

Gestion et partage des données de recherche

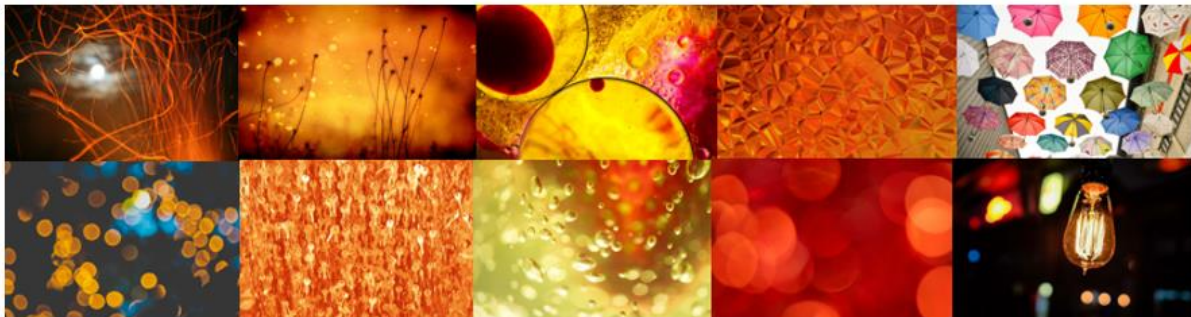
<https://www.fosteropenscience.eu/node/2328>

<https://www.fosteropenscience.eu/node/2644>



**Open Science
Training Courses**

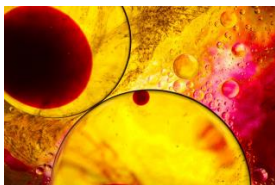
<https://www.fosteropenscience.eu/toolkit>



This project has received funding from the European Union's Horizon2020 programme for research, technological development and demonstration under agreement no 741839.

La taxonomie FOSTER définit la science ouverte comme le mouvement visant à rendre la recherche, les données et la diffusion scientifiques accessibles à tous les niveaux d'une société en quête d'informations.

Cela semble une bonne chose, mais que signifie la science ouverte (SO) dans un sens pratique ? Les dix cours de science ouverte de FOSTER répondent à certaines des questions les plus courantes que vous pourriez vous poser sur la mise en pratique de la science ouverte. Chaque cours dure environ 1 à 2 heures et vous recevrez un certificat à la fin. Les cours comprennent des conseils pratiques pour se lancer dans la SO ainsi que des informations sur les outils et les ressources spécifiques à la discipline que vous pouvez utiliser. Il n'y a pas d'ordre précis dans les cours - il suffit d'explorer les sujets que vous souhaitez approfondir à votre propre rythme.



Gestion et partage des données de recherche

Version imprimable : 10.5281/zenodo.2630562

Version en ligne : <https://www.fosteropenscience.eu/node/2328>

+ Evaluer le caractère FAIR des données

Version en ligne : <https://www.fosteropenscience.eu/node/2644>

Dans cette leçon on se concentre sur les données que l'on peut partager et sur la manière de le faire le plus efficacement possible.

La recherche basée sur les données est de plus en plus courante dans un large éventail de disciplines universitaires, de l'archéologie à la zoologie, et couvre des domaines aussi bien artistiques que scientifiques. Pour soutenir une bonne recherche, nous devons veiller à ce que les chercheurs aient accès à des données de qualité.

À l'issue de ce cours, vous serez capable de :

- comprendre les différences entre les données ouvertes, fermées et partagées
- être en mesure de prendre des décisions concernant les données que vous pouvez partager
- gérer vos données suivant les principes FAIR
- savoir ce qu'est un plan de gestion des données
- savoir comment obtenir un impact maximal du partage de vos données de recherche

Sommaire

1. Que sont les données de recherche ?
2. Données ouvertes, partagées et fermées
3. Rendre les données FAIR
4. Plan de Gestion des Données
5. Exemples disciplinaires de PGD publics
6. Outils de planification de la gestion des données
7. Attentes des financeurs concernant la planification de la gestion et du partage des données
8. Quand et comment partager vos données ?
9. Conseils pour obtenir le meilleur impact du partage de vos données de recherche
10. Testez vos connaissances
11. Ressources supplémentaires

1. Que sont les données de recherche ?

L'Université de Leeds décrit les données de recherche comme « toute information qui a été collectée, observée, générée ou créée pour valider les résultats originaux de la recherche ».

