

De la gestion des données vers le partage et la réutilisation

Espace institutionnel Entrepôt de données DAT'AMU

Guichet unique de la Donnée du Site Aix - Marseille (GDsAM)

Données de recherches

24 Mai 2024

Contenu de l'intervention

Objectifs :

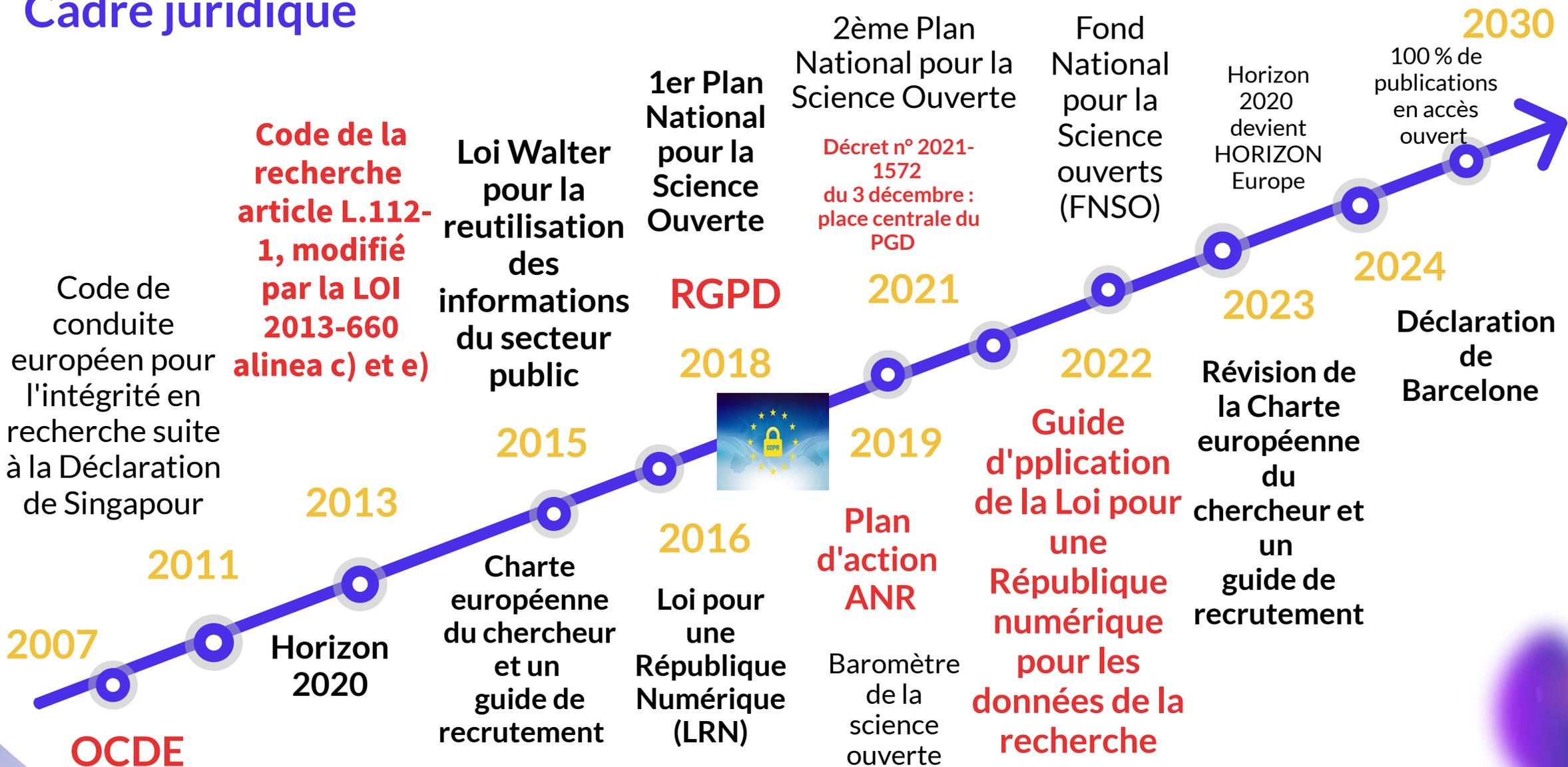
- Ajouter une plus value à vos données.
- Valoriser vos jeux de données grâce à un data paper.

1 Définir ce que sont vos données et modéliser votre processus de recherche →

2 Documenter vos données et métadonnées en suivant les principes FAIR →

3 Lier et partager vos données de recherche avec vos publications. →

Cadre juridique





2022 TROIS GRANDS PRINCIPES à retenir :



"(...) les données produites ou reçues dans le cadre de la recherche publique sont juridiquement considérées comme des « documents administratifs » ou des informations publiques."

"Les données de recherche issues de l'activité courante des établissements de recherche et détenues par eux **sont ainsi soumises, sauf exceptions, aux principes d'ouverture par défaut et de libre réutilisation fixés par le CRPA.**"

"(...) sauf exceptions, les établissements d'enseignement supérieur et de recherche **sont tenus de publier spontanément en ligne et de rendre librement réutilisables les informations publiques qu'elles produisent ou reçoivent (...)**"

Guide
d'application de
la Loi pour une
République
numérique pour
les données de la
recherche

Francine Filoche, Chargée de mission données de la recherche
Département ingénierie documentaire du SCD AMU _ Cellule
Science Ouverte



Que recherchez-vous ?



Charte de transparence sur le recours aux animaux à des fins scientifiques et réglementaires en France

MODÈLE ANIMAL, AIX-MARSEILLE U TRANSPARENCE

MODIFIÉ PAR YACINE AIT MOKHTAR LE MAR 07/05/2024 - 11:01

L'expérimentation animale dans la recherche demeure une indispensable, encadrée par des dispositifs éthiques et règlementaires de protection, au respect et à la bienveillance des animaux utilisés.

La recherche biologique et médicale vise à l'accroissement continu des connaissances scientifiques, le progrès des thérapies en médecine humaine et vétérinaire et une meilleure protection de l'homme, de l'animal et de l'environnement.

La recherche, qu'elle soit fondamentale ou appliquée, requiert des modèles pour progresser. Elle nécessite une approche intégrée qui s'appuie sur la complémentarité des méthodes expérimentales basées sur l'informatique (*in silico*), la biochimie, la culture de cellules et de tissus (*in vitro*) et le recours aux animaux vivants (*in vivo*), ainsi que sur les données cliniques. Les modèles *in silico* et *in vitro* sont très largement utilisés, mais ne sont en mesure de remplacer les modèles animaux que partiellement en l'état actuel des connaissances. Les modèles utilisant des animaux restent donc nécessaires pour mener à bien une recherche de qualité.

Dans le développement et la production des médicaments, qu'ils soient humains ou vétérinaires, le recours aux animaux répond également à des exigences réglementaires de sécurité et d'efficacité. Dans le domaine vétérinaire, les études sont réalisées sur les espèces animales auxquelles les produits sont destinés.

En conformité avec les dispositions légales du Code rural et du Code civil, et au titre de leurs convictions, les signataires de cette Charte reconnaissent que les animaux sont des êtres vivants doués de sensibilité. Le recours aux animaux en recherche scientifique et médicale est encadré par une réglementation spécifique qui les protège, veille à leur bien-être et impose le respect de règles éthiques.

Chaque citoyen a droit à une information complète, claire et exacte sur les raisons et les conditions de l'utilisation d'animaux à des fins scientifiques ou réglementaires, sur le cadre réglementaire de cette utilisation ainsi que sur les progrès scientifiques et médicaux qui en découlent.



Droits et obligations du chercheur

Propriété intellectuelle des données

Recommandations et obligations de diffusion

Communicabilité des données

Accès, sécurité et licences

Intégrité scientifique et éthique

■ PROPRIÉTÉ INTELLECTUELLE DES DONNÉES DE RECHERCHE (1/2)

Règle générale

- Attribution de la propriété intellectuelle des données à l'établissement de tutelle des producteurs de données



Si partenariat

- Nécessité d'établir au préalable un **accord de consortium**
- Cas de collaborations entre secteurs public et privé (déploiement industriel, commercialisation...)
- Cas de collaborations internationales





Droits et obligations du chercheur

Propriété intellectuelle des données

Recommandations et obligations de diffusion

Communicabilité des données

Accès, sécurité et licences

Intégrité scientifique et éthique

■ PROPRIÉTÉ INTELLECTUELLE DES DONNÉES DE RECHERCHE (2/2)

"[...] il existe une différence importante de régime juridique applicable aux œuvres de type écrits scientifiques et aux données. Si l'on entend le terme « données de recherche » au sens strict (c'est-à-dire d'informations générées dans le cadre d'un processus de recherche), alors le chercheur ne sera pas considéré comme un « auteur », car les données ne sont généralement pas en tant que telles des « œuvres » protégeables par le droit d'auteur. Un droit de propriété intellectuelle spécifique existe néanmoins pour la protection des bases de données, mais son fonctionnement est différent de celui du droit d'auteur. Ce droit dit de « producteur de base de données » n'appartient pas à l'origine aux personnes physiques qui réalisent la base, mais à l'entité qui a effectué des « investissements substantiels » pour rendre cette opération possible. Dans la plupart des hypothèses, ce seront donc les établissements de tutelle des chercheurs qui auront la qualité de « producteurs » et posséderont les droits attachés aux bases de données de recherche. Or la loi République numérique a explicitement « neutralisé » le droit des bases de données des administrations pour faire primer le principe de libre réutilisation. Le nouvel article 11 du texte indique ainsi :

Il en résulte que les données produites par les chercheurs sont bien comprises dans le principe d'ouverture par défaut. La situation sera donc très différente de celles des écrits scientifiques et autres créations produites par les chercheurs dans le cadre de leurs activités."

Text button



source

Lionel Maurel.

La réutilisation des données de la recherche après la loi pour une République numérique.

La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques,

Presses Universitaires de Provence, 2018,
9791032001790.

hal-01908766



Sous réserve de droits de propriété intellectuelle détenus par des tiers, les droits des administrations mentionnées au premier alinéa de l'article L. 300-2 du présent code, au titre des articles L. 342-1 et L. 342-2 du code de la propriété intellectuelle [c'est-à-dire le droit de producteur de bases de données], ne peuvent faire obstacle à la réutilisation du contenu des bases de données que ces administrations publient en application du 3° de l'article L. 312-1-1 du présent code.

■ COMMUNICABILITÉ DES DONNÉES

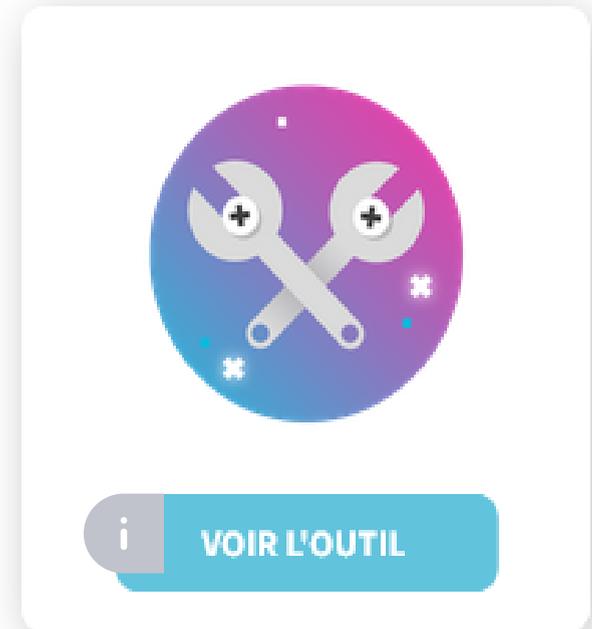
La communicabilité des données peut être conditionnée par :

- la nature ou le type des données
- l'origine des données
- leur(s) utilisation(s).

Elle peut être empêchée temporairement ou définitivement.

Toute restriction doit être mentionnée et expliquée dans le plan de gestion de données.

*Survoler pour voir
les exemples :*



Cliquer pour voir le détail



2eme niveau par champ d'application en fonction du RGPD

Ouverture des données de recherche

Guide d'analyse du cadre juridique en France

Le présent guide est issu des réflexions d'un groupe de travail inter-organismes animé par l'INRA. Il ne prétend pas à l'exhaustivité et est fourni uniquement à titre d'information. Il ne saurait en tout état de cause se substituer aux politiques d'établissements, au respect des dispositions législatives ou réglementaires et au respect de la jurisprudence applicable en la matière. Ce guide peut évoluer.

Membres du groupe de travail : BÉGAUD Nicolas (INRA), CASTETS-BÉNAUD Céline (IUT), CHASSANG Guenther (Inserm), Membre de la Plateforme Genomical Society), DANVANT Marine, TREFF CAFFIN Laurence (Inra), GARDON Nathalie (co-animatrice, INRA), MARTIN Caroline (Agreenium), MARTELLETTI Andrea (stagiaire INRA, M2 droit et Informatique), MENDOZA-CAMINADE Alexandra (IUT), MORELLETTE Nathalie (co-animatrice, INRA), NEIRAC Claire (Erad), avec la participation d'Inra? (Benjamin JEAN, Laure KASSEEM).



