé dans sa discipline
3. Institutionnel
4. Généraliste

➤ [Sélectionner un entrepôt thématique de confiance, COSO, 2024](#)

➤ [Logigramme Recherche Data gouv](#)

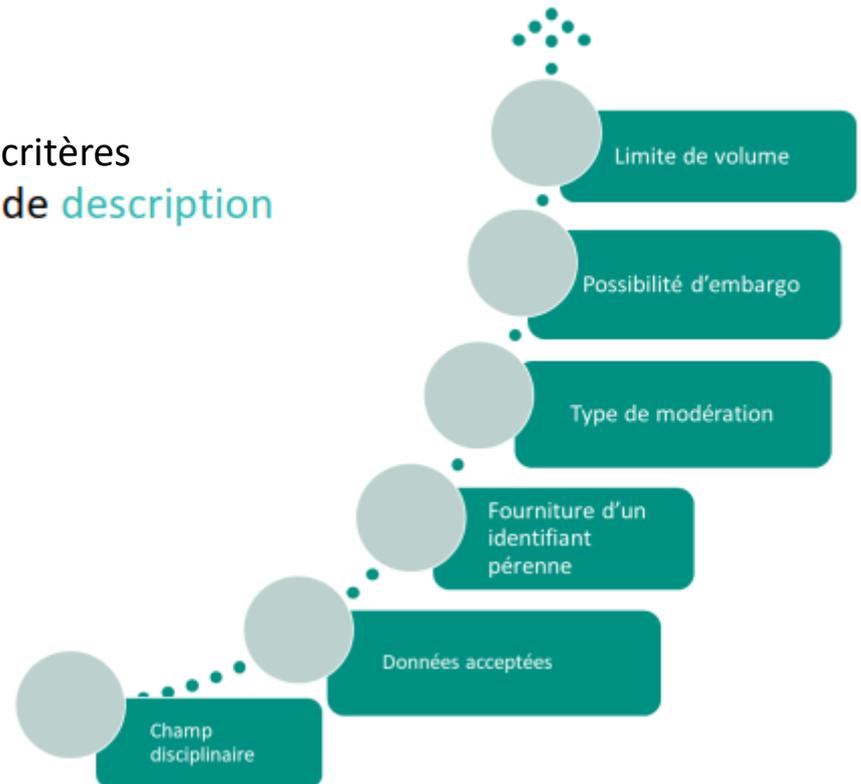


➤ [Sélectionner un entrepôt thématique de confiance, COSO, 2024](#)

critères
d'exclusion



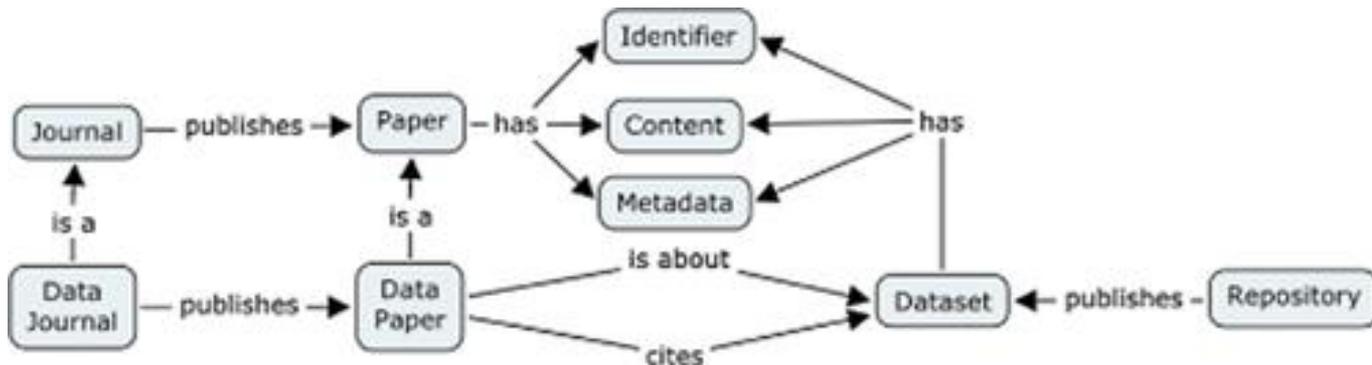
critères
de description



[Liste d'entrepôts de confiance à retrouver sur Recherche Data Gouv et sous format xls \(avril 2025\)](#)