On peut également citer la définition issue de la Queensland University of Technology : " Le terme de données de la recherche désigne les données sous forme de faits, d'observations, d'images, de résultats de programmes informatiques, d'enregistrements, de mesures ou d'expériences sur lesquels un argument, une théorie, un test ou une hypothèse, ou un autre produit de la recherche est basé. Les données peuvent être numériques, descriptives, visuelles ou tactiles. Elles peuvent être brutes, nettoyées ou traitées, et peuvent être conservées dans tout format ou support. "

ARDC, « What Is Research Data », 2019. <https://ardc.edu.au/wp-content/uploads/2020/01/What-is-research-data.pdf>

Les données de recherche peuvent inclure des éléments comme...

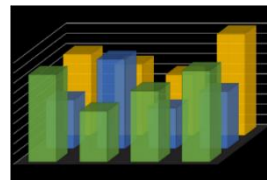
... des données brutes

Les données brutes sont celles qui sont capturées par des instruments et des capteurs tels que les télescopes, les téléphones intelligents et les satellites.



... des visualisations, modèles et algorithmes

Les chercheurs génèrent également des ressources numériques telles que des modèles et des algorithmes pour les aider à analyser, visualiser et présenter les données brutes de manière significative.



... des images, fichiers audio et vidéo

Ne pas oublier que les images numériques sont aussi des données. C'est également le cas de tous les fichiers audio ou vidéo capturés au cours de la recherche, tels que les interviews enregistrées.



... à peu près tout !

Essentiellement, les données de recherche peuvent être à peu près tout ce que les chercheurs produisent ou avec quoi ils travaillent au cours de leurs recherches.



Dans [cette vidéo](#) sont détaillés des exemples disciplinaires de différents types de données

2. Données ouvertes, partagées et fermées

2.1. Niveau d'ouverture

Données ouvertes

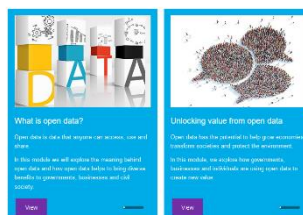
L'Open Data Institute (ODI) définit les données ouvertes comme celles auxquelles tout le monde peut accéder, que l'on peut utiliser et partager. Selon l'ODI, les données ouvertes doivent faire l'objet d'une licence afin que chacun sache comment il peut les utiliser et s'il peut les transformer, les combiner et les partager avec d'autres, même à des fins commerciales.

L'ODI propose une introduction sur tous les aspects des données ouvertes dans son parcours pédagogique « [Open Data Essentials](#) ». Nous vous recommandons vivement de revoir ces modules.

Open data essentials

Welcome to the Open Data Institute's e-Learning programme developed for the European Commission. This programme has been designed to enable you to discover what open data is and how it is changing the lives of citizens on our planet. There are 13 lessons for you to explore covering the essentials of open data, how to plan and measure success and how to implement an open data programme successfully. This programme is free and should take a maximum of 2-3 hours to complete.

Don't miss to discover each of the available lessons:



Données partagées

Comme pour les données ouvertes, les données partagées peuvent être rendues largement accessibles mais peuvent être soumises à certaines conditions telles que la réutilisation non commerciale ou la réutilisation avec attribution. Il est important de noter que toutes les données partagées ne doivent pas nécessairement être accessibles à tous. Parfois, les données partagées ne sont mises à la disposition que de groupes spécifiques tels que les pairs d'une autre université.

Données fermées

Si les chercheurs traitent des données très sensibles - telles que des données personnelles ou commerciales - il peut être impossible de les partager. Toutefois, même dans ce cas, une description des métadonnées associées aux données de recherche devrait être partagée. Le partage de données sensibles peut également être facilité par l'utilisation d'espaces sécurisés où seuls les utilisateurs autorisés bénéficient d'un accès contrôlé.

2.2. Modèle de données ouvertes 5 étoiles

Si vous décidez de rendre vos données ouvertes, le [modèle 5 étoiles Open Data](#) de Tim Berners-Lee donne un bon aperçu de ce que vous devez prendre en considération d'un point de vue pratique.