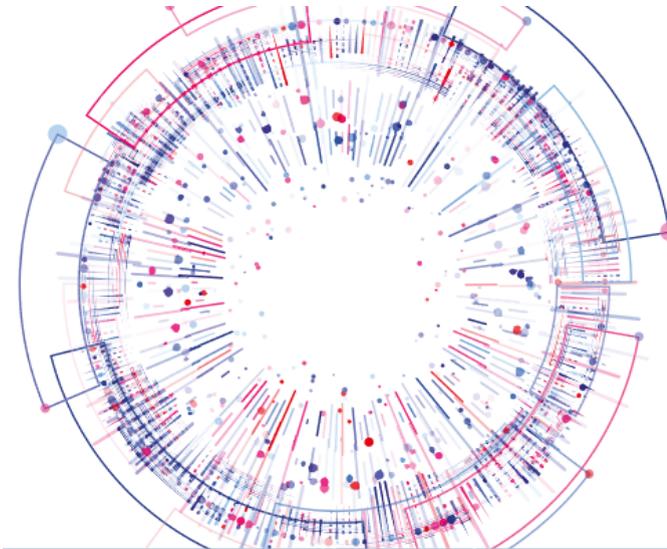
Contenu sous licence ouverte



Avec le soutien du Comité pour la science ouverte

V2 - Décembre 2017

Francine Filoche, Chargée de mission données de la recherche
Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte



GUIDE PRATIQUE

DE LA PUBLICATION EN LIGNE ET DE LA RÉUTILISATION DES DONNÉES PUBLIQUES (« OPEN DATA »)

DOCUMENT ÉLABORÉ PAR LA CADA ET LA CNIL EN ASSOCIATION AVEC ETALAB

CNIL
COMMISSION NATIONALE INFORMATIQUE & LIBERTÉS

Cada | Commission d'accès aux documents administratifs



InSHS

Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte

GUIDE POUR LA RECHERCHE

VERSION 2

FEVRIER 2021



Text button





« Le RGPD encadre le traitement des données personnelles sur le territoire de l'Union européenne.

Ce nouveau règlement européen s'inscrit dans la continuité de la Loi française Informatique et Libertés de 1978 et renforce le contrôle par les citoyens de l'utilisation qui peut être faite des données les concernant.

Il harmonise les règles en Europe en offrant un cadre juridique unique aux professionnels. Il permet de développer leurs activités numériques au sein de l'UE en se fondant sur la confiance des utilisateurs. »

Source : CNIL. RGPD : de quoi parle-t-on ? 10 avril 2018.
<https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>

Les données à caractère personnel

C'est toute information relative à une personne physique susceptible d'être identifiée, directement ou indirectement.

Par exemple :

un nom, une photo, une empreinte, une adresse postale, une adresse mail, un numéro de téléphone, un numéro de sécurité sociale, un matricule interne, une adresse IP, un identifiant de connexion informatique, un enregistrement vocal, etc.

Peu importe que ces informations soient confidentielles ou publiques.

Lorsque ce type de données apparaît durant le projet il est important de **faire appel au Comité d'éthique, à un-e DPO/DPD, un-e référent-e RGPD** pour s'assurer de la **conformité et du respect éthique de ces données**.

La **collecte des données est impactée**, car elle peut nécessiter un **consentement écrit/oral** mais aussi l'**établissement d'une fiche d'information et de traitement** indiquant les objectifs de la collecte, le possible partage, stockage etc.

Le **stockage** de ce type de données nécessite un **stockage crypté ou chiffré** par exemple (pas de cloud, pas de stockage maison).

Le **traitement de ce type de données demande un travail d'anonymisation**. En effet, ces données doivent être rendues anonymes de manière à rendre impossible toute identification de la personne concernée : noms masqués, visages floutés, etc. Seules les données ayant été anonymisées peuvent être ensuite partagées.

Attention : s'il est possible par recoupement de plusieurs informations (âge, sexe, ville, diplôme, etc.) ou par l'utilisation de moyens techniques divers, d'identifier une personne, les données sont toujours considérées comme personnelles.

La **conservation des données personnelles ne peut être indéfinie**. Une durée de conservation doit être déterminée par le responsable de traitement en fonction de l'objectif ayant conduit à la collecte de ces données. Ce principe de conservation limitée des données personnelles est prévu par le **RGPD** et la **loi Informatique et Libertés**.

Le **droit d'auteur** protège les **œuvres et publications des enseignants-chercheurs et chercheurs et salariés de l'université** dès leur création sans nécessité d'obtenir un titre de propriété intellectuelle.

En cas de consortium ou de travail entre équipes internationales, la question de la propriété des données doit être réglée a priori.

Les données qui ne sont pas des œuvres de l'esprit sont des informations publiques.
Pour ces données, **la propriété appartient à l'établissement auquel il est rattaché.**

Les **données et les codes issus de la recherche sont considérés comme des documents administratifs** : cela implique un droit d'accès sur demande, une obligation de diffusion gratuite et une libre réutilisation.

Le droit d'auteur concerne également le logiciel mais les dispositions sont différentes :

- toutes les personnes ayant contribué à la mise en œuvre du logiciel (architecture, rédaction du code, documentation, interface web...) sont considérées comme auteurs du logiciel.

Les droits patrimoniaux, (i.e la possibilité d'exploiter les résultats) des résultats des recherches menées par les enseignants-chercheurs, chercheurs et/ou salariés de l'université dans le cadre de leur mission **appartiennent à l'université.**

Ainsi, tous les droits relatifs à l'utilisation et à l'exploitation d'un logiciel développés par des personnels de l'université appartiennent à l'université qui est libre d'en disposer tant que cela ne va pas à l'encontre des droits moraux des auteurs.

Attention :

Il faut également penser aux **accords de réutilisation** lorsque l'on n'est pas détenteur des données (utilisation de bases de données à accès restreints, sur abonnement etc.)

Ce sont des informations qui révèlent la prétendue **origine raciale ou ethnique**, les **opinions politiques**, les **convictions religieuses** ou **philosophiques** ou l'**appartenance syndicale**, ainsi que le traitement des données génétiques, des **données biométriques** aux fins d'identifier une personne physique de manière unique, des données concernant la **santé** ou des données concernant la **vie sexuelle** ou l'**orientation sexuelle** d'une personne physique.

Le **règlement européen interdit de recueillir ou d'utiliser ces données**, sauf exceptions listées dans les cas suivants :

- si la personne concernée a donné son consentement exprès (démarche active, explicite et de préférence écrite, qui doit être libre, spécifique, et informée) ;
- si les informations sont manifestement rendues publiques par la personne concernée ;
- si elles sont nécessaires à la sauvegarde de la vie humaine ;
- si leur utilisation est justifiée par l'intérêt public et autorisé par la CNIL ;
- si elles concernent les membres ou adhérents d'une association ou d'une organisation politique, religieuse, philosophique, politique ou syndicale.

Il est important de faire appel au **Comité d'éthique** avant la collecte et de **s'assurer d'obtenir une autorisation pour collecter et utiliser et sécuriser ce type de données**. La collecte ne se réalisera qu'après autorisation du Comité d'éthique.

(Source : CNIL_ Commission Nationale Informatique et Libertés accompagne les professionnels dans leur mise en conformité et aide les particuliers à maîtriser leurs données personnelles et exercer leurs droits. Elle analyse l'impact des innovations technologiques et des usages émergents sur la vie privée et les libertés. Elle travaille en étroite collaboration avec ses homologues européens et internationaux pour élaborer une régulation harmonisée.)



- _ L'information précise sur le traitement, la finalité, l'utilisation des données, la durée de conservation doit être faite aux personnes concernées par le traitement ;
 - _ l'information doit être transparente et facilement accessible
 - _ droit d'accès à ses données (Cf. article 15)
 - droit d'être informé d'une violation des données en cas de risque élevé pour les personnes concernées.
 - droit d'opposition (Cf. article 21)
 - droit de rectification (Cf. article 16) : une personne peut demander de modifier les données la concernant.
 - droit à l'effacement (Cf. article 17)
 - droit à la portabilité une personne peut demander à recevoir les données qui la concernent dans un format structuré et lisible par machine et de les transmettre à un autre responsable de traitement.
- Ce droit ne s'exerce pas au traitement nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique dont est investi le responsable de traitement.
- droit à une utilisation restreinte de ses données (Cf. article 23)

L'Analyse d'Impact sur la vie Privée des personnes

La CNIL peut réaliser une analyse d'impact sur la vie privée des personnes -sujets de votre recherche.

Cette analyse permet d'évaluer les risques d'un traitement de données sur la vie privée de personnes concernées.

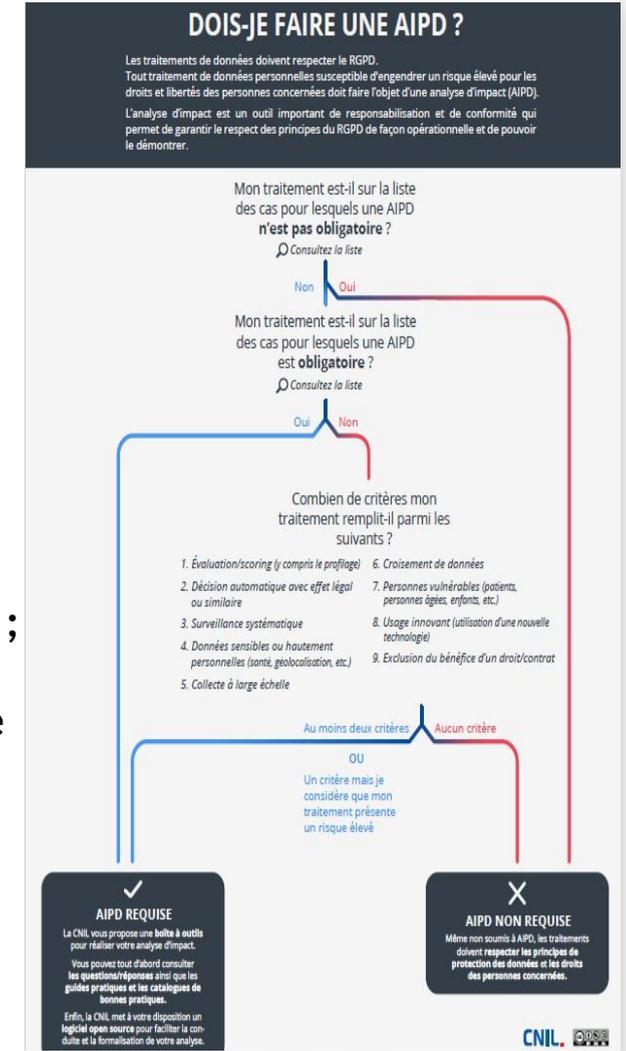
Elle est réalisée par le responsable du traitement (et par la délégation, le responsable scientifique du projet) en lien avec le délégué à la protection des données (DPO) et le responsable de la sécurité des systèmes d'information.

Elle est OBLIGATOIRE LORSQUE LE TRAITEMENT EST SUSCEPTIBLE D'ENGENDRER DES RISQUES ELEVES et doit notamment être réalisée dès lors qu'elle remplit au moins 2 critères suivants :

_ Surveillance automatique ; _ Données sensibles ; _ Traitement à grande échelle ; _ Croisement de données ; _ Personnes vulnérables (patients, personnes âgées, enfants, etc) ; _ Evaluation /scoring (y compris profilage) ; _ Décision automatique avec effet légal ; _ Usage innovant ou utilisation NTIC ; _ Exclusion du bénéfice d'un droit, d'un contrat.

Si vous êtes dans ce cas, utiliser l'outil de ticketing du DPO AMU :

<https://www.univ-amu.fr/fr/public/donnees-personnelles-et-saisine-du-dpo>



Enjeux et bonnes pratiques des données de recherche

UNE BONNE GESTION

- Sauvegarder les données au cours du processus de publication
- Rendre les données réutilisables en les documentant

UNE BONNE CONSERVATION

- Protéger les données
- Se prémunir de la perte et altération
- Permettre une réutilisation dans un nouveau projet
- Conserver une trace de la recherche

UNE BONNE DIFFUSION

- Permettre un accès le plus large possible
- Eviter l'appropriation des données
- Contribuer à la qualité et visibilité de la recherche
- Favoriser les collaborations entre les acteurs scientifiques

Francine Filoche, Chargée de mission données de la recherche
Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte

1. Définir ce que sont vos données et "modéliser" votre processus de recherche

Périmètre de vos typologies de matériaux de recherche
(*Produits de recherche*)

Francine Filoche, Chargée de mission données de la recherche
- Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte



Les données de la recherche : plusieurs définitions

« 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 nécessaire pour valider les résultats de la recherche. »

Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics - 2006

«Loi pour une République Numérique (2016)
L'article 38 de la loi mentionne: «..les fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche»

PLoi pour une République Numérique, 18 juillet 2016

« Enregistrements factuels (chiffres, textes, images, son, vidéo) utilisés comme sources primaires pour la recherche et qui sont habituellement acceptés par la communauté scientifique comme étant nécessaire pour valider les résultats de la recherche. »

Plan pour la Science Ouverte, 04 juillet 2018

« Les données de la recherche désignent les informations, et en particulier les faits ou chiffres collectés pour être analysés et traités pour alimenter des réflexions, discussions ou calculs. Dans un contexte de recherche, ces données sont, par exemple, des statistiques, des résultats d'expériences, des mesures, des observations sur le terrain, des résultats d'enquêtes, des enregistrements d'entrevues ou des images. Il s'agit plus spécifiquement de données disponibles sous forme numérique. »

Lignes directrices sur les règles pour le libre accès aux publications scientifiques et pour le libre accès aux données de la recherche dans Horizon 2020

« Par informations de recherche, nous entendons les informations (parfois appelées métadonnées) relatives à la mise en œuvre et à la communication de la recherche.

Il s'agit notamment des métadonnées bibliographiques telles que les titres, les résumés, les références, les données sur les auteurs, les données sur l'affiliation et les données sur les lieux de publication, des métadonnées sur les logiciels de recherche, les données de recherche, les échantillons et les instruments, des informations sur le financement et les subventions et des informations sur les organisations et les contributeurs à la recherche.

Les informations de recherche se trouvent dans des systèmes tels que les bases de données bibliographiques, les archives de logiciels, les entrepôts de données et les systèmes d'information sur la recherche des établissements.»

Déclaration de Barcelone sur l'information ouverte, avril 2024

Quelle est la nature de vos données ?

Données ayant en général été publiées
Données extraites, triées et agrégées

Données de références

Résultats d'un traitement, d'une combinaison ou d'une sélection de données brutes
Elles sont reproductibles

Données d'observations

Capturées ou collectées en temps réel, unique et impossible à reproduire, importance de documenter la méthode de collecte

Données expérimentales

Créées en laboratoire par exemple, avec une méthode définie
Elles peuvent être reproductibles

Données dérivées

Données de simulation

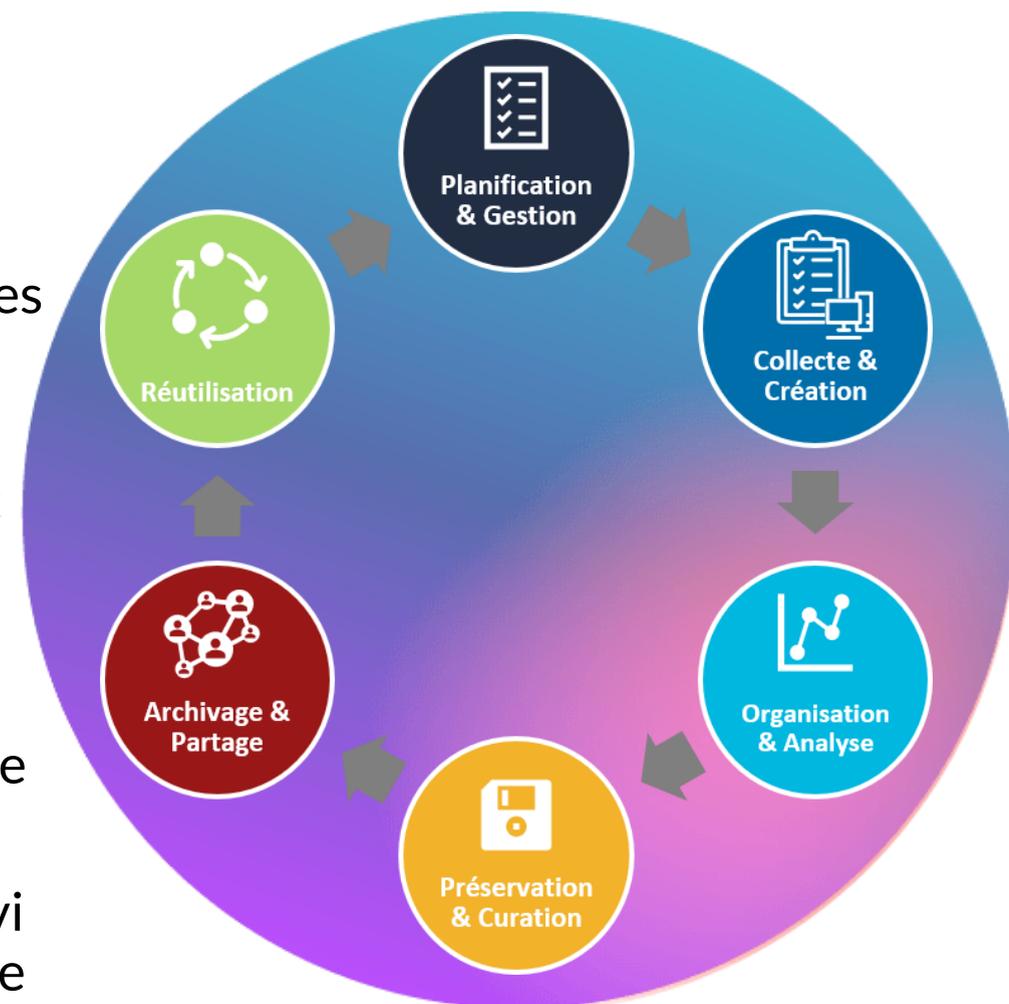
Générées à partir d'un modèle
Elles peuvent être reproductibles si le modèle est bien documenté

Dans quel état ou à quel stade allez-vous collecter ou générer des données ?



Le PGD accompagne le cycle de vie des données

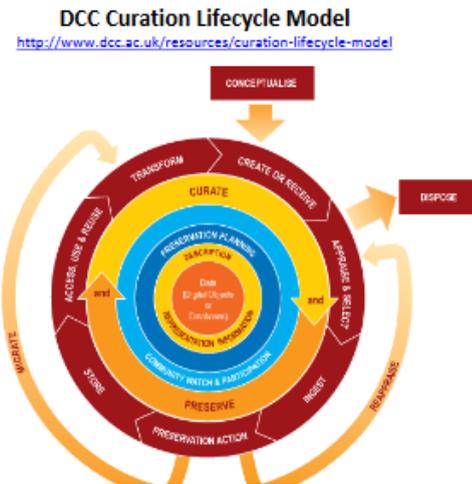
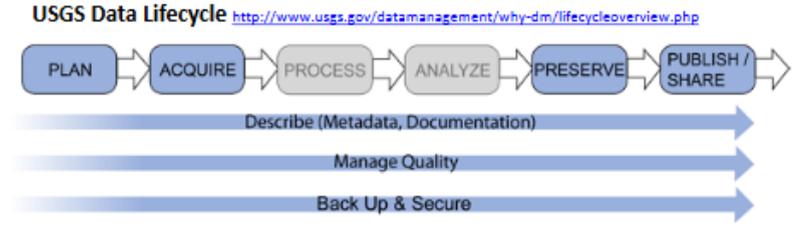
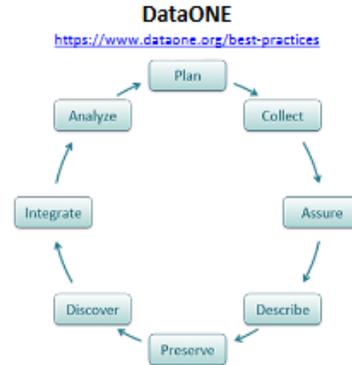
- **Création** : collecter/créer les données
- **Stockage** : définir un espace de stockage des données
- **Traitement** : entrer, numériser transcrire etc. décrire les données et gérer et sauvegarder les données
- **Analyse** : analyser et interpréter les données, produire les résultats de recherche, citer les sources de données
- **Archivage** : préparer les données pour la conservation, migrer les données vers le meilleur format et un support adapté, définir les accès et promouvoir les données archivées
- **Partage** : définir les droits d'auteur et d'utilisation, contrôler les métadonnées et la documentation liée, faire des copies de sauvegarde
- **Réutilisation** : Réaliser des analyses secondaires, un suivi de recherche, une nouvelle recherche, des évaluations de la recherche...



Renseigner les rubriques en suivant le cycle de vie



Autres exemples



The active management of data throughout the lifecycle

(Jones S, 2015.
http://www.consortiomadrona.es/noticias_eventos/2015/Jo-madaPGD/sarah.pdf)

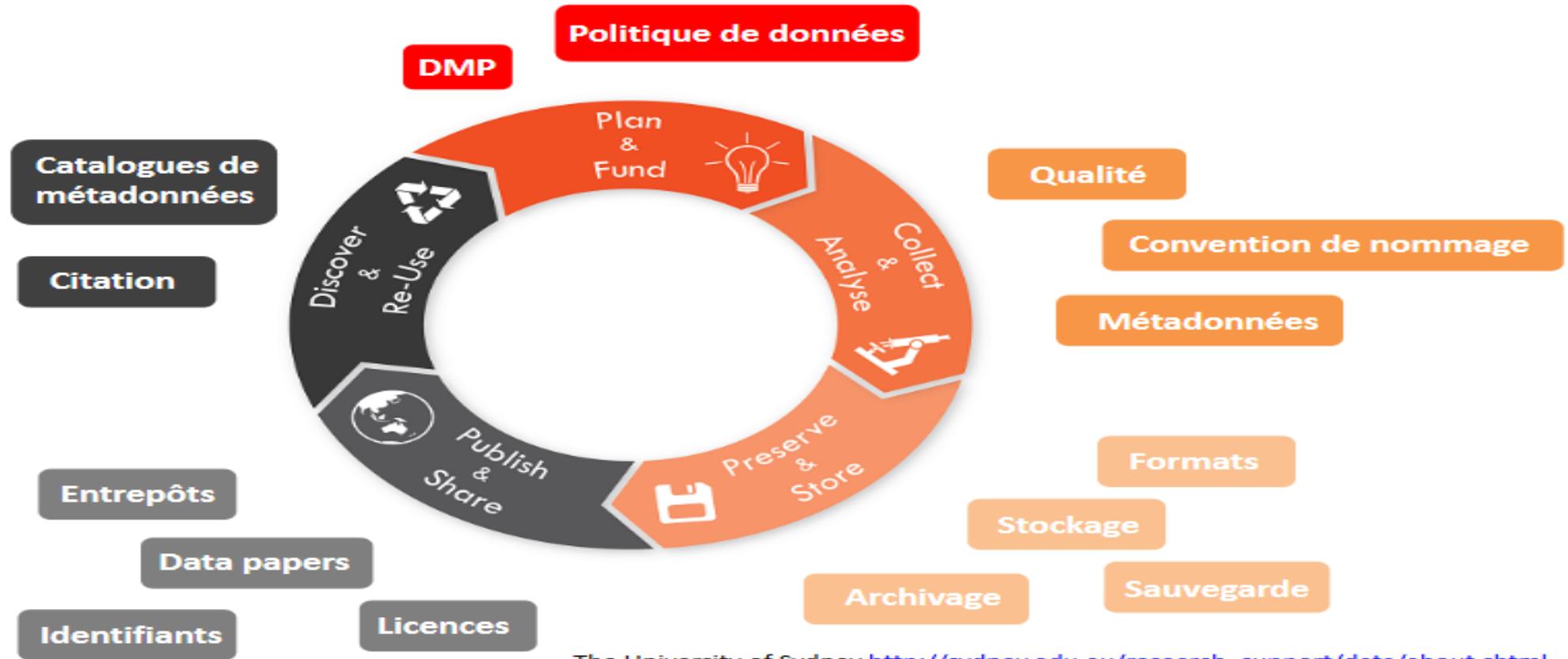
Research Data Lifecycle Diagram : JISC and Bonner McHardy

(Licence CC-BY-NC-ND)
<https://www.jisc.ac.uk/guides/research-data-management>



Bonnes pratiques et cycle de vie

Les bonnes pratiques de gestion des données s'appliquent à chaque étape du cycle de vie



The University of Sydney http://sydney.edu.au/research_support/data/about.shtml

Francine Filoche, Chargée de mission données de la recherche

- Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte



Accompagnement : Atelier de la Donnée à AMU

■ ACTEURS ET CONTRIBUTEURS (2/2)

Les **universités, infrastructures et organismes de recherche** émettent souvent des recommandations à destination de leurs communautés de recherche.

De même les **financeurs** (comme l'ANR ou la Commission européenne) et certains éditeurs peuvent proposer des conseils ou donner des consignes précises (par exemple : obligation de rédaction d'un PGD dans les 6 mois suivant le début du projet pour les projets financés par l'ANR).



Dans le cadre de **Recherche Data Gouv**, les **ateliers de la donnée** sont en proximité géographique des équipes de recherche pour leur apporter une première expertise dans la gestion raisonnée des données de recherche.

Voir le site des ateliers de la donnée



Cliquer pour en savoir plus

Francine Filoche, Chargée de mission données de la recherche

- Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte

2. Documenter vos métadonnées en suivant les principes FAIR

- Encourager à archiver durablement les données et métadonnées
- Faciliter leur accès et/ou leur téléchargement
- Spécifier les conditions d'accès et d'utilisation
- Définir les conditions d'accès aux données
- En cas de d'accès restreint, rendre accessibles les métadonnées pour signaler l'existence des données



Faciles à trouver

- Faciliter la découverte des données par les humains et les systèmes informatiques
- Décrire et indexer des données et des métadonnées
- Déposer les données dans un entrepôt
- Attribuer un identifiant unique et pérenne (DOI)
- Décrire les données par les métadonnées



Accessibles



- Mettre en avant les caractéristiques qui rendent les données réutilisables pour de futures recherches ou d'autres finalités
- Associer une licence de diffusion
- Décrire en détail les données, les contextualiser, les rendre compréhensibles



Intéropérables

- Téléchargeable, utilisable, intelligible, combinable avec d'autres données, par les humains et les machines
- Formats ouverts
- Mises à disposition du code source du logiciel
- Standards des métadonnées et vocabulaires contrôlés



Réutilisables





A promotional graphic for "Stories of Data: Open Science Talk". It features a pair of white headphones on the left. To the right of the headphones are two logos: "Listen On Anchor" (purple) and "Listen on Spotify" (green). Further right is a stylized blue eye with circuit-like patterns. At the bottom left is the "EDSC-Pillar" logo. The background is white with a blue audio waveform at the bottom.