Les *data papers*

Data paper : publication scientifique décrivant un ensemble de données particulier ou un groupe d'ensembles de données, publié sous la forme d'un article révisé par des pairs dans une revue scientifique ([Typologie des documents](#) - HAL)



Candela, L., Castelli, D., Manghi, P. and Tani, A. (2015), Data Journals: A Survey. J Assn Inf Sci Tec, 66: 1747-1762. <https://doi.org/10.1002/asi.23358>



Data in Brief (Elsevier) [Guide for authors](#)

Article Structure

We ask all authors to use our Data in Brief template when preparing their manuscript. All Data in Brief articles must contain the following sections:

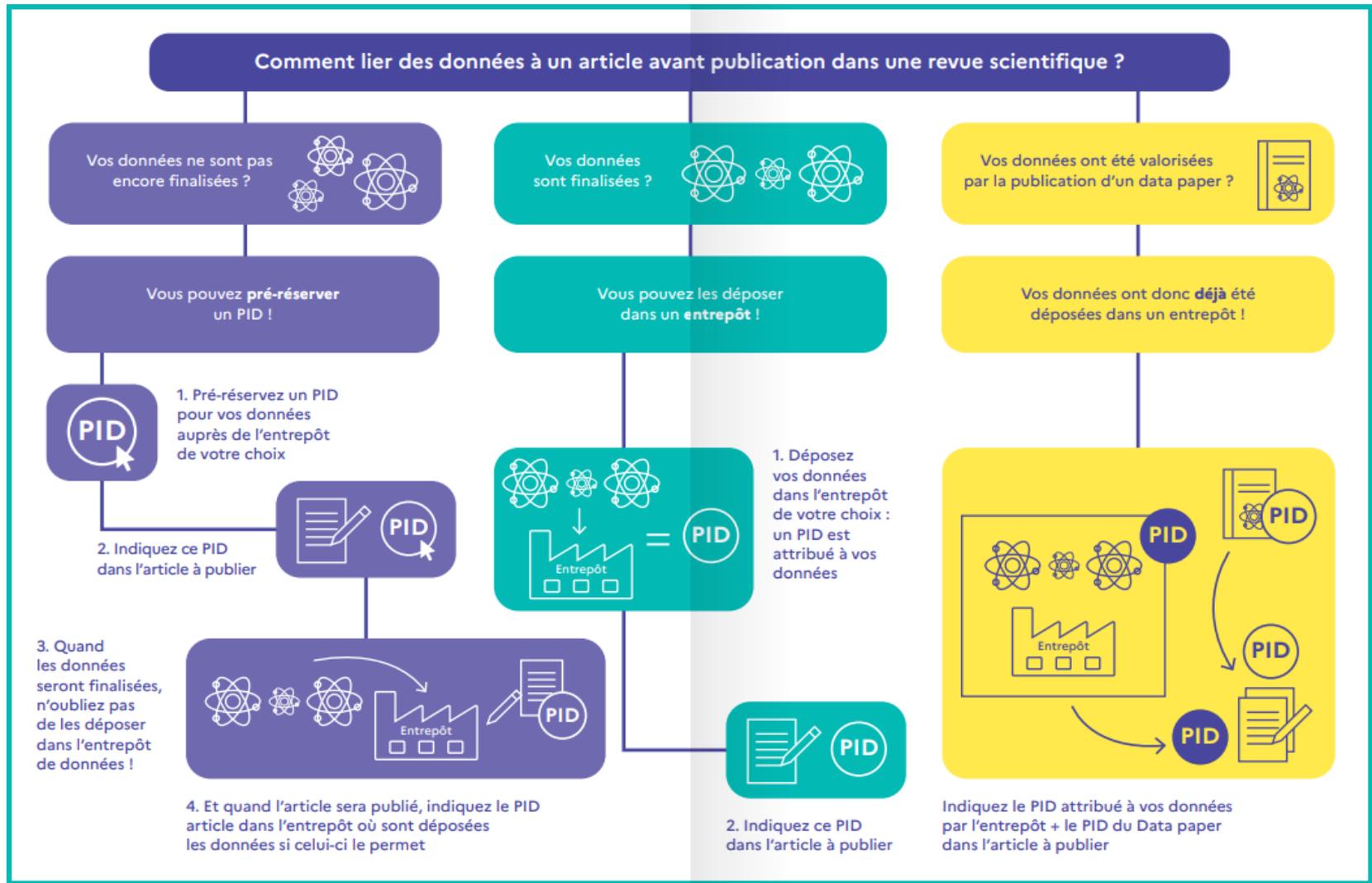
- 1 Title, authors and author affiliations
- 2 Abstract
- 3 A completed Specifications Table
- 4 3-5 bullet points that explain why the data is valuable to other researchers
- 5 A direct link to the deposited data
- 6 A complete description of the experimental design and materials and methods
- 7 References