L'ouverture des données ne se limite pas à rendre ses données accessibles en ligne. Il s'agit également de fournir un accès intelligent afin que la communauté puisse comprendre et réutiliser efficacement vos données. Gardez à l'esprit que vous pourriez être l'un des futurs réutilisateurs de vos données et que l'effort que vous faites pour les contextualiser vous profite également ! Même si vous ne prévoyez pas d'appliquer une licence ouverte à vos données, la plupart des conseils du modèle 5 étoiles restent de bonnes pratiques.

Voici des exemples pour chaque étoile ajoutée au modèle :

	Données ouvertes 1 étoile : Mettre ses données à disposition sur le Web (quel que soit le format) sous une licence ouverte.
	Données ouvertes 2 étoiles : Les rendre disponibles sous forme de données structurées (par exemple, tableau Excel à la place d'un scan de tableau de données).
	Données ouvertes 3 étoiles : Utiliser des formats non propriétaires (par exemple, des valeurs séparées par des virgules en .csv au lieu d'Excel pour les tableaux).
	Données ouvertes 4 étoiles : Utiliser les Uniform Resource Identifiers (URI) pour désigner des éléments afin que les utilisateurs puissent pointer vers vos données.
	Données ouvertes 5 étoiles : Relier ses données à d'autres données pour fournir un contexte.

Les données ouvertes 5 étoiles ne sont pas le fruit du hasard !

L'élaboration d'un **plan de gestion des données** au début de votre prochain projet de recherche peut permettre de s'assurer que des mesures adéquates ont été mises en place pour fournir un accès intelligent.



Pour plus d'informations, voir la partie 4 sur le Plan de Gestion des Données un peu plus loin.

2.3. Qui décide si les données de recherche doivent être ouvertes, fermées ou partagées sous conditions ?

Les chercheurs ont un rôle clé à jouer pour décider quelles données peuvent être partagées, mais ils ne sont pas les seuls acteurs impliqués dans cette décision. Les décisions concernant l'endroit où sont stockées vos données à court et à long terme peuvent également avoir un impact sur les données que vous pouvez partager et avec qui.

Participants à la recherche

Si la recherche implique de travailler avec des sujets humains, il faut s'assurer d'obtenir un consentement éclairé. Celui-ci doit informer les participants à la recherche de tout projet de partage de leurs données (au sein de votre équipe de recherche ou à plus grande échelle) ainsi que de tout projet de conservation à plus long terme de leurs données en vue de leur réutilisation.



Le mieux est d'en parler à un comité d'éthique et de s'assurer que tout projet de réutilisation est intégré dans les formulaires de consentement que vous utiliserez. Si vous prévoyez d'utiliser un formulaire de consentement standard disponible auprès de votre institution, lisez-le d'abord, pour vous assurer qu'il ne contient pas de déclarations générales sur le refus de partager les données ou de promesses de détruire les données à la fin du projet.

Il faut prendre en considération toute procédure de nettoyage et/ou d'anonymisation des données qui devra être effectuée pour faciliter le partage dès le début. En gardant l'esprit que ces actions peuvent être très coûteuses, il faut s'assurer de demander un budget suffisant pour ces activités dans les demandes de subvention.

Ce [dossier](#) fait le point sur les données à caractère personnel en France.

La Commission Nationale de l'Informatique et des Libertés (CNIL) propose également de nombreuses ressources sur le [règlement général sur la protection des données \(RGPD\)](#) et le recueil du consentement des personnes.

Pour plus d'informations sur l'obtention du consentement éclairé, aller directement à la leçon sur la [protection des données et l'éthique](#).

Collaborateurs de recherche

Si vous participez à un projet de recherche en collaboration - que ce soit avec d'autres institutions universitaires, des partenaires industriels ou la science citoyenne - vous devez vous assurer que vos partenaires acceptent le partage des données. Cela doit être clarifié pendant la phase de conception du projet et intégré dans tout accord de consortium élaboré pour régir le projet. Le service chargé des subventions ou des contrats peuvent vous aider. Gardez à l'esprit que vos partenaires peuvent uniquement accepter de partager des jeux de données particuliers et demander que d'autres soient gardés confidentiels. Il est conseillé de préciser ces restrictions dans la proposition de recherche. Vous devez également



convenir de la date à laquelle, pendant la durée du projet, les jeux de données sélectionnés seront partagés et avec qui, et documenter ces décisions afin que tous les partenaires aient une idée claire de ce qui se passera et de la date à laquelle cela se produira.

Infrastructure de données de recherche

Lorsque vous planifiez votre recherche, réfléchissez à l'infrastructure de données de recherche qui sera utilisée pour rendre les données accessibles et aux personnes qui devront y avoir accès.

Il peut être tentant d'utiliser Google docs ou Dropbox pour partager facilement les données avec les collaborateurs pendant la phase active du projet. Cependant, n'oubliez pas d'examiner si vous allez travailler avec des données personnelles ou commerciales sensibles. Si c'est le cas, ce ne sont pas les meilleures options pour un partage de données sécurisé.

Il est toujours bon de passer un peu de temps avec une équipe des services informatiques pour savoir quelle infrastructure de partage de données peut être prise en charge en interne. Dans la plupart des cas, l'infrastructure fournie par votre institution assurera une meilleure sécurité et une meilleure sauvegarde. Il est fréquent que les collaborateurs d'autres institutions puissent y avoir accès.

N'oubliez pas de discuter de l'ampleur et du/des formats des données que vous prévoyez de partager et de réfléchir à la fréquence d'accès (quotidienne, mensuelle, annuelle). Ces aspects permettront de déterminer les meilleures options de stockage et de partage des données pendant la durée du projet et au-delà.

Par exemple, [Huma-Num propose un ensemble de services](#) à la disposition des chercheurs en Sciences Humaines et Sociales pour stocker et partager leurs données de manière sécurisée.



Entrepôts de données de recherche

Lorsque vous choisissez un entrepôt pour les données sélectionnées, en vue d'une conservation et d'un partage à long terme, assurez-vous que l'entrepôt répond à vos besoins. Par exemple, si vos données seront partagées avec une communauté de recherche spécifique, assurez-vous que l'entrepôt permet bien aux chercheurs de demander l'accès et d'être authentifiés.



La plupart des entrepôts de données ont des politiques qui précisent les limites relatives à la taille du dépôt de données ou les restrictions sur les formats qu'ils acceptent. Soyez conscient de tout processus de standardisation entrepris par l'entrepôt (c'est-à-dire lorsque les données déposées sont migrées vers des formats privilégiés). Dans de nombreux cas, la standardisation peut affecter la réutilisation des données.

Par exemple, si une feuille de calcul Excel présentée dans une publication est enregistrée au format PDF, elle sera disponible en tant qu'enregistrement de ce qui a été présenté dans l'article, mais perdra une grande partie des fonctionnalités nécessaires pour prendre en charge la validation et la réutilisation. Toute formule appliquée à des cellules individuelles de la feuille de calcul sera perdue.

Cela signifie également que la réutilisation des données sera plus difficile et nécessitera de les ressaisir dans un nouveau tableur.

Un exemple d'entrepôt proposé par l'infrastructure GBIF : <https://www.gbif.org/dataset/search>

Réutilisateurs des données secondaires

Si vous craignez que vos données soient utilisées à mauvais escient et que cela vous dissuade de les partager, vous pouvez toujours rendre les données accessibles tout en demandant aux réutilisateurs potentiels d'accepter des conditions générales d'utilisation équitable.

Par exemple, vous pouvez demander aux chercheurs qui souhaitent réutiliser vos données de cocher une case indiquant qu'ils ne tenteront pas d'identifier les participants anonymes.



La [licence d'utilisateur final \(EUL\)](#) du service de données britannique fournit un exemple de la manière dont cette approche est mise en pratique pour une archive de données nationale.

Cet [article](#) fait le point sur les règles de consultation des archives contenant des données personnelles en France.

3. Rendre les données FAIR

3.1. Contexte

Pour soutenir la recherche axée sur les données, nous devons veiller à ce que les chercheurs aient accès à des données de qualité ainsi qu'aux autres résultats connexes, notamment les logiciels, méthodes et publications associés.

Comme la recherche dans tous les domaines dépend de plus en plus de l'analyse informatique, il est également nécessaire de s'assurer que les résultats de la recherche puissent être trouvés et compris par les machines comme par les personnes.

À la lumière de ces éléments, les [principes de données FAIR](#) ont été introduits en 2016 pour aider à fournir un ensemble de principes directeurs à prendre en compte par toutes les parties prenantes de l'écosystème de la recherche afin de garantir que les résultats de la recherche soient Faciles à trouver, Accessibles, Interopérables et finