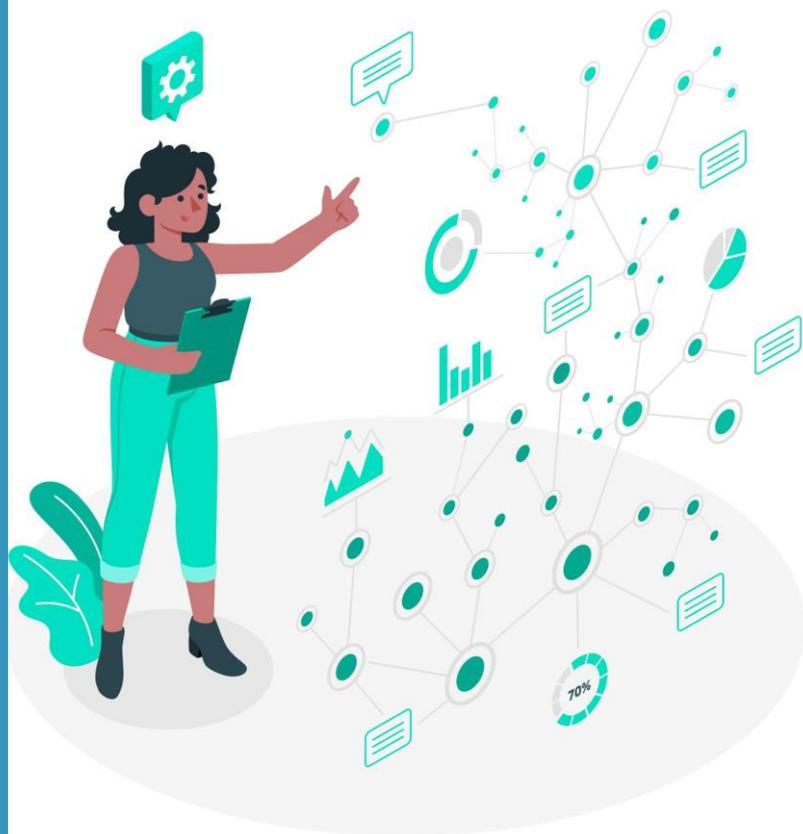


FORMER LES
PROFESSIONNELS DE
L'INFORMATION ET
DE LA
DOCUMENTATION
AUX DONNÉES DE LA
RECHERCHE EN
45 MINUTES

2020



Une production du gTIGRE (groupe de travail inter-établissements Grand Est)



Les auteurs

- Laetitia Bracco (Université de Lorraine)
- Mathilde Barthe (Médial)
- Stéphanie Cheviron (Université de Strasbourg)
- Agnès Faller (Université de Reims Champagne-Ardenne)
- Madeleine Hubert (Bibliothèque nationale et universitaire de Strasbourg)
- Sylvie Steffann (Université de Haute-Alsace)
- Noël Thiboud (URFIST de Strasbourg)
- Jean-Baptiste Vu Van (Université de technologie de Troyes)

Présentation de la formation

Ce document est un outil clef en main pour réaliser, en interne, une formation basique aux données de la recherche en 45 minutes. Elle a été pensée pour des personnels des bibliothèques ou centre de documentation qui sont novices dans le domaine et ont besoin d'être sensibilisés. En effet, dans le monde universitaire, les données sont un élément à connaître, au moins de nom, par tout agent en situation d'accueil dans une bibliothèque.

Les parties « l'essentiel » permettent de former des agents aux bases en 45 minutes. La formation peut être étendue à 1h30 en utilisant également les parties « pour développer ». Il est également possible de faire la formation en 45 minutes puis de distribuer la trame pour que les participants lisent les parties « pour développer » ultérieurement.

Un quizz est proposé à la fin pour évaluer l'acquisition des connaissances. Il existe également une version de ce quizz directement intégrable dans Moodle, déposée en même temps que la présente trame de formation dans Zenodo. Les questions peuvent être posées en une fois à la fin ou distillées au fur et à mesure de la formation, à la convenance de chaque formateur.

Compétences visées :

Avoir des notions sur :

- Le paysage de la recherche en France
- Les enjeux de la Science Ouverte et le cadre réglementaire
- La définition des données de la recherche
- Les principaux concepts sur les données (DMP/PGD, FAIR, entrepôts et moissonnage, métadonnées, jeu de données, identifiants pérennes...)
- Les principales questions pouvant être posées en banque de prêt

Connaître dans son établissement :

- La politique de l'établissement en matière de Science Ouverte
- Les personnes référentes, pour pouvoir réorienter l'utilisateur le cas échéant
- Les services proposés par l'établissement (archive ouverte, entrepôt, ...)

Public visé : toute personne amenée à faire du service public.

Ce document est librement réutilisable selon la licence CC-BY.