Quelques listes de revues publiant des data papers

- Publier un data paper : où et comment ? [DATACC](#)
- Publier un data paper, [CoopIST-Cirad](#)
- J. Schöpfel, D. J. Farace, H. Prost, A. Zane. “Data papers as a new form of knowledge organization in the field of research data”, ISKO France, Oct 2019, [halshs-02284548](#)

Voir aussi [Data & Corpus](#) : revue spécialisée dans les data papers et données en SHS

Partager les données liées aux publications scientifiques



Les licences associables aux données de recherche

Un principe à faire respecter : la paternité / l'attribution (BY)



[Etalab](#) : toute utilisation est possible tant que la **paternité** est mentionnée
(équivalente à CC-BY)



[ODC-ODbL](#) : Open Database License : partage, réutilisation et adaptation
sous condition d'**attribution** et de licence à l'**identique**
(équivalente à CC-BY-SA)

- La [fiche pratique Dataactivist](#) ou les [fiches pratiques du GTSO](#)
- L'aide à la décision : <https://choosealicense.com/non-software/>

Pour conclure

Le Minimum à FAIR

Cliquer sur les illustrations pour avoir plus d'informations



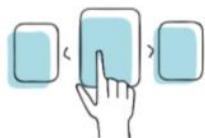
Publié le : 21/11/2024



Centre de ressources
supports pédagogiques, e-formations
Recherche Data Govv



Choisir



un **entrepôt de données** adapté dans lequel déposer ses données

Vérifier



que l'entrepôt dans lequel les données sont déposées leur attribue un **identifiant pérenne**

Documenter



un maximum les **données** et utiliser des vocabulaires contrôlés de sa communauté disciplinaire

Définir



clairement les **droits d'accès** aux données au moment du partage

Privilégier



des **formats de fichiers ouverts**, non-propriétaires

Attribuer



une **licence de réutilisation** à ses jeux de données

Mentionner



les **productions liées aux données** (publications, thèses, autres jeux de données, codes et logiciels...)

[Le minimum à FAIR](#) (Doranum, 2024)

le minimum à faire pour rendre des données FAIR

Communauté scientifique

Archiviste



Ingénieur projet



Informaticien



Service juridique



Professionnel de l'IST



Délégué à la Protection des Données





Recherche Data Gov

« Un écosystème au service du partage et de l'ouverture des données de la recherche »

- *Ateliers de la donnée*
- *Centres de ressources*
- *Centres de références thématiques*



[Lien vers l'évaluation](#)

15 questions
Validé avec 75% de réussite
(2 tentatives)

[Open Badge Urfist - Introduction aux données de la recherche](#)

Ressources généralistes

- [Dorandum](#)
 - [Parcours interactif sur les données de la recherche](#) (juin 2020, mis à jour 2023)
- [Guide de bonnes pratiques sur la gestion des données de la recherche](#), Atelier Données CNRS (2023, v2)
- Espace Callisto du [Centre de ressources Compétences – Recherche Data Gouv](#) (2022-)
- C. Borgman, [Qu'est-ce que le travail scientifique des données ?](#), (2015, traduction française 2020)
- [Passeport Données de la recherche](#) (2024)

Ressources disciplinaires

- [Centres de référence thématiques](#)
Recherche Data Gouv

Astronomie et astrophysique



Système Terre et Environnement



Sciences Humaines et Sociales



→ catalogue/entrepôt : <https://data.progedo.fr/>



Huma-Num^{IR*} → catalogue/entrepôt : <https://nakala.fr/>

Biologie-Santé



- Espace Callisto Doranum avec des cours [thématiques et disciplinaires](#) : bioinformatique, informatique théorique, sciences de l'environnement, économie, droit...
- [Dataacc](#) : gestion des données en physique et chimie
- [Guide de bonnes pratiques pour la gestion des données de recherche en bio-imagerie](#)
- Guide du consortium [MASA](#)
- Guide [Données et recherches participatives](#)