F comme Findable

Rendre les données localisables, notamment par le biais **de métadonnées**

- ❖ Décrire le processus permettant de retrouver les données (fourniture de **métadonnées**) Décrire les items identifiant les données et se référer aux mécanismes **d'identification standard**
- ❖ Utilisez-vous des identifiants pérennes et uniques tels que les **DOI** ?
- ❖ Conventions de **nommage** utilisées
- ❖ Décrire l'approche de recherche par **mots clés**
- ❖ Décrire la méthode utilisée pour un **versionnage** clair
- ❖ Spécifiez des **normes** pour la **création de métadonnées** (le cas échéant)
- ❖ *S'il n'y a pas de normes dans votre discipline* décrire **quel type de métadonnées** sera créé et comment



A comme Accessible (1)

- Spécifiez quelles données seront **ouvertement** disponibles. Si certaines données sont maintenues **fermées**, précisez le motif de cette restriction
- Précisez la manière dont les données seront **mises à disposition**
- Précisez quelles **méthodes ou outils logiciels** sont nécessaires pour accéder aux données. Une documentation sur le logiciel est-elle nécessaire pour accéder aux données incluses ?

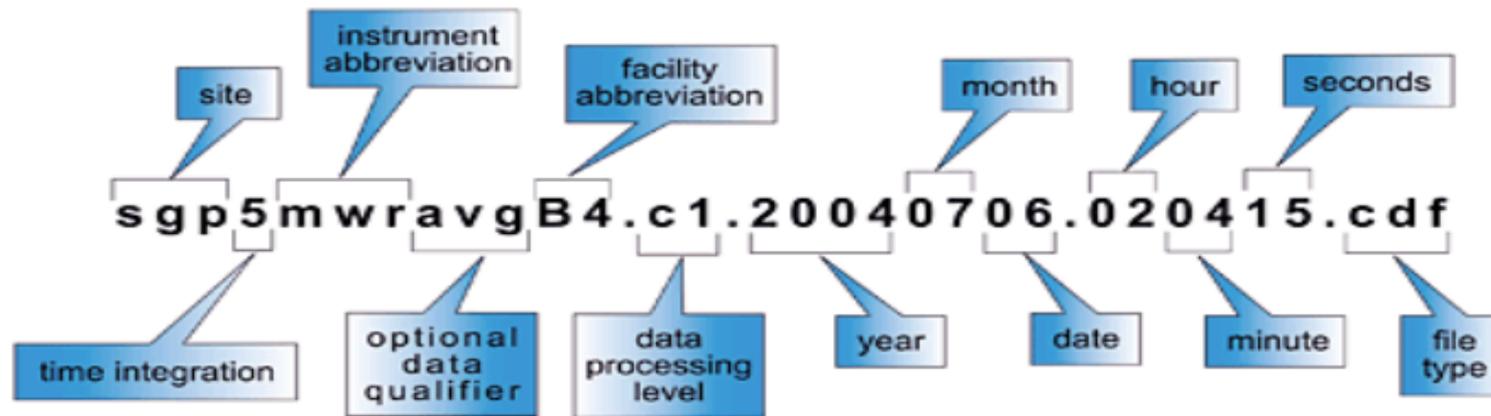


A comme Accessible (2)

- Est-il possible d'inclure un logiciel adapté (par exemple doté **d'un code source ouvert**)?
- Précisez **où** les données et **les métadonnées** associées, **la documentation** y afférente et **le code** sont *déposés*
- Précisez dans **quelles conditions** l'accès sera fourni en cas de **restrictions**

Des noms de fichiers clairs et harmonisés

- Exemple de structure de nom (U.S. Dpt.of Energy, ARM Climate Res.Facility)



Le fichier `sgp5mwravgB4.c1.20040706.020415.cdf` contient des données de radiomètre à micro-ondes avec un temps d'intégration moyen de 5 minutes, du site Vici Southern Great Plains(USA) à partir du 6 juillet 2004. Le niveau des données est "c1" indiquant qu'elles ont été dérivées ou calculées via un processus de transformation à valeur ajoutée <https://dis.arm.gov/data/docs/plan>



Les FORMATS OUVERTS vs FERMÉS (propriétaires)

- **ODT** (OpenDocument Text)
- **RTF** (Rich Text Format)
- **TXT** (Text)

TEXTE



- **DOC** (Microsoft Word)
- **PAGES** (Mac)
- **WP** (WordPerfect)

- **BMP** (Windows BitMaP)
- **GIF** (Rich Text Format)
- **JPG** (Joint Photographic Expert Group)
- **PNG** (Portable Network Graphics)

IMAGE



- **JP2** (Joint Photographic Expert Group 2000)
- **TIF** (Tagged Image File Format)

- **CSV** (Comma-separated values)
- **ODS** (OpenDocument Spreadsheet)

TABLEUR



- **XLS** (Microsoft Excel)
- **NUMBERS** (Mac)

- **MKV** (Matroska)
- **NUT**
- **OGM** (OGG Media)

VIDEO



- **AVI** (Audio Video Interleave)
- **FLV** (Flash Video)
- **MOV** (QuickTime Movie)
- **MP4** (MPEG-4 Part 14)
- **QT** (QuickTime Movie)
- **WMV** (Windows Media Video)

AUDIO



- **FLAC** (Free Lossless Audio Codec)
- **MP3** (MPEG-1/2_Audio_Layer_III)
- **OGG**
- **WAV** (Windows Media Audio)

- **AAC** (Advanced Audio Coding)
- **RAM** (Real Audio Metadata)
- **WMA** (Windows Media Audio)

- **ODP** (OpenDocument Presentation)

PRÉSENTATION



- **KEY** (Keynote Mac)
- **PPT** (Microsoft PowerPoint)

- **7z** (Seven Zip)
- **TAR** (Tape Archiver)
- **ZIP**

ARCHIVAGE



- **ACE**
- **RAR** (Roshal Archive)

- **EPS** (Encapsulated PostScript)
- **ODG** (OpenDocument Graphics)
- **SVG** (Scalable Vector Graphics)

DESSIN



- **AI** (Adobe Illustrator Artwork)
- **CDR** (CorelDRAW)
- **FH** (Adobe FreeHand)

Cap sur l'interopérabilité

Utiliser des langages communs

- Standards de métadonnées
- Terminologies : ontologies, thesaurus, classification...
- Identifiants pérennes
- Formats ouverts
- Outils
- Protocoles ...



I comme Interopérable



- ✧ Évaluez l'interopérabilité de vos données.
- ✧ Spécifiez les **vocabulaires** de données et de métadonnées, **normes** ou méthodologies que vous suivrez pour **faciliter l'interopérabilité**.
- ✧ Indiquez si vous utiliserez un **vocabulaire standard** pour tous les types de données présents dans votre ensemble de données, afin de permettre **l'interopérabilité interdisciplinaire**
- ✧ *Sinon*, allez-vous fournir un **mappage associant les ontologies les plus utilisées ?**

Rubrique . Les exigences légales et éthiques



22 février 2021

Charte de transparence sur le recours aux animaux à des fins scientifiques et réglementaires en France

La recherche biologique et médicale vise à l'accroissement continu des connaissances scientifiques, le progrès des thérapies en médecine humaine et vétérinaire et une meilleure protection de l'homme, de l'animal et de l'environnement.

La recherche, qu'elle soit fondamentale ou appliquée, requiert des modèles pour progresser. Elle nécessite une approche intégrée qui s'appuie sur la complémentarité des méthodes expérimentales basées sur l'informatique (in silico), la biochimie, la culture de cellules et de tissus (in vitro) et le recours aux animaux vivants (in vivo), ainsi que sur les données cliniques. Les modèles in silico et in vitro sont très largement utilisés, mais ne sont en mesure de remplacer les modèles animaux que partiellement en l'état actuel des connaissances. Les modèles utilisant des animaux restent donc nécessaires pour mener à bien une recherche de qualité.

Dans le développement et la production des médicaments, qu'ils soient humains ou vétérinaires, le recours aux animaux répond également à des exigences réglementaires de sécurité et d'efficacité. Dans le domaine vétérinaire, les études sont réalisées sur les espèces animales auxquelles les produits sont destinés.

En conformité avec les dispositions légales du Code rural et du Code civil, et au titre de leurs convictions, les signataires de cette Charte reconnaissent que les animaux sont des êtres vivants doués de sensibilité. Le recours aux animaux en recherche scientifique et médicale est encadré par une réglementation spécifique qui les protège, veille à leur bien-être et impose le respect de règles éthiques.

Chaque citoyen a droit à une information complète, claire et exacte sur les raisons et les conditions de l'utilisation d'animaux à des fins scientifiques ou réglementaires, sur le cadre réglementaire de cette utilisation ainsi que sur les progrès scientifiques et médicaux qui en découlent.

C'est dans ce but que les signataires de cette charte, établissements de recherche, laboratoires développant des médicaments destinés à l'homme ou aux animaux, leurs partenaires et leurs représentants, prennent quatre engagements :

1. Expliquer les raisons et les conditions du recours aux animaux à des fins scientifiques et réglementaires

- Nous nous engageons à respecter un principe de transparence et d'information sur nos pratiques en matière de recours aux animaux à des fins scientifiques ou réglementaires.
- Nous informons les personnes travaillant dans nos établissements et les personnes externes que nous réalisons ou participons à la réalisation d'études sur animaux. Nous en précisons les raisons et les conditions, chaque fois que cela est possible sans porter atteinte à la confidentialité des personnes et des essais.
- Lorsque nous communiquons sur nos avancées, nous précisons la part des modèles animaux et non animaux dans ces progrès.

2. Diffuser l'information auprès du grand public et des médias

- Nous nous engageons à ce que l'information sur le recours aux animaux à des fins scientifiques ou réglementaires soit accessible au grand public et aux médias.
- Ainsi, l'information sur le recours aux animaux en recherche fondamentale ou appliquée et pour le développement de médicaments humains et vétérinaires sera compréhensible par tous et d'un accès facile sur les différents supports de communication des établissements, y compris sur les sites internet institutionnels.

3. Faciliter les échanges d'information avec le grand public et des médias

- Nous nous engageons à faciliter et encourager les échanges avec le grand public et son information.
- Différentes méthodes d'interaction seront mises en place telles que la définition d'un point de contact à destination des médias et des personnes ou la mise en place de rencontres avec le public.

4. Produire chaque année un document sur les progrès en matière d'information du public

- Il est nécessaire de réaliser un suivi des engagements pour garantir la réussite de la charte.
- Dans ce but, nous ferons part des actions menées consécutivement à la signature de la charte dans un document de synthèse annuel.



Avec le soutien de : Académie Vétérinaire de France



Pratiquer une recherche intègre et responsable

Guide

DÉONTOLOGIE

CHARTRE DÉONTOLOGIQUE DU CNRS



Comité d'éthique du CNRS
www.cnrs.fr/comets
Mars 2017



Charte nationale de déontologie des métiers de la recherche

La charte nationale de déontologie des métiers de la recherche concerne l'ensemble des femmes et des hommes (désignés dans le texte par le terme générique « chercheur ») d'un établissement ou d'un organisme, permanents ou non, qui contribuent à l'activité de recherche et s'engagent à respecter, dans le cadre des missions de recherche et d'appui à la recherche qui leur incombent, les principes d'intégrité qui y sont formulés.

1. Respect des dispositifs législatifs et réglementaires

Tout chercheur se tient informé des dispositifs législatifs et réglementaires qui régissent les activités professionnelles et veille au respect des textes étudant, s'agissant notamment des recherches sur l'être humain, l'animal et l'environnement.

2. Fiabilité du travail de recherche

Les chercheurs doivent respecter les engagements pris dans le cadre de leur unité de recherche ou dans le cadre de contrats spécifiques. Les méthodes mises en œuvre pour la réalisation du projet de recherche doivent être les plus appropriées.

La description détaillée du protocole de recherche, dans le cadre de cahiers de laboratoire ou de tout autre support, doit permettre la reproductibilité des travaux expérimentaux.

Tous les résultats bruts (qui appartiennent à l'institution) ainsi que l'analyse des résultats doivent être conservés de façon à permettre leur vérification.

Les conclusions doivent être fondées sur une analyse critique des résultats et les applications possibles ne doivent pas être amplifiées de manière injustifiée. Les résultats doivent être communiqués dans leur totalité de manière objective et honnête.

Tout travail de recherche s'appuie naturellement sur des études et résultats antérieurs. L'utilisation de ces sources

French National Charter for Research Integrity

The French National Charter for Research Integrity concerns all women and men (referred to in the text by the generic term 'researcher') forming part of an institution or body, whether permanent staff or not, who contribute to a research activity and who undertake to respect the principles of integrity expressed herein, in the context of those research projects in which they are involved, either directly or indirectly.

1. Compliance with legislative and regulatory requirements

All researchers should ensure that they remain informed and up-to-date concerning the legislation and regulations which govern their professional activities and that they comply with any such legislation, notably that covering research on human subjects, animals and the environment.

2. Reliability of research work

Researchers must respect the commitments undertaken by their research unit or for any specific contract. They must always use the most appropriate methods to conduct any research project.

A detailed description of the research protocol, through the use of laboratory notebooks, or any other media, must enable the replication of the experimental work.

All raw data (which are the property of the institution) together with their analysis must be conserved to enable their verification.

The conclusions must be based on a critical analysis of the results and possible applications should not be unjustifiably amplified. The results should be communicated in their entirety in an objective and honest manner.

All research naturally relies on previous studies and results. The use of these sources must be correctly cited in any scientific production, publication and communication. In certain situations, their use may

TX-DFD-40

1

MAJ 29/04/20



Francine Filoche, Chargée de mission données de la recherche
- Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte



- Encourager à archiver durablement les données et métadonnées
- Faciliter leur accès et/ou leur téléchargement
- Spécifier les conditions d'accès et d'utilisation
- Définir les conditions d'accès aux données
- En cas de d'accès restreint, rendre accessibles les métadonnées pour signaler l'existence des données



Accessibles



Faciles à trouver

- Faciliter la découverte des données par les humains et les systèmes informatiques
- Décrire et indexer des données et des métadonnées
- Déposer les données dans un entrepôt
- Attribuer un identifiant unique et pérenne (DOI)
- Décrire les données par les métadonnées



Intéropérables

- Téléchargeable, utilisable, intelligible, combinable avec d'autres données, par les humains et les machines
- Formats ouverts
- Mises à disposition du code source du logiciel
- Standards des métadonnées et vocabulaires contrôlés

Reproductible

R

Réutilisables



- Mettre en avant les caractéristiques qui rendent les données réutilisables pour de futures recherches ou d'autres finalités
- Associer une licence de diffusion
- Décrire en détail les données, les contextualiser, les rendre compréhensibles



- La confiance dans les résultats repose entre autres sur le fait que les expériences ou les calculs soient reproductibles.
- En fonction des disciplines il est compliqué de reproduire des résultats ou répliquer des résultats.
- Il s'agit ici pour les disciplines concernées de détailler les campagnes de calcul, les méthodes, les solutions et outils, les codes, les standards respectés, les environnements étudiés, les exécutions de tâches, les mesures appliquées, les compilations de tests et séries de répliquions effectuées, la gestion des résultats, les intégrations...



- **Reproductibilité** avec un même instrument, une méthode identique, dans un environnement donné est essentielle pour valider les résultats d'une réxpérience (ex: biologie computationnelle).
- Nécessité de maîtriser les outils pour des campagnes de calculs pour applications de mmodélisation, simulations paramétriques, analyses de données expérimentales....
- Développement de code de calcul open source



La Qualité des Données

- C'est le but ultime de la gestion des Données, tant en interne que pour leur diffusion
- Elle participe « ...des processus de préparation des données préalable à la fouille de données elle-même. Ces activités, que l'on regroupe sous le terme de *curation* de données, englobent la **sélection, vérification, normalisation, annotation, reformatage, enrichissement, structuration des données collectées** ; le but ultime est de disposer de *données qualifiées* pour être soumises à un processus de fouille scientifiquement valide. »

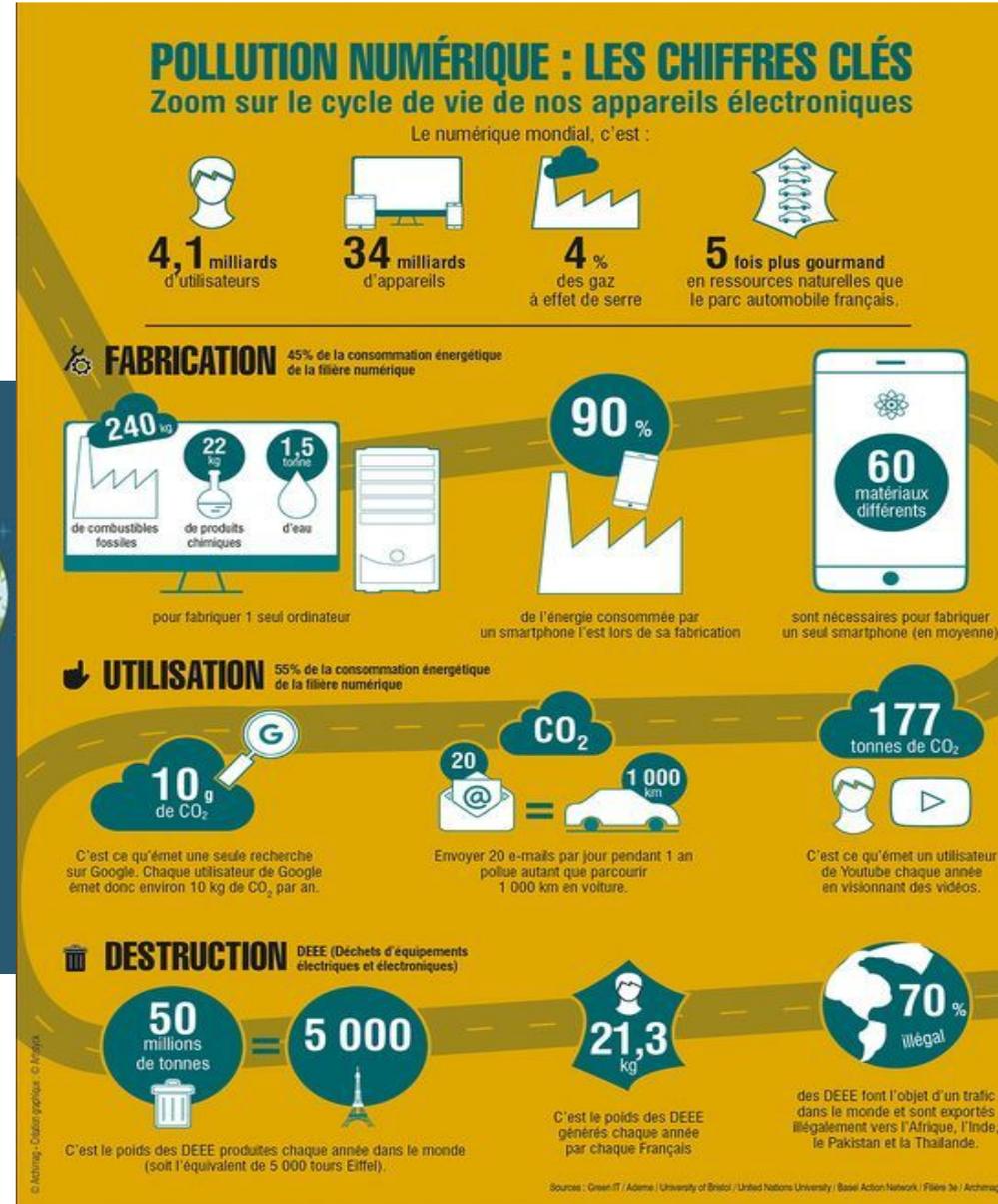
3. Lier et partager vos données de recherche avec vos publications

Cimétière de données à exhumer : le disque dur

On estime que les publications permettent d'accéder à environ 10 % des données, le reste demeurant disponible mais non utilisé sur les disques durs d'ordinateurs



Penser à trier et sélectionner le plus significatif



Principe « Aussi ouvert que possible, aussi fermé que nécessaire »

STOCKAGE ET SAUVEGARDE



Serveur sécurisé
de l'institution

Espace collaboratif
de travail



PARTAGE



Entrepôt généraliste
ou disciplinaire



Ressource :

✓ [DoRANum - Stockage, partage et archivage : quelles différences ?](#)



Définition et
cycle de vie

Mesures de
sauvegarde

Nommage
des fichiers

Formats
de fichiers

■ MESURES DE SAUVEGARDE À METTRE EN PLACE

Une sauvegarde efficace signifie qu'il faut **dupliquer** et **stocker** les données à différents endroits sur différents supports selon une temporalité pertinente pour le projet. L'idéal est d'appliquer la **règle du 3-2-1**.

Dans tous les cas, il faut organiser et **planifier ces sauvegardes** en veillant à bien gérer les **versions**. À chaque point d'étape du projet, sélectionnez les données à sauvegarder, à supprimer. Les différents états des données sont conservés en corrélation avec les différentes étapes de traitement, ce qui permet de revenir à une version antérieure si besoin.



Survoler pour voir le détail

Cela nécessite aussi de définir un **hébergement** et une **politique de sauvegarde** adaptés aux besoins du projet concernant les spécificités de stockage des données (par exemple en cas de données sensibles, de grosse volumétrie...). Cela peut être sur des serveurs locaux (machines virtuelles), un cloud institutionnel avec accès sécurisé... Il est recommandé d'éviter au maximum les outils du type One Drive, Google Drive, Dropbox, etc. Ne pas hésiter à se rapprocher de son établissement afin de connaître les espaces de stockage sécurisés mis à disposition.



MESURES DE SAUVEGARDE À METTRE EN PLACE

La **règle du 3-2-1** signifie :

- garder 3 exemplaires des données,
- sur 2 supports ou technologies différents,
- dont 1 se trouve hors site.

à différents endroits sur différents supports pour respecter la **règle du 3-2-1**.

bien
es à
Q
tir à



Survoler pour voir le détail

exemple en cas de données sensibles, (machines virtuelles), un cloud au maximum les outils du type rocher de son établissement afin on.

Principe « Aussi ouvert que possible, aussi fermé que nécessaire »



PARTAGE

ARCHIVAGE PÉRENNE



Entrepôt généraliste
ou disciplinaire



Plateforme d'archivage
(ex : CINES)

Ressource :

- ✓ [DoRANum - Stockage, partage et archivage : quelles différences ?](#)

Où déposer ?

re3data.org

On a vu l'importance de Zenodo au plan européen

Voir aussi l'annuaire international **RE3DATA**

On peut consulter ce répertoire

<https://www.re3data.org/browse/> (en lien avec DataCite)

- Par sujet
- Par type de contenu
- Par pays

La requête est matérialisée par une succession de cercles concentriques, du plus général vers le plus spécifique



Entrepôts de données

- Data repository, digital repository
- Un **entrepôt de données** est un réservoir de données de recherche, brutes ou dérivées, qui peuvent être retrouvées et réutilisées grâce à une description par des métadonnées. Un identifiant pérenne ou numéro d'accès est attribué à chaque jeu de données (d'après Aventurier P, 2013)
- Types d'entrepôts
 - Disciplinaire: [SEANOE](#), [PANGAEA](#), [Réseau Quetelet](#), [Ortolang](#)
 - Multidisciplinaire : [Zenodo](#), [Figshare](#), [Dryad](#)
 - Institutionnel : [Edinburgh DataShare](#), [Merritt](#), [3TU Datacentrum](#)
 - Spécifique d'un projet de recherche: [Scientific Drilling Database](#)
 - Banque de données de référence : [GenBank](#), [UniProt](#), [COD](#)



MA CONFORMITÉ AU RGPD | RESSOURCES | TEXTES OFFICIELS | LA CNIL | 🔍

🏠 > La CNIL adopte un référentiel sur les entrepôts de données de santé

La CNIL adopte un référentiel sur les entrepôts de données de santé

26 octobre 2021

La création d'un entrepôt de données de santé nécessite le respect de certaines formalités. La CNIL a adopté un nouveau référentiel pour simplifier les procédures en proposant un cadre adapté aux pratiques.

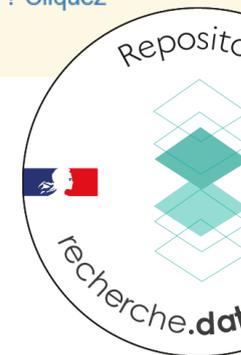
RÉFÉRENTIEL

RELATIF AUX TRAITEMENTS DE DONNÉES A
CARACTÈRE PERSONNEL MIS EN ŒUVRE A
DES FINS DE CRÉATION D'ENTREPÔTS DE
DONNÉES DANS LE DOMAINE DE LA SANTÉ



Stratégie de partage des données et conservation à long terme

Bienvenue ! Pour déposer des données, rendez-vous dans votre espace institutionnel ou dans l'espace générique, puis connectez-vous via le menu en haut à droite. (Vous ne savez pas où déposer ? Cliquez ici.)



DAT'AMU _ Entrepôt de données de recherche d'Aix Marseille Université. DAT'AMU _ Aix Marseille University research data warehouse.

(Aix-Marseille Université)

[Recherche Data Gouv >](#)

[Contact](#) [Partager](#)

Welcome to the DAT'AMU warehouse platform of Aix-Marseille University !

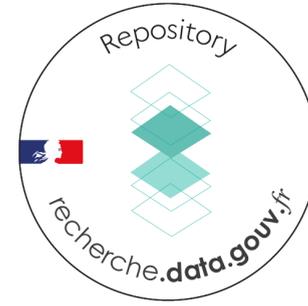
Bienvenue sur la plateforme entrepôt DAT'AMU d'Aix-Marseille Université !



Francine Filoche, Chargée de mission données de la recherche
- Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte



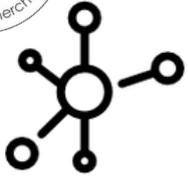
Atelier de la Donnée au SCD
Guichet de la Donnée du site d'Aix-Marseille
Fabien BORGET
21/11/2023



Les principes directeurs de Recherche Data Gouv :

Multidisciplinaire

Toutes les thématiques de la recherche sont concernées



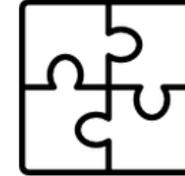
Confiance

Modération et objectif de certification Core Trust Seal



Mutualisation

Les moyens logiciels et l'infrastructure seront mutualisés



Visibilité et reconnaissance des productions et des acteurs



Meilleure visibilité internationale des données de la recherche française



Subsidiarité

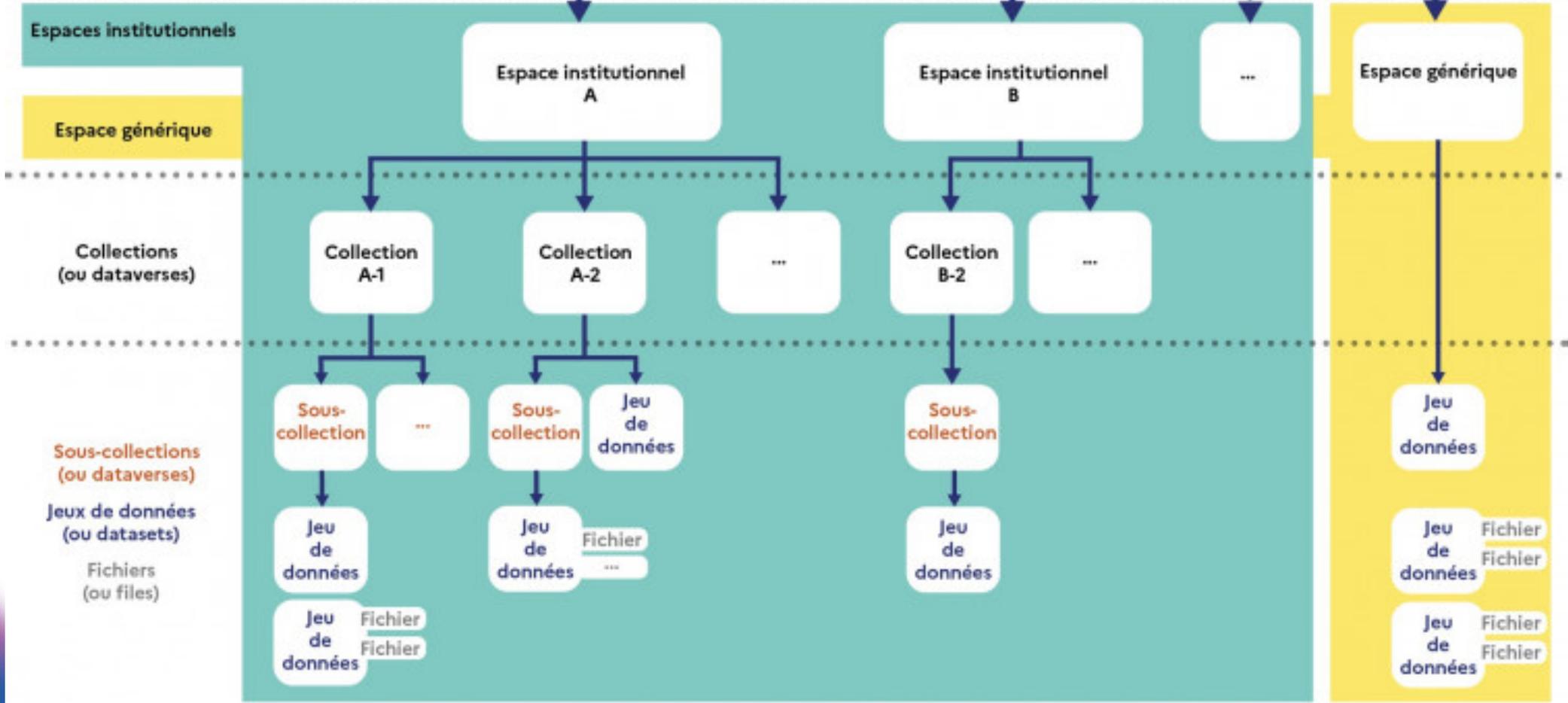
- Avec les autres entrepôts thématiques
- De l'accompagnement en proximité



Organisation de l'entrepôt Recherche Data Gouv

Page d'accueil de l'entrepôt

Entrepôt Recherche Data Gouv



Stratégie de partage des données et conservation à long terme

Qu'est-ce qu'un espace institutionnel ?

Un espace institutionnel est une collection créée à la demande d'un établissement de recherche français et dont l'administration et la curation lui sont déléguées. Un espace institutionnel contient des jeux de données pouvant être réunis en collections ou souscollections.

Qu'est-ce que l'espace générique ?

L'espace générique désigne la collection dans laquelle les déposants ne disposant pas d'un espace institutionnel peuvent déposer leurs jeux de données. L'espace générique peut contenir des collections d'unités dans le cas particulier où aucune tutelle ne dispose d'espace institutionnel.

Pour créer une collection d'unité, remplir le formulaire de demande.

Sur la page d'accueil, **un carrousel permet d'accéder directement aux collections associées aux établissements** (espaces institutionnels), et à l'espace générique.



Stratégie de partage des données et conservation à long terme

Bienvenue ! Pour déposer des données, rendez-vous dans votre espace institutionnel ou dans l'espace générique, puis connectez-vous via le menu en haut à droite. (Vous ne savez pas où déposer ? Cliquez ici.)

Espace Générique



Recherche Data Gouv >

Espace commun utilisable par les membres de la communauté scientifique française qui ne peuvent pas institutionnel.

Cet espace contient des données de tous types et de toutes disciplines. Les tests sont à effectuer dans le [bac à sable](#).

This common space can be used by members of the French scientific community who cannot deposit and This space contains data of all types and from all disciplines. Tests should be performed in the [sandbox](#).

Chercher dans cette collection... Recherche avancée

- Collections (1)
- Jeux de données (56)
- Fichiers (4 508)

Année de publication
2024 (10)
2023 (37)
2022 (10)

1 à 10 de 57 résultats

A 2 month-long annotated skylight polarization images database.
7 mai 2024
POUGHON, Leo; AUBRY, Vincent; MONNOYER, Jocelyn; VIO annotated skylight polarization images database.", <https://doi.org/10.21203/rs.3.rs-3120312/v1>
A public database composed of real-world color polarimetric images of the Timone, Marseille, with improved calibration in regard of previous work (http://www.robots.ox.ac.uk/~vgk/research/09_Polarization/)

Aix*Marseille université Socialement engagée
DAT'AMU _ Entrepôt de données de recherche d'Aix Marseille Université. DAT'AMU _ Aix Marseille University research data warehouse.

Bienvenue ! Pour déposer des données, rendez-vous dans votre espace institutionnel ou dans l'espace générique, puis connectez-vous via le menu en haut à droite. (Vous ne savez pas où déposer ? Cliquez ici.)



DAT'AMU _ Entrepôt de données de recherche d'Aix Marseille Université. DAT'AMU _ Aix Marseille University research data warehouse.

(Aix-Marseille Université)

Recherche Data Gouv >



Contact Partager

Welcome to the DAT'AMU warehouse platform of Aix-Marseille University !

Bienvenue sur la plateforme entrepôt DAT'AMU d'Aix-Marseille Université !

Aix-Marseille Université

Université Aix-Marseille recherche DAT'AMU _ Entrepôt de données de recherche d'Aix Marseille Université. DAT'AMU _ Aix Marseille University research data warehouse.

Chercher dans cette collection... Recherche avancée

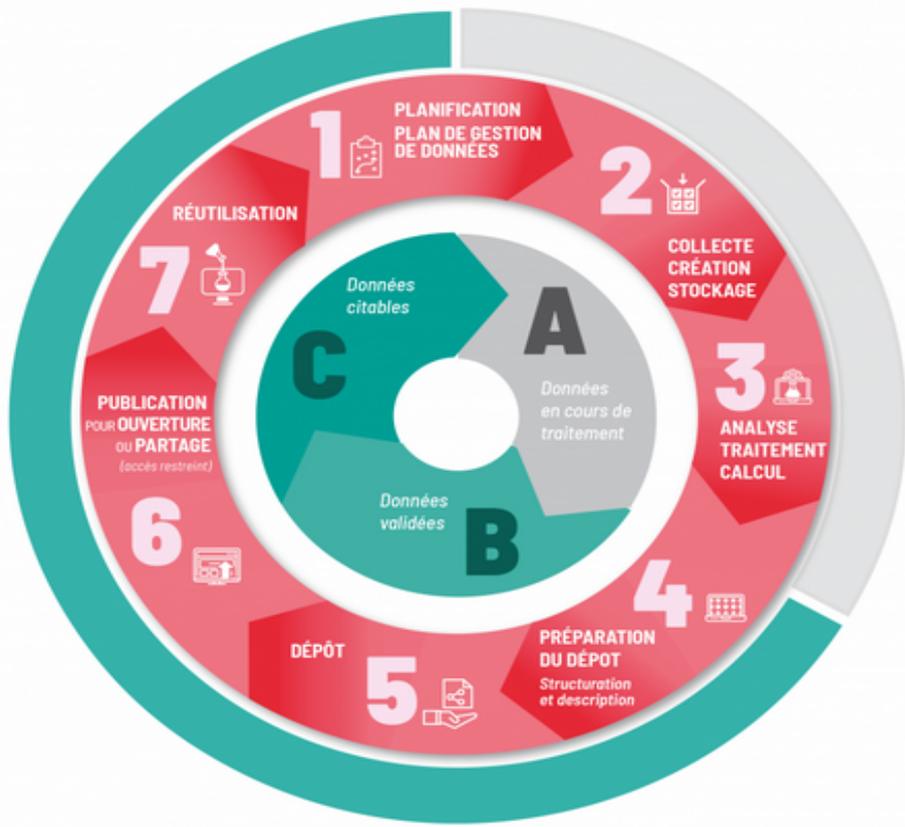
Lire la suite de Description [+]

Ajouter des données

- Collections (0)
- Jeux de données (0)
- Fichiers (0)

Cette collection ne contient actuellement aucune collection, aucun jeu de données ou fichier. Veuillez vous authentifier pour voir si vous pouvez y ajouter du contenu.

Navigation bar with logos for Espace Générique, Arts et Métiers Sciences et technologies - Arts et Metiers Institute of Technology, cirad Centre de Coopération Internationale en Recherche Agronomique pour le Développement, and INRAE Data INRAE.



Etapes : 4-5-6-7

Recherche Data Gouv met à disposition des jeux de données de recherche achevés. Ces jeux de données sont ouverts ou partagés en accès restreint quand des exceptions légitimes encadrées par la loi l'imposent, par exemple en ce qui concerne le secret professionnel, les secrets industriels et commerciaux, les données personnelles.

Les ateliers de la donnée accompagnent les équipes de recherche tout au long du cycle de vie de la donnée, par contre l'entrepôt et le catalogue de données de Recherche Data Gouv ne soutiennent que la partie aval du cycle de vie de la donnée relatif à la diffusion de jeux de données achevés.

Le premier tiers de ce cycle qui couvre les étapes de collecte, création, stockage et de traitement des données pour aboutir à des données scientifiquement validées, se déroule au sein des laboratoires en mobilisant des services de stockage et de calcul. L'entrepôt de données de Recherche Data Gouv n'est pas une solution de stockage pour les données en cours de traitement.

L'entrepôt de données Recherche Data Gouv intervient à partir du deuxième tiers du cycle de vie des données, au niveau du dépôt et de la publication des données scientifiques validées et de la curation des jeux de données en effectuant une vérification des métadonnées et des fichiers de données déposés dans l'entrepôt.

L'entrepôt et le catalogue Recherche Data Gouv interviennent également au niveau du troisième tiers du cycle de vie des données, qui consiste à exposer des données citables, à travers l'ouverture ou le partage (accès restreint) des données et leur réutilisation.

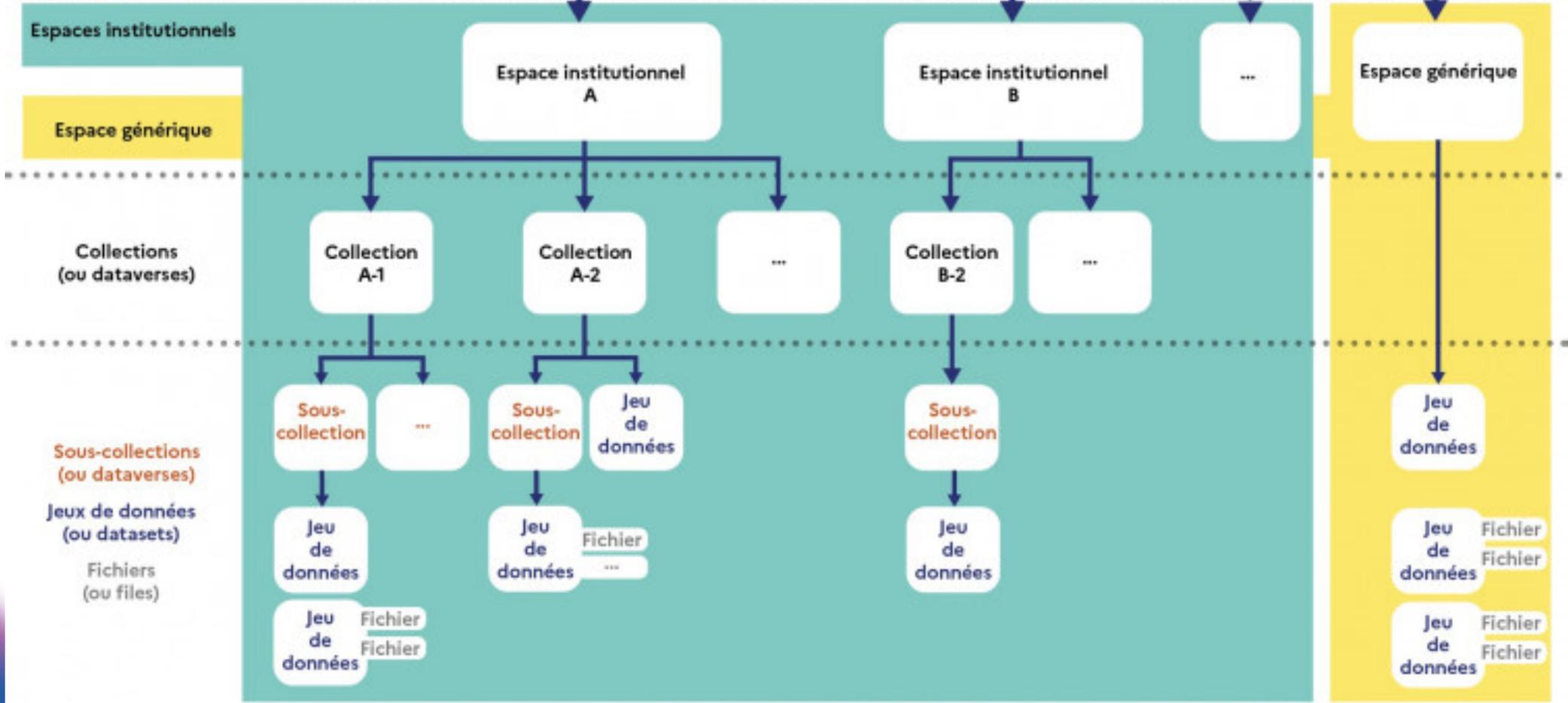
L'entrepôt de données de Recherche Data Gouv offre un entrepôt pluridisciplinaire souverain pour la publication des jeux de données. Cette offre s'adresse à toutes les communautés qui ne disposeraient pas encore d'un entrepôt thématique reconnu.

Quand les données sont déjà partagées ou ouvertes grâce à un entrepôt thématique français ou international, le catalogue de Recherche Data Gouv se charge de les repérer et de les signaler.

Organisation de l'entrepôt Recherche Data Gouv

Page d'accueil de l'entrepôt

Entrepôt Recherche Data Gouv



Qu'est-ce qu'un jeu de données ?

Un jeu de données est un ensemble cohérent de données brutes ou dérivées, auquel sont associées des métadonnées qui permettent de décrire ces données (présentation, date de publication, identifiant pérenne, mots-clefs, couverture géographique, couverture temporelle, etc.).

Les données qui composent un jeu de données peuvent être de natures différentes (textes, chiffres, images, etc.).

Un fichier peut être relié au jeu de données.

Un DOI est attribué automatiquement à chaque jeu de données déposé dans l'entrepôt Recherche Data Gouv.

Qu'est-ce qu'une collection ?

Une collection est un espace dédié aux jeux de données d'un collectif (laboratoire, projet de recherche, etc.) ou d'une thématique.

Chaque collection peut contenir :

- des sous-collections
- des jeux de données
- des fichiers

L'essentiel sur l'entrepôt Recherche Data Gouv

[https://recherche.data.gouv.fr/fr/categorie/9/guide/lessentiel-sur-lentrepo... 2](https://recherche.data.gouv.fr/fr/categorie/9/guide/lessentiel-sur-lentrepo...)



Glacier surface flow velocity products from satellite data (UGA, CNRS, IRD, INRAE, Grenoble-INP)

2 mars 2023



This collection p
collection is rela

Collection

Annual glacier surface fl

2 mars 2023 - G



Rabatel, Antoin
Sentinel-2 data

Jeu de données

Glacier ice flow velocity is
project is to create innovat

ALPES_wFLAG_wKT_A

2 mars 2023 - A



Network Comm
This file include
hydrological year

the hyd...

Fichier

Demande d'une collection pour une unité de recherche

Merci de compléter ce formulaire de demande d'une collection et de l'envoyer à admin-recherchedatagouv@inrae.fr, uniquement dans le cas où aucune de vos tutelles ne dispose d'un espace institutionnel dans Recherche Data Gouv.

Si l'une de vos tutelles dispose d'un espace institutionnel, merci de contacter directement les administrateurs de l'espace. Vous pouvez les contacter via « Contact » qui est affiché sur l'interface de l'entrepôt au niveau de cet espace.

Le centre de ressources entrepôt-catalogue vous adressera une réponse avec copie aux contacts des tutelles, précisant que sous 15 jours, à part opposition de l'un d'eux, la collection sera créée.

UNITÉ DE RECHERCHE

Intitulé de l'unité :

Nom(s) de la ou des tutelles :

Nom, Prénom du directeur de l'unité :

Adresse mail du directeur de l'unité :

Nom, Prénom du référent * :

Adresse mail du référent :

Avez-vous informé les tutelles de votre unité de votre demande ? :

Précisez les noms et les coordonnées des personnes informées au sein de vos tutelles :

-
-

*Personne en charge du suivi de la demande côté unité

INFORMATIONS SUR LES RESSOURCES DEDIEES A LA COLLECTION

Fournir une estimation des ressources qui pourront être dédiées à l'administration, l'accompagnement des déposants et la curation de la collection demandée.

Voici les préconisations du nombre d'ETP à mobiliser en fonction de la quantité de jeux de données déposés

- Hypothèse 1 - basse : moins de 25 dépôts de jeux de données à traiter par semaine
- Hypothèse 2 - moyenne : entre 25 et 50 dépôts de jeux de données à traiter par semaine
- Hypothèse 3 - haute : plus de 50 dépôts de jeux de données à traiter par semaine

	Hypothèse 1	Hypothèse 2	Hypothèse 3
Administration fonctionnelle + accompagnement des déposants	0,2	0,5	1
Curation	0,2	0,5	1

Ressources dédiées à la gestion de la collection

Hypothèse retenue : 1 2 3

Administration et accompagnement (nombre d'ETP) :

Curation (nombre d'ETP) :

DESCRIPTION DE LA COLLECTION

Nom de la collection :

Identifiant de la collection :

Recommandations générales :

o Intitulé court non ambigu.

Rédiger un fichier READ-ME pour accompagner un jeu de données et les fichiers qu'il contient est un bon moyen de **rendre l'information compréhensible et réutilisable**.

Modèle de fichier README : 

Le **READ-ME** permet d'**indiquer 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 les fichiers associés, ou des fichiers accessibles à long terme sur des services d'hébergement. Cela peut être la provenance, les critères de sélection, les conditions de réutilisation des données collectées.

Pour les données produites vous pouvez indiquer le contexte de création, les méthodes utilisées, les protocoles, le suivi et le contrôle de qualité par exemple. Plus généralement, indiquez le mieux possible les méthodes, protocoles, plan d'échantillonnages, dictionnaires des variables, équipements utilisés, les guides d'enquêteurs, technique, d'anonymisation, schéma de BDD etc.



Les Identifiants pérennes PID (*Permanent Identifiers*)



Digital Object Identifier. Attribué par Crossref pour les publications classiques et par **Datacite** pour les Données de la recherche (en France, par l'intermédiaire de l'**INIST**). Utilisé dans **Zenodo, Dryad, KNB...**

Systeme ARK

Archival Records and Knowledge . Système d'archivage open source et de préservation numérique. Orienté OAIS : les fonctions d'acquisition, d'archivage et de diffusion des données fonctionnent sur des paquets d'informations regroupant le contenu et les métadonnées dans des conteneurs contigus. Très utilisé par la **BNF**

Handle System®

Issu notamment Corporation for National Research Initiatives (CNRI). Présent dans les sites de l'univers **D-Space**, dans **Isidore, Nakala, Eudat...**



Un identificateur d'objet numérique (DOI) est une chaîne alphanumérique affectée à l'identification unique d'un objet. Il est lié à une description de métadonnées de l'objet ainsi qu'à un emplacement numérique, tel qu'une URL, où tous les détails concernant l'objet sont accessibles.



Pour créer de nouveaux DOI et les affecter à un contenu, il faut s'affilier à DataCite ou coopérer avec l'un de ses membres (l'INIST en France). Grâce à l'interface Web ou à l'API de l'entrepôt de métadonnées DataCite, on est en mesure de soumettre un nom, une description de métadonnées suivant le schéma de métadonnées DataCite et au moins une URL de l'objet pour créer un DOI. Une fois créées, les informations sur un DOI sont disponibles les différents services de Datacite, OAI-PMH et autres.

<https://www.datacite.org/>

Stratégie de partage des données et conservation à long terme



Cycle de vie et types de PID



Utilité des PID

PID Objet

PID Contributeur

Lier auteurs, institutions, publications, données et logiciels

EXEMPLE



Rendre visible

Partager

Découvrir

Citer

Vérifier



chacun des jeux de données !

Comment lier des données à un article avant publication dans une revue scientifique ?

Vos données ne sont pas encore finalisées ?



Vous pouvez **pré-réserver** un PID !

PID

1. Pré-réservez un PID pour vos données auprès de l'entrepôt de votre choix

2. Indiquez ce PID dans l'article à publier

PID

3. Quand les données seront finalisées, n'oubliez pas de les déposer dans l'entrepôt de données !



PID

4. Et quand l'article sera publié, indiquez le PID article dans l'entrepôt où sont déposées les données si celui-ci le permet

Vos données sont finalisées ?



Vous pouvez les déposer dans un **entrepôt** !



PID

1. Déposez vos données dans l'entrepôt de votre choix : un PID est attribué à vos données

PID

2. Indiquez ce PID dans l'article à publier

Vos données ont été valorisées par la publication d'un data paper ?



Vos données ont donc **déjà** été déposées dans un **entrepôt** !

PID



PID

PID

Indiquez le PID attribué à vos données par l'entrepôt + le PID du Data paper dans l'article à publier



R comme Réutilisable (1)

Augmenter la réutilisation des données (grâce à des licences explicites)*

- ❖ Précisez les modalités d'autorisation de partage des données avec pour objectif de permettre leur réutilisation la plus large possible
- ❖ Précisez quand les données seront disponibles pour réutilisation. Le cas échéant, précisez pourquoi et pour quelle période un embargo sur les données est nécessaire

**On fera plus loin allusion aux licences Creative Commons, ODBL et éventuellement Etalab*

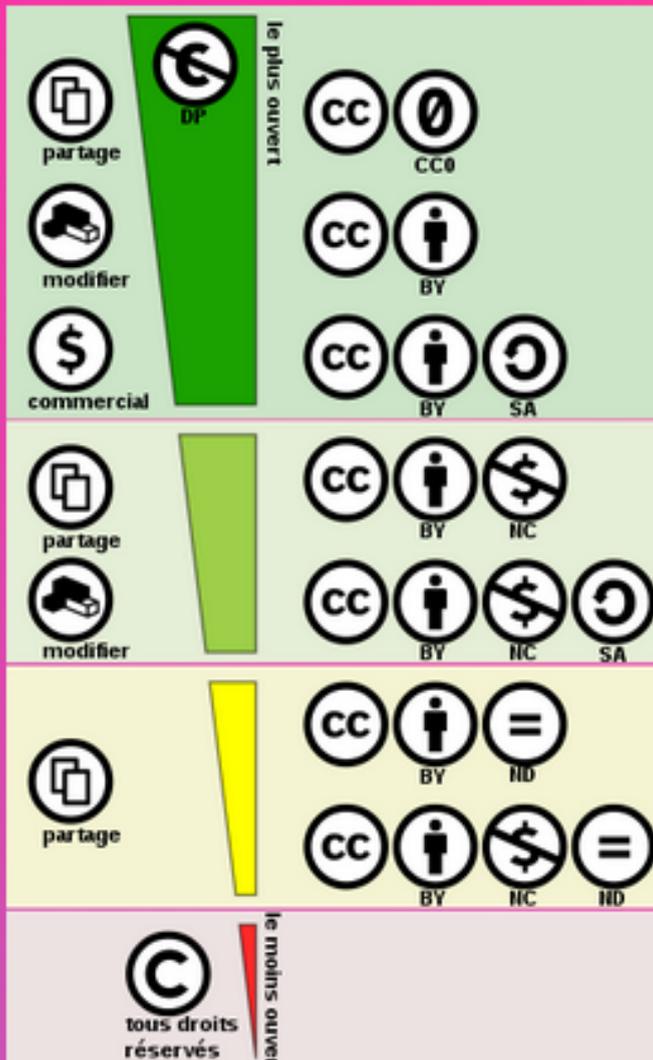




R comme Réutilisable (2)

- Préciser si les données produites et / ou utilisées dans le projet peuvent être utilisées par des tierces parties, en particulier après la fin du projet ?
- Si la réutilisation de certaines données est restreinte, expliquez pourquoi
- Décrire les processus de garantie de la qualité des données
- Spécifiez la durée pendant laquelle les données resteront réutilisables

5. Stratégie de partage, ouverture des données et préservation à long terme



Simon Villeneuve, Les familles de licences Creative Commons

Les licences de diffusion

Sans licence, les données ne sont pas véritablement ouvertes.

Une licence ouverte garantit à tous le droit d'**utiliser**, de **partager** et d'**accéder** à vos données avec la **sécurité juridique** nécessaire aux producteurs et aux réutilisateurs des données.

Outil pour choisir sa licence CC

Rubrique 3. Stockage et sauvegarde des données en cours

A propos des licences

Francine Filoche, Chargée de mission données de la recherche

- Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte

Les licences de réutilisation dans le cadre de l'



et de la loi pour une République numérique

Publié le 22/11/2018
Mis à jour le 26/01/2023

Stratégie de partage des données et conservation à long terme



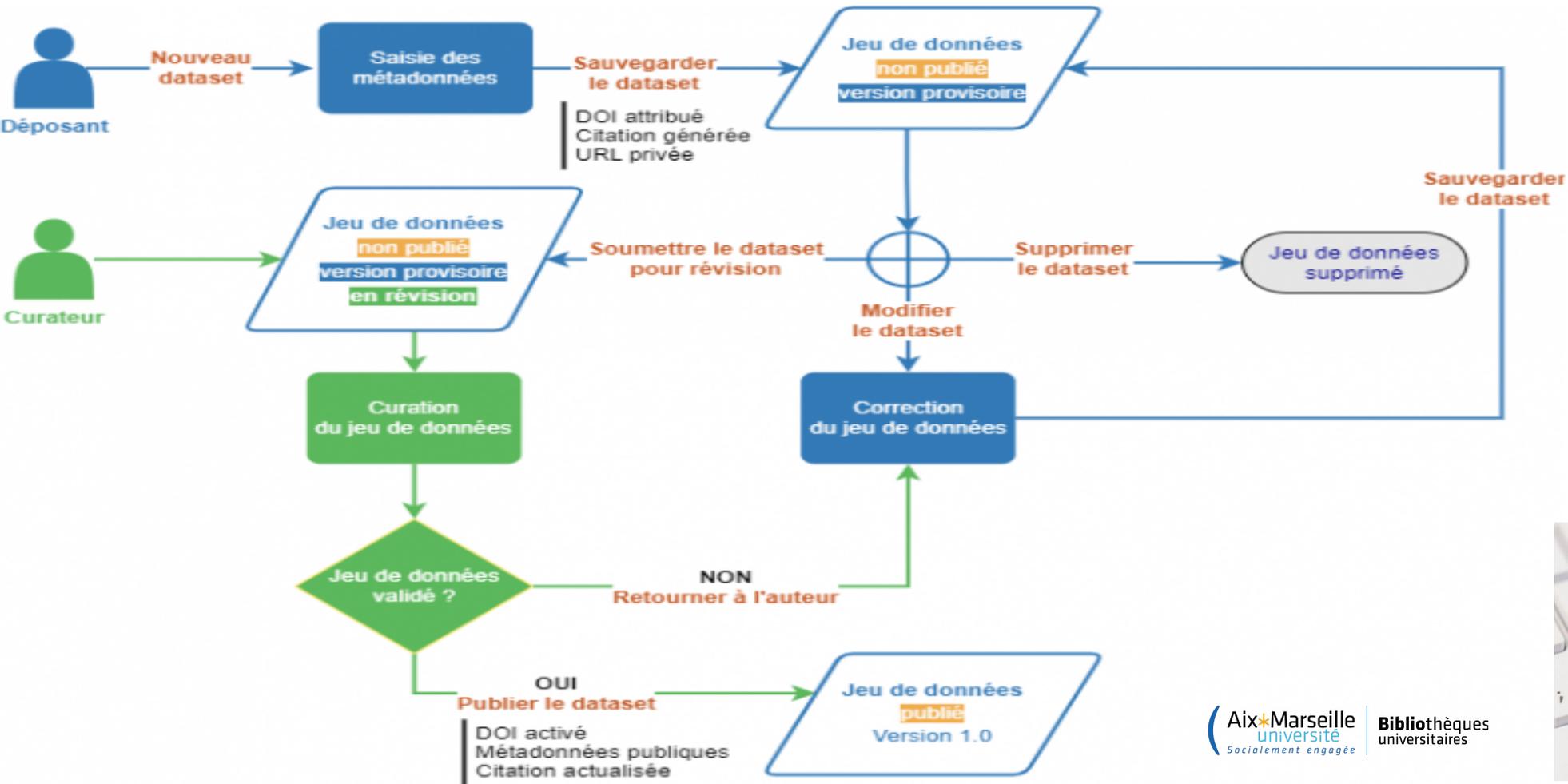
Pourquoi déposer dans un entrepôt ?

- Mise à disposition des données
- Réutilisation des données par le déposant et ses pairs
- Citabilité des données
- Visibilité accrue des données et des chercheurs
- Transparence des données
- Lien bidirectionnel entre des jeux de données et un article (ou data paper)
- Conservation des données

Publication d'un jeu de données avec curation a priori

Le déposant a un rôle de collaborateur pour les jeux de données qu'il crée dans la collection et ne peut pas les publier directement. Il doit les soumettre à un curateur de la collection pour validation et publication.

Première version d'un jeu de données



Stratégie de partage des données et conservation à long terme : exemple dans RechercheDataGouv

<https://entrepot.recherche.data.gouv.fr/dataverse/espacegenerique>





Contexte et cycle de vie des données

Réutilisation et citation des données (chercheur)

Réutilisation et citation des données (réutilisateurs)

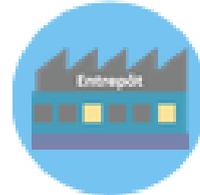
Data paper et Data journal

Exposition et visualisation des données

■ RÉUTILISATION ET CITATION DES DONNÉES 1/2

■ DU CÔTÉ DU CHERCHEUR

Pour que les données de recherche qu'il a produites soient réutilisées dans de bonnes conditions, le chercheur doit adopter plusieurs bonnes pratiques :



Rendre ses données FAIR

Documenter ses jeux de données avec un fichier lisez-moi.txt

Les déposer dans un entrepôt

Appliquer une licence de diffusion

Bien renseigner les métadonnées

Associer le(s) logiciel(s) nécessaire(s) à leur lecture et compréhension

Appliquer un identifiant pérenne



Contexte et cycle de vie des données

Réutilisation et citation des données (chercheur)

Réutilisation et citation des données (réutilisateurs)

Data paper et Data journal

Exposition et visualisation des données

■ RÉUTILISATION ET CITATION DES DONNÉES 2/2

■ DU CÔTÉ DES RÉ-UTILISATEURS

Il existe plusieurs manières pour un chercheur de trouver des jeux de données réutilisables :



Survoler pour voir le détail

Dans tous les cas, il est à noter que les ré-utilisateurs doivent s'engager à respecter certaines règles :

- 1 Respecter la propriété intellectuelle des auteurs telle que mentionnée dans la licence
- 2 Citer les données si la licence l'exige (il est recommandé de toujours citer ses sources)
- 3 Lier les données aux publications.



Contexte et cycle de vie des données

Réutilisation et citation des données (chercheur)

Réutilisation et citation des données (réutilisateurs)

Data paper et Data journal

Exposition et visualisation des données

DATA PAPER ET DATA JOURNAL

La rédaction et la publication d'un **data paper** sont une bonne façon pour un chercheur de valoriser ses données de recherche.

Cliquer pour voir le détail

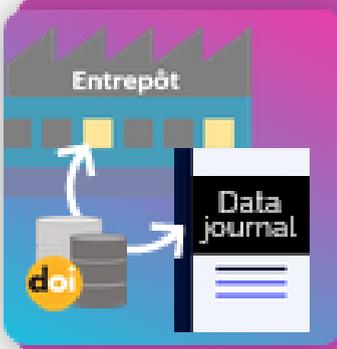


Qu'est-ce qu'un Data Paper ?
Qu'est-ce qu'un Data journal ?

Exemple de data paper



Exemples de data journals



Entrepôt

doi

Data journal

Comment lier les données au Data Paper ?

Un data paper peut être publié soit dans un **data journal** (revue dédiée à ce type de publication), soit dans une **revue classique** qui accepte les data papers.

Conclusion : Allier Open Data et FAIR data

Open Data

- contenus librement accessibles
- peuvent être utilisés, modifiés et partagés par n'importe qui dans n'importe quel but
- mis à disposition dans le cadre d'une licence ouverte
- ne doivent pas imposer de restrictions au delà de CC-BY et CC-BY-SA

ET

FAIR Data

- données qui suivent les principes FAIR
- lisibles et utilisables par des humains et des machines
- prise en compte des questions juridiques et éthiques
- outils et technologies adéquats

Merci pour votre attention

Contact spécifique pour toute demande de RDV pour une relecture
et ou
un accompagnement à la rédaction d'un PGD dans DMP OPIDOR :

scd-dmp-assistance@univ-amu.fr
francine.filoche@univ-amu.fr

Autres interrogations **en matière de gestion de données :**

guichet-donnees.univ-amu.fr

Pour toute demande concernant la Science Ouverte et le dépôt dans HAL,
contactez la Cellule Science Ouverte du SCD AMU :

scd-science-ouverte@univ-amu.fr

Francine Filoche, Chargée de mission données de la recherche



- Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte



Aix-Marseille
université
Socialement engagée

Bibliothèques
universitaires

AMU, le CNRS, le CEA, la MMSH, Open Edition, la PUD-AMU, le SCD et CEDRE (ainsi que l'ensemble de ses partenaires)
ont le plaisir de vous **annoncer**

le lancement d'un nouveau service dédié à la gestion et à l'ouverture des données de Recherche :
le Guichet de la Donnée du site d'Aix-Marseille (GDsAM).

Le GDsAM répond à tous besoins d'informations, de conseils, de formations, d'accompagnements juridique ou opérationnel,
et même de soutiens directs aux projets de recherche de vos laboratoires.

Le guichet vous accompagne également, pour tout ce qui concerne la gestion des données de Recherche
en vue de leur FAIRisation pour faciliter une future réutilisation.

Ce guichet permet de mettre en contact des experts du cycle de vie de la donnée de Recherche
quels que soient la discipline et le statut du demandeur
(chercheur, enseignant-chercheur, ingénieur, étudiant)

qui en éprouve le besoin dans le cadre d'un projet, qu'il soit disciplinaire ou interdisciplinaire.

GDsAM permet de créer le dialogue entre experts et demandeurs à travers une adresse unique :
guichet-donnees.univ-amu.fr

GDsAM fait partie du réseau des Ateliers de la Donnée, dispositif essentiel de la politique nationale de la Science Ouverte. Le GDsAM est
labellisé « en trajectoire » par Recherche Data Gouv.(<https://recherche.data.gouv.fr/fr>), l'écosystème français de la donnée de recherche
fortement soutenu par le ministère de l'Enseignement Supérieur et de la Recherche.

Pour toute information complémentaire, **contactez : Gdsam-contact@groupe.renater.fr**



Quelques ressources en plus 1/3

- **re3data** : registre mondial des entrepôts de données de recherche
- **Recherche Data Gouv** : Voir DAT'AMU dans l'entrepôt national français interdisciplinaires à choisir si pas d'entrepôt disciplinaire
- **RDA Metadata Directory** : un répertoire de standards dans différentes disciplines
- **DCC** (Digital Curation Centre)
- Guide pour la recherche, CNRS INSHS, 2021
- **FAIRSharing** : un répertoire de standards en science de la vie
- Coûts de la gestion de données, DORANUM
- **Espace stockage humaNum** : Nakala / **DataDryad**
- Choisir un entrepôt de confiance
- **Licences Creatives Commons**
- **Comprendre le RGPD, CNIL**
- **Protection des données personnelles et RGPD dans la recherche** : conséquences, obligations, implications
- **Liste indicative de formats de fichiers**
- **Data Management Plan Catalogue**



Quelques ressources en plus 2/3

- **Décret n°2021-1572 du 3 décembre 2021** relatif au respect des exigences de l'intégrité scientifique (PGD art. 6)
- Code de la recherche, article L533-4
- Code de la recherche, Chapitre 2, articles L112- à L112-6 (Organisation de l'accès libre aux données de la recherche)
- Guide d'analyse du cadre juridique en France sur l'ouverture des données de la recherche
- Arbre de décision, Puis je diffuser mes données ? , <http://www.bibliotheque-numerique.fr/DonneesDiffusables.php>
- Charte française de deontologie des metiers de la recherche
- Choisir sa licence Creative Commons :
- Guide pratique de la publication en ligne et de la réutilisation des données publiques (open data), CADA/ CNIL
- Guide de santé de la CNIL
- Analyse d'impact sur la vie privée, évaluation des traitements de données
- Site Ouvrir la science, Ministère de l'Enseignement supérieur et de la Recherche, France
- Métadonnées, standards, formats, , DORA NUM
- Faire entrer la science ouverte dans son projet ANR : un guide pratique
- Espace d'entraînement DMP PIDOR pour la rédaction de PGD
- Page d'accès création d'un compte pour l'espace d'entraînement DMP OPIDOR



- **Livret** d'aide à l'acculturation des enjeux de l'ouverture des données, Jeu "A l'asso", AMU 
- Glossaire des données de la recherche 
- Passeport données de la recherche, science ouverte, gouv.fr 
- Ouverture des données : ...aussi ouvert que possible...aussi fermé que nécessaire, Dipso, INRAE 
- Le point sur les données de la recherche, BD, Université Guyane Française 
- Accès aux guichet de la donnée AMU 
- Tutoriel Données de la recherche , SCD AMU 
- Comité éthique AMU 
- Charte Science Ouverte AMU 

Merci pour votre attention

Contact spécifique pour toute demande de RDV pour une relecture
et ou
un accompagnement à la rédaction d'un PGD dans DMP OPIDOR :

scd-dmp-assistance@univ-amu.fr francine.filoche@univ-amu.fr

Autres interrogations en matière de gestion de données :

guichet-donnees.univ-amu.fr

Pour toute demande concernant la Science Ouverte et le dépôt dans HAL,
contactez la Cellule Science Ouverte du SCD AMU :

scd-science-ouverte@univ-amu.fr

Francine Filoche, Chargée de mission données de la recherche

 - Département ingénierie documentaire du SCD AMU _ Cellule Science Ouverte



Aix*Marseille
université
Socialement engagée

Bibliothèques
universitaires