Table des matières

1. Avoir des notions sur.....	3
1.1. Le paysage de la recherche en France	3
a. L'essentiel.....	3
b. Pour développer	4
1.2. Les enjeux de la Science Ouverte et le cadre réglementaire	6
a. L'essentiel.....	6
b. Pour développer	6
1.3. Ce que sont les données de la recherche.....	7
a. L'essentiel.....	7
b. Pour développer	8
1.4. Les principaux concepts à connaître sur les données de la recherche.....	9
2. Accompagner/aider : savoir répondre aux principales questions posées à l'accueil	11
2.1. Comment trouver des jeux de données	11
a. L'essentiel.....	11
b. Pour développer	11
2.2. Comment les citer.....	12
a. L'essentiel	12
2.3. Comment faire un plan de gestion des données (PGD ou DMP)	12
a. L'essentiel	12
3. Dans mon établissement	13
3.1. La politique de l'établissement en matière de Science Ouverte	13
3.2. Les personnes référentes, pour pouvoir réorienter l'utilisateur le cas échéant	14
4. Evaluer ses connaissances	15

1. Avoir des notions sur...

1.1. Le paysage de la recherche en France

a. L'essentiel

Financée sur projets, la recherche publique en France bénéficie des crédits des universités, des organismes de recherche publics et des agences de financement, dont l'Agence nationale de la recherche (ANR), ou encore de dotations provenant de l'Union européenne.



Depuis peu, ces financeurs (ANR et Horizon 2020, le programme-cadre européen de financement de la recherche et de l'innovation 2014-2020) exigent que les porteurs de projets appliquent les principes de la "Science Ouverte" (*Open Science*), à savoir le dépôt des publications dans une archive ouverte, la rédaction d'un plan de gestion des données de la recherche et, si possible, l'ouverture des données de la recherche.

La recherche est évaluée par des instances nationales ou européennes (les mêmes qui allouent les crédits) mais aussi et surtout par les pairs (les chercheurs et experts de la même discipline) dans le cadre des publications scientifiques. Publier ses travaux dans les revues ou collections reconnues de sa discipline est décisif pour faire valider les résultats de ses recherches, puis les diffuser et les faire connaître à la communauté scientifique, mais aussi obtenir la reconnaissance de ses pairs et progresser dans sa carrière.

Qu'est-ce que la Science Ouverte ?

Cette expression désigne l'ensemble des principes qui contribuent à faciliter l'accès aux données et aux résultats de la recherche financée sur fonds publics. Si les principaux enjeux sont l'accès

libre (sans la barrière de l'abonnement à la revue ou du paiement de l'article) aux publications et aux jeux de données, la Science Ouverte consiste aussi à développer de nouvelles infrastructures sur des formats dits ouverts, à revoir les principes qui fondent la bibliométrie (dont est notamment issue la notion de facteur d'impact) en forgeant de nouveaux indicateurs (ou métriques), de nouveaux principes pour l'évaluation et la gestion de la carrière des chercheurs, ou encore à favoriser la "science participative" (*Citizen Science*).



b. Pour développer

La recherche publique est essentiellement financée sur projets. Obtenir des financements pour effectuer des travaux de recherche est une préoccupation majeure des unités de recherche et des chercheurs afin de prendre en charge des frais de mission (congrès, colloques, autres), mettre en place ou remplacer des appareils, investir dans des équipements scientifiques lourds ou des infrastructures de recherche, développer des applications, faire appel à de nouvelles compétences (temps ingénieur, temps technicien et autres), mais aussi pour faire connaître les résultats de la recherche par le biais des publications.

La publication, décisive pour le chercheur ou l'équipe de recherche, joue un triple rôle de validation de la découverte ou des résultats obtenus (c'est le comité scientifique de la revue ou de la collection qui joue ce rôle d'instance de validation), de reconnaissance par les pairs (des chercheurs de la même discipline qui vont citer cette publication) et de communication auprès de la communauté scientifique. On publie pour prendre date (attester de l'antériorité d'une découverte), pour contribuer à l'avancée des connaissances scientifiques sur une problématique, mais aussi pour faire connaître ses travaux à sa communauté (voire au-delà) et être cité.

Les équipes de recherche labellisées (UR (ex-EA) ou UMR) fonctionnent grâce aux crédits alloués par les universités (la partie recherche de la dotation de fonctionnement), les organismes de recherche publics (le CNRS pour les UMR) ou les agences de financement, dont l'Agence nationale de la recherche (ANR), créée en 2005. Elles peuvent aussi recevoir d'autres dotations provenant des régions, d'associations caritatives, de l'industrie et de l'Union européenne.

Les agences de financement subventionnent des projets qui sont sélectionnés selon des critères internationaux et qui s'inscrivent dans les priorités nationales des différents champs de la recherche et de l'innovation. Ces agences prennent en charge tout ou partie de ces dépenses (on parle de dépenses éligibles), mais, en contrepartie, elles imposent des règles, notamment ces dernières années, qui intègrent le respect des principes de la Science Ouverte. En particulier l'ANR et Horizon 2020, le programme-cadre de financement de la recherche et de l'innovation de l'Union européenne (2014-2020), exigent le dépôt des publications dans une archive ouverte et la rédaction d'un plan de gestion des données de la recherche.

Une fois qu'elles ont obtenu les crédits nécessaires au financement de leurs travaux ou projets, les équipes de recherche sont tenues de rendre compte de la manière dont elles ont employé ces subventions et sont évaluées dans ce sens. En France, l'organisme chargé de l'évaluation de la recherche publique est le Hcéres (Haut conseil de l'évaluation et de la recherche et de l'enseignement supérieur). La publication des travaux figure parmi les principaux critères d'évaluation. Publier les résultats de ses travaux peut avoir un coût, surtout si l'on veut que ces derniers soient largement reconnus et cités. Les chercheurs et unités de recherche sont évalués sur la quantité, mais aussi et surtout sur la qualité et l'impact (ou rayonnement) de leurs publications.

L'objectif est souvent de publier ses travaux dans une revue à fort impact (cet impact est mesuré par des indicateurs comme l'*Impact Factor*, *CiteScore* et autres méthodes de calcul) afin d'obtenir des citations (le fait pour un article d'être cité par d'autres articles). Cette méthode d'évaluation de la recherche fait l'objet de nombreux débats et de réflexions dans le cadre de la Science Ouverte. La soumission d'un article à un comité de lecture peut avoir un coût (ce qu'on appelle les *Article Processing Charges* ou frais de traitement des articles appliqués par les éditeurs). Paradoxalement, publier dans une revue totalement ou partiellement accessible en ligne a aussi un coût.

Les revues en *Open Access* (gratuité pour le lecteur final) peuvent occasionner des frais de publication pour l'auteur. Si elle fait l'économie des frais d'impression, de stockage et de

diffusion, l'édition numérique occasionne des coûts importants de maintenance des serveurs et de développement des applications. Dans le secteur des sciences humaines et sociales (SHS), les éditeurs demandent des contributions importantes aux porteurs de projets d'édition pour la parution des ouvrages (actes de colloque, ouvrages collectifs ou monographies).

Une des préoccupations des agences de financement est de contenir ces coûts en développant les principes de la Science Ouverte.

Bref glossaire de la recherche :

ANR : placée sous la tutelle du ministère de l'enseignement supérieur, de la recherche et de l'innovation (MESRI), l'Agence nationale de la recherche est l'agence française de financement de la recherche sur projets. Elle a pour missions entre autres de "financer et promouvoir le développement des recherches fondamentales et finalisées, l'innovation technique et le transfert de technologies ainsi que le partenariat entre secteur public et le secteur privé". Elle exerce aussi une mission d'évaluation de l'offre de recherche et de l'impact des subventions qu'elle distribue sur la production scientifique nationale.

CNRS : placé sous la tutelle du ministère de l'enseignement supérieur, de la recherche et de l'innovation (MESRI), cet EPST (établissement public à caractère scientifique et technique) est le plus grand organisme public français de recherche scientifique. Le Centre national de la recherche scientifique est organisé en instituts eux-mêmes organisés en un millier d'unités de recherche et de service labellisées. Ces unités sont pour la plupart gérées avec d'autres structures, dont les universités, et sont labellisées pour une durée de cinq sous la forme d'unités mixtes de recherche (UMR).

UR (anciennement EA) : une unité de recherche (ex-équipe d'accueil) est la dénomination administrative d'une structure de recherche qui regroupe les chercheurs travaillant sur une thématique de recherche commune ou un même axe de recherche. Ces unités sont "labellisées" par le MESRI duquel elles obtiennent également un financement et par lequel elle sont évaluées tous les cinq ans.

Hcéres : « le Haut Conseil de l'évaluation de la recherche et de l'enseignement supérieur (Hcéres) est l'autorité administrative indépendante chargée d'évaluer l'ensemble des structures de l'enseignement supérieur et de la recherche, ou de valider les procédures d'évaluations conduites par d'autres instances » (source : site web du Hcéres).

UMR : unité mixte de recherche, unité de recherche cogérée par un établissement d'enseignement supérieur (comme une université) et un organisme de recherche (comme le CNRS), labellisée pour cinq ans.

1.2. Les enjeux de la Science Ouverte et le cadre réglementaire

a. L'essentiel

Émanation de la communauté scientifique, en particulier de la communauté des physiciens dès les années 1990 et relayé à présent par les institutions, le mouvement de la Science Ouverte poursuit les objectifs suivants :



- **Communiquer** directement les résultats de ses travaux à ses pairs et au grand public (publications et données) ;
- **Accroître la visibilité de ses travaux** pour augmenter leur citabilité et leur impact ;
- **Améliorer la qualité du travail de recherche par une meilleure gestion des données produites** (description, partage, réutilisation) ;
- Pour les investisseurs publics, **reprendre en main le circuit de la diffusion des résultats de la recherche** afin de contrer les tendances monopolistiques des grands groupes éditoriaux et d'ouvrir au contribuable l'accès aux résultats de la recherche (Loi pour une République numérique ; Plan pour la Science Ouverte) ;
- Cultiver la transparence et les principes de l'intégrité scientifique afin de lutter contre les fraudes scientifiques et de renforcer la confiance des citoyens dans la recherche.

b. Pour développer

Parti de la communauté scientifique, le mouvement de l'*Open Science* (Science Ouverte) correspond d'abord à une ouverture de la recherche en direction des pairs, qui est à présent relayée par les gouvernements et les institutions publiques pour une ouverture en direction de la société (acteurs économiques et acteurs de la société civile).

En France, l'article 30 de la loi pour une République numérique promulguée le 7 octobre 2016 donne toutes les garanties nécessaires au chercheur pour qu'il puisse déposer ses travaux (articles) dans une archive ouverte. Puis, le Plan national pour la Science Ouverte présenté par la ministre Frédérique Vidal, le 4 juillet 2018, vise l'accès ouvert pour les publications et pour les données issues de recherches financées sur projets. Ce plan entend « mettre en place une obligation de diffusion en accès ouvert des articles et ouvrages issus des recherches financées par appel à projets sur fonds publics ».

Mais au-delà du partage et de la réutilisation des publications et des données, ce mouvement s'étend aussi aux questions relatives à l'évaluation (métriques et impact) et à la médiation scientifique (vulgarisation, *Citizen Science*, etc.). Le deuxième axe du Plan national pour la Science Ouverte concerne les données de la recherche. Il oblige notamment les porteurs de projets financés sur fonds publics à diffuser ouvertement leurs données (sauf exceptions légales) et demande la généralisation des Plans de gestion de données (PGD).

L'élaboration d'un plan de gestion de données est déjà obligatoire pour les projets ANR depuis 2019 ; de même pour les projets inscrits dans le programme Horizon 2020. Pour publier un article, certains éditeurs peuvent aussi exiger l'accès aux données qui ont permis d'aboutir aux résultats de la recherche présentée dans l'article soumis (par exemple, l'éditeur [PLOS ONE](#)).

Certains établissements exigent aussi à présent que leurs équipes de recherche appliquent les principes de la Science Ouverte (les données brutes ou traitées de la recherche sont la propriété de l'établissement). Plus largement, nous voyons de plus en plus d'établissements apporter un soutien financier à l'édition ouverte, nommer des chargés de mission Science Ouverte, adhérer aux mouvements et manifestes nationaux et internationaux et rédiger des chartes ou politiques favorables à la Science Ouverte.

La question des données de la recherche vise aussi à améliorer les pratiques de recherche en réfléchissant notamment à leur description, à leur structuration et à leur conservation à long terme. Le partage des données (quand cela est possible) permet aussi leur réutilisation (et donc un gain de temps et une réduction des coûts). C'est bien pour répondre à ces différents objectifs que les porteurs de projet sont dorénavant incités à rédiger un Plan de gestion des données (PGD ou DMP = *Data Management Plan*). Loin d'être une contrainte administrative de plus, l'élaboration d'un PGD est un outil précieux pour la gestion d'un projet de recherche.

Liens utiles

- Plan national pour la Science Ouverte : <https://www.ouvri.la-science.fr/plan-national-pour-la-science-ouverte/>
- Comité pour la Science Ouverte (CoSo) : <https://www.ouvri.la-science.fr/presentation-du-comite/>
- Loi pour une République numérique : <https://www.ouvri.la-science.fr/guide-application-loi-republique-numerique-article-30-ecrits-scientifiques-version-courte/>

1.3. Ce que sont les données de la recherche

a. L'essentiel



Plusieurs définitions coexistent. En voici deux qui sont couramment utilisées :

- Définition [COOPIST-CIRAD](#) : “Les données de la recherche, ou données de recherche, ou données scientifiques, peuvent être définies comme l'ensemble des informations collectées, observées ou créées sous une forme numérique dans le cadre d'un projet de recherche”.
- Définition Couperin : “ensemble des informations scientifiques produites ou collectées dans le cadre d'un projet de recherche, les données de la recherche

peuvent être des photos, des mesures, des sons, etc. Elles sont nécessaires comme éléments probants afin de valider les résultats de la recherche et doivent être accompagnées d'informations qui les documentent, telles que des protocoles expérimentaux, des descriptifs méthodologiques ou des métadonnées. Ces données peuvent être diffusées dans des entrepôts généralistes ou spécialisés” (source : [Améliorer son projet ANR grâce à la Science Ouverte](#), p.4-5)

Elles sont de natures diverses : textuelles, numériques, multimédia, informatiques... Et de niveaux d'élaboration différents : brutes, structurées ou analysées. Un exemple : les photos de vos vacances sont un jeu de données brutes ; une fois que le lieu est ajouté dans la description de chaque fichier, c'est un jeu de données structurées ; une fois qu'elles ont été triées et que seulement certaines ont été gardées pour réaliser un album, c'est un jeu de données analysées.

Les données de la recherche sont collectées, observées ou obtenues par expérimentation à des fins d'analyse, pour produire et valider les résultats d'une recherche originale.

b. Pour développer

Une définition plus complète est fournie par l'[OCDE](#) (*Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, 2007) :

« 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. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche ».

On peut classer les données de la recherche en fonction de leur nature ou forme :

- Texte (questionnaires, retranscription d'entretien, procédures, protocoles, journal de bord) ;
- Numérique (réponses à des tests ou enquêtes, données médicales, mesures et relevés) ;
- Multimédia (sons, vidéos, images), codes de développement (codes sources, algorithmes, scripts) ;
- Propres à des disciplines spécifiques.

En fonction de leur niveau d'élaboration :

- Données brutes (donnée collectées, mais non organisées et non mises en forme) ;
- Données traitées (elles ont été organisées, structurées pour être prêtes à être analysées) ;
- Données analysées (telles qu'elles apparaissent dans les publications).

Pourquoi s'intéresse-t-on tant à ces données ?

Les données constituent la matière première de l'activité de recherche. Ce sont les données tangibles qui permettent d'élaborer de nouvelles hypothèses, d'en consolider d'autres, de faire progresser les connaissances sur tel ou tel domaine de recherche. Leur bonne gestion est donc essentielle pour :

- Ne pas les perdre : la sécurisation des données à l'aide d'outils et de solutions d'hébergement dédiées est un enjeu important ;
- Garantir leur origine : la donnée a un cycle de vie qu'il convient de documenter à chaque étape du parcours ;
- Les retrouver : les données peuvent représenter des masses d'information qu'il convient d'organiser et de structurer ;
- Les exploiter : le chercheur à l'origine de la collecte des données pourra y revenir et ouvrir leur accès à d'autres chercheurs.

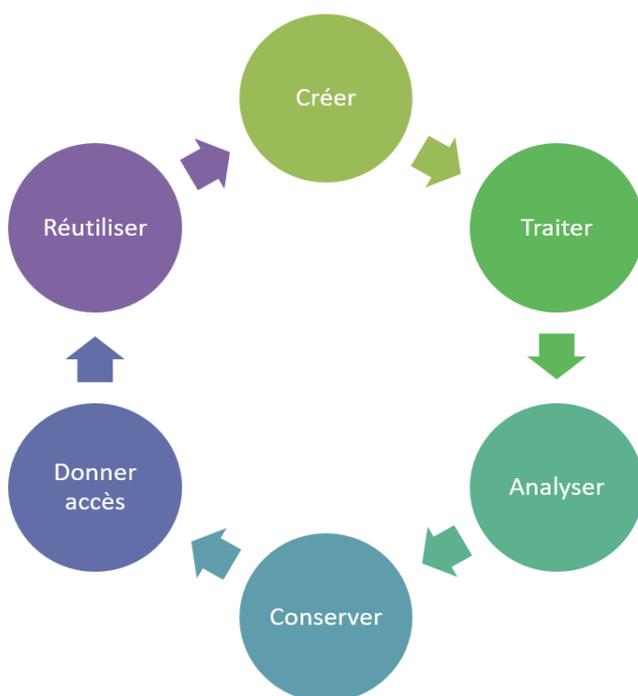
1.4. Les principaux concepts à connaître sur les données de la recherche

Toutes ces notions sont essentielles et font donc partie des 45 minutes de la formation basique.



Archive ouverte : plateforme où les documents déposés sont en accès ouvert. Elle peut être nationale (comme HAL – Hyper Article en Ligne), institutionnelle ou disciplinaire (comme arXiv). Les chercheurs peuvent y déposer une notice décrivant leurs publications (titre, auteurs, résumé...) et y associer ou nom le fichier du texte intégral selon les droits dont ils disposent. Les archives ouvertes ne sont pas destinées à recevoir des données de la recherche, sauf cas très particulier (source : [Améliorer son projet ANR grâce à la Science Ouverte](#)).

Cycle de vie des données : le cycle de vie des données correspond aux différentes étapes au cours desquelles, au sein d'un projet de recherche, les données sont élaborées ou transformées.



Data Management Plan (DMP) ou plan de gestion des données de la recherche (PGD) : document unique qui décrit, pendant et à l'issue du projet de recherche, les étapes de production, de traitement, de description, de diffusion ou protection et d'archivage (conservation) des données. Il est demandé par tous les financeurs de la recherche. Le chercheur qui coordonne le projet est généralement le rédacteur du PGD, mais il ne l'écrit pas seul : il peut solliciter les services d'appui à la recherche pour l'aider.

Data paper : publication qui décrit un jeu de données scientifiques, notamment à l'aide d'informations structurées appelées métadonnées. Contrairement aux articles de recherche classiques, les *data papers* fournissent une voie formalisée au partage des données plutôt que tester des hypothèses ou présenter de nouvelles analyses. (Source : <https://doranum.fr/data-paper-data-journal/fiche-synthetique/>).

Entrepôts : les entrepôts sont des plateformes sur lesquelles sont déposés, décrits et conservés des jeux de données de la recherche. Les entrepôts sont généralistes ou disciplinaires. (Source : [Améliorer son projet ANR grâce à la Science Ouverte](#)).



FAIR : principes d'ouverture des données, qui visent à les rendre Faciles à trouver (Findable), Accessibles (Accessible), Interopérables (Interoperable), Réutilisables (Reusable). (Source : [Améliorer son projet ANR grâce à la Science Ouverte](#)).

Identifiant pérenne ou « PIDs » (*Persistent IDentifiers*) : chaîne de caractères attribuée de manière pérenne à un objet. Il en existe 2 grands types : les identifiants objet pour les productions scientifiques (publications et données) et les identifiants contributeur pour les auteurs et les institutions. On parle souvent de PID pour *Persistent IDentifier* en anglais. Un exemple : le DOI (*Digital Object Identifier*) pour les articles. (Source : <https://doranum.fr/identifiants-perennes-pid/fiche-synthetique/>).

Jeux de données : agrégations de données organisées pour former un ensemble cohérent. Les jeux de données numériques doivent être formatés de telle sorte qu'ils soient communicables, interprétables et adaptés à un traitement informatisé. Exemple : jeu de données de relevés hygrométriques ayant servi à faire un article scientifique. (Source : [COOPIST-CIRAD](#)).

Métadonnées : informations nécessaires à la description de données, en général structurées selon une norme. Le titre d'un ouvrage dans un catalogue de bibliothèque en est un exemple.

Moissonnage : processus selon lequel des métadonnées stockées (dans une base de données, dans un entrepôt) sont récupérées par une autre base de données qui les affiche à son tour.

2. Accompagner/aider : savoir répondre aux principales questions posées à l'accueil

Les données de la recherche occupent une place de plus en plus grande au sein des universités : production toujours plus grande de données qui nécessitent des infrastructures techniques adaptées, question de l'ouverture de ces données pour améliorer la reproductibilité de la recherche, accompagnement des doctorants et des chercheurs dans ces domaines.



Les personnels des bibliothèques ont de plus en plus de missions d'appui à la recherche, qui ne cessent de se développer depuis plusieurs années. Au cours d'une plage de service public, tout agent pourra, petit à petit, être confronté à des questions sur les données de la recherche. S'il est certain que des personnels sont spécifiquement formés dans ce domaine, il est nécessaire que chaque agent faisant du service public soit sensibilisé afin de pouvoir répondre aux questions de premier niveau dans le domaine.

2.1. Comment trouver des jeux de données

Des étudiants en Master 2ème année, des doctorants et des chercheurs peuvent être amenés à chercher des jeux de données pour leurs recherches.

a. L'essentiel

Deux outils peuvent être utilisés pour chercher des jeux de données : [Google Dataset Search](#) et [OpenAIRE](#). Dans les deux cas, la recherche se fait très facilement avec des mots-clefs, comme dans une recherche bibliographique classique.

b. Pour développer

Des outils plus complexes peuvent être utilisés pour trouver des jeux de données. Ils permettent notamment d'identifier des entrepôts de données spécialisés. A cet effet, on peut par exemple consulter :

- [Re3data](#)
- [FAIRsharing](#)

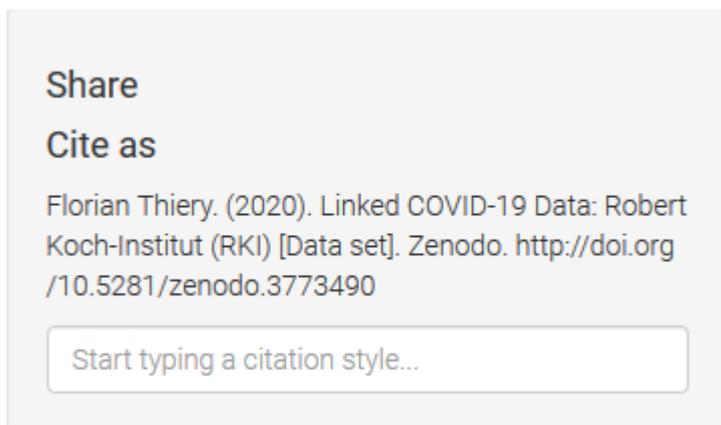
Dans les deux cas, ces annuaires d'entrepôts sont destinés à chercher des entrepôts de données, et non des jeux de données directement. La recherche dans ces deux bases doit obligatoirement se faire en anglais. C'est pourquoi il est nécessaire, avant de se lancer dans une recherche, de définir les mots-clefs qui vont permettre de délimiter les contours du sujet, comme dans toute recherche bibliographique.

Pour trouver les bonnes terminologies, on peut utiliser [Loterre](#). Par exemple, un chercheur souhaite trouver des données en dynamique moléculaire : en cherchant ce terme dans Loterre, on apprend qu'un mot-clef pertinent en anglais peut être "Molecular Dynamics Simulations".

Une fois en possession des mots-clefs pertinents en anglais, la recherche peut être lancée dans les annuaires d'entrepôts.

2.2. Comment les citer

a. L'essentiel



Share

Cite as

Florian Thiery. (2020). Linked COVID-19 Data: Robert Koch-Institut (RKI) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3773490>

Start typing a citation style...

De manière générale, il faut citer le titre, le ou les auteurs, la date de publication, le cas échéant la version, l'organisme de recherche, le DOI.

Les entrepôts de données proposent des modèles de citation pour citer correctement les données, sur le modèle de ce qui se fait dans l'édition scientifique. Il suffit alors de copier/coller la citation.

Zenodo, par exemple, propose même de reformater la citation en fonction d'un style dont il suffit de taper le nom dans le champ texte sous la citation.

Si l'on dispose du DOI, on peut également utiliser le [DOI Citation Formatter](#) : après avoir choisi le style de citation voulu, elle est automatiquement générée.

2.3. Comment faire un plan de gestion des données (PGD ou DMP)

a. L'essentiel

Le site dmp.opidor.fr mis en ligne par l'Inist permet de faire en ligne son DMP. Le chercheur peut choisir un modèle de DMP parmi tous les modèles proposés par les établissements de recherche, les financeurs. Il est également possible de télécharger le modèle en PDF ou en docx. Un des avantages de DMP OPIDoR est de permettre le travail collaboratif sur le DMP.



Voir un exemple de plan de gestion de données rédigé : celui du projet [Hospitam](#).

3. Dans mon établissement

Cette partie de la trame est à remplir par chaque établissement, en fonction de ses spécificités. Voici néanmoins des exemples.



3.1. La politique de l'établissement en matière de Science Ouverte

Quelques exemples de politiques dans des universités du Grand Est :

Université de Strasbourg

Les chercheurs sont invités à déposer leurs publications sur l'archive ouverte [UnivOAK](#). Un DMP doit être obligatoirement rédigé pour tous les projets financés par l'IDEX. Pour toute demande d'information concernant la Science Ouverte, les chercheurs peuvent s'adresser au pôle d'appui à la diffusion de la recherche du service des bibliothèques. Un projet d'entrepôt institutionnel de données est en cours.

Université de Lorraine

Les chercheurs doivent déposer leurs publications dans l'archive ouverte institutionnelle [HAL-UL](#). En ce qui concerne les données de la recherche, les chercheurs sont accompagnés par le comité opérationnel des données de la recherche (transversal, regroupant des représentants Directions de la Documentation, du Numérique, de la Recherche et Valorisation, ainsi que la MSH Lorraine, du pôle de recherche OTELo et de l'Inist) pour toutes les questions relatives à la gestion des données et à leur versement dans des entrepôts. Un projet d'entrepôt institutionnel de données est en cours.

Université de Reims

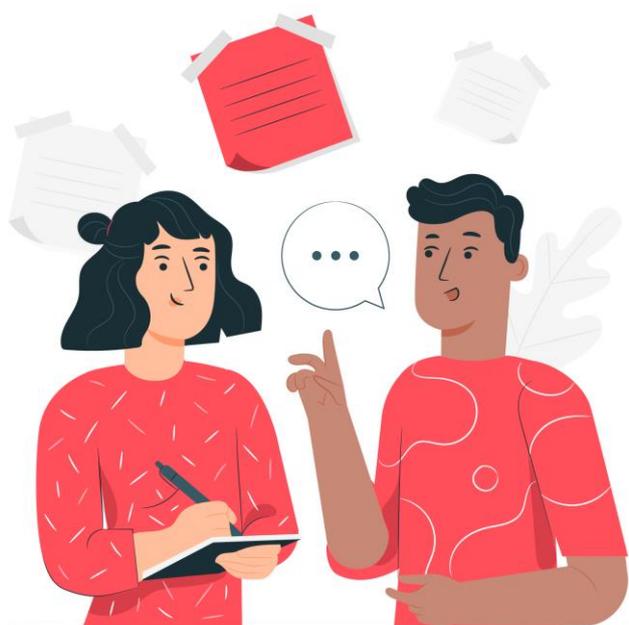
Les chercheurs sont incités à déposer leurs publications dans l'archive ouverte institutionnelle [HAL-URCA](#). Ils s'adressent pour le moment aux directeurs de pôles pour toutes les questions relatives à la gestion des données.

Les questions à se poser :

- L'établissement a-t-il une politique d'archive ouverte ?
 - Si oui, dispose-t-il d'une archive ouverte institutionnelle ?
 - Y a-t-il des mesures d'incitation au dépôt, voir un mandat de dépôt ?
- L'établissement a-t-il une politique de gestion des données de la recherche ?
 - Si oui, dispose-t-il d'un entrepôt institutionnel ?

3.2. Les personnes référentes, pour pouvoir réorienter l'utilisateur le cas échéant

La gestion des données implique plusieurs acteurs de l'établissement. Chaque établissement peut ainsi remplir cette partie en ajoutant les noms des personnes concernées.



- Pour les questions relatives à la **gestion des données de manière globale** (PGD, diffusion en accès ouvert) : les services d'appui à la recherche au sein des bibliothèques (voir la liste sur [SOS-PGD](#)) ;
- Pour les questions relatives au **stockage** : les services informatiques ;
- Pour les questions de **sécurité informatique** : le responsable sécurité des services informatiques (RSSI) ;
- Pour l'**archivage** : l'archiviste, les bibliothèques ;
- Pour les **aspects juridiques** : le service juridique ;
- Pour ce qui concerne les **données personnelles et le RGPD** : le délégué à la protection des données (DPO ou DPD).

4. Evaluer ses connaissances

1. Qu'est-ce que la Science Ouverte ?

- Une participation citoyenne, ouverte à tous, aux travaux de recherche
- **Une philosophie de partage des savoirs indépendante des éditeurs**
- Un mouvement limité à la France
- **Des principes partagés à l'international**
- La mise en ligne de travaux n'ayant que peu de valeur académique
- **La mise à disposition de publications académiques et de données scientifiques**
- Un phénomène passager
- **Un élément appelé à devenir essentiel dans la carrière des chercheurs**

2. Qu'est-ce qu'Horizon 2020 ?

- Le nom d'un voyage organisé vers Mars
- Le programme du Ministère de l'Enseignement supérieur autour des données de la recherche
- **Le programme européen pour la recherche et le développement pour la période 2014-2020**
- Un manifeste signé par les principales universités et grands organismes de recherche mondiaux en faveur de la Science Ouverte

3. A qui profite la Science Ouverte ? (Question d'appariement, à remettre dans l'ordre)

Le chercheur

Protection du droit d'auteur

Meilleure visibilité

La communauté scientifique

Améliorer la qualité du travail de recherche

Transparence et intégrité

Les agences de financement

Maîtrise du circuit de diffusion des résultats de la recherche

Accessibilité de la recherche au grand public

4. Les données de la recherche sont...

- **Des enregistrements factuels (chiffres, textes, images, sons...)**
- **Nécessaires pour valider le résultat d'une recherche**
- Aussi des objets matériels (souches bactériennes, souris...)
- Aussi des publications scientifiques (articles, livres...)

5. Quels sont les bénéfices de la conservation et de l'exposition des données de la recherche ?

- Augmenter le taux de citation d'une production scientifique
- Permettre la réutilisation de résultats de recherche par d'autres
- Pouvoir toujours publier dans des revues à haut facteur d'impact
- Valider le contenu d'une production scientifique
- Obtenir plus de financements

6. Pour trouver des données de recherche, je peux consulter... (plusieurs réponses possibles)

- Des annuaires d'entrepôts de données
- Des catalogues de bibliothèques
- Des annuaires de jeux de données
- Des entrepôts de jeux de données
- Des archives ouvertes

7. Copies d'écran : lequel est un jeu de données ?

The screenshot shows the Zenodo interface for a dataset. At the top, there is a search bar and navigation links for 'Upload' and 'Communities'. The dataset title is 'Coronavirus Twitter Data: A collection of COVID-19 tweets with automated annotations', dated March 31, 2020. It is categorized as a 'Dataset' and 'Open Access'. Statistics show 3,307 views and 7,767 downloads. The authors listed are Huang, Xiaolei; Jamison, Amelia; Broniatowski, David; Quinn, Sandra; and Dredze, Mark. A description states that the dataset contains tweets related to COVID-19 with original data, dates, keywords, and geolocation. A URL is provided: <http://twitterdata.covid19datasources.org/index>. The dataset is indexed in OpenAIRE. A table lists the files: 'covid19_2020_02_06.json.gz' (8.3 MB). The publication date is March 31, 2020, and the DOI is 10.5281/zenodo.3993873.

The screenshot shows the HAL archive interface. At the top, there are logos for CCSD, HAL, Episciences.org, Sciencesconf.org, and Support. The main header features the 'Archive ouverte de l'Université de Lorraine' logo and the 'UNIVERSITÉ DE LORRAINE' logo. The page title is 'tel-01751806, version 2' and 'Thèse'. The author is Sandrine Ollinger. The title of the thesis is 'Le raisonnement analogique en lexicographie, son informatisation et son application au Réseau Lexical du Français'. The abstract (Résumé) discusses the contemporary lexicography, its informatization, and its application to the French Lexical Network. The page also shows a list of files: 'theseOllingerS.pdf' and 'marquepage.pdf'. The HAL ID is 'tel-01751806, version 2'.

8. Pour pouvoir citer un jeu de données, j'ai besoin...

- D'obtenir l'autorisation des auteurs
- De payer un abonnement
- **D'un DOI (*Digital Object Identifier*)**
- Je ne peux pas citer un jeu de données tout seul, je dois citer tout l'entrepôt dans lequel il se trouve

9. A quoi sert un plan de gestion de données ?

- A obtenir des financements
- **A planifier toutes les opérations nécessaires pour produire des données de qualité durant le projet**
- **A respecter les demandes des financeurs**

10. Le PGD est un document collaboratif, quels acteurs peuvent participer à sa rédaction ?

- **Le porteur du projet**
- Le secrétaire administratif
- **Le professionnel de l'information et de la documentation**
- **L'informaticien**
- **Le délégué à la protection des données**
- L'éditeur
- Le financeur

11. Suivre le mouvement de la Science Ouverte permet aux chercheurs de :

- **Valoriser leur production scientifique**
- **Diffuser leurs résultats auprès d'une communauté locale, nationale et internationale**
- Mettre des brevets sur les données de recherches produites dans le cadre du financement public d'un projet
- **Accéder plus facilement aux publications scientifiques de la communauté**
- **Pouvoir réutiliser des travaux scientifiques issus de différents domaines**

Pour aller plus loin, vous pouvez consulter le guide d'autoformation aux données de la recherche produit par le gTIGRE : <https://zenodo.org/record/3920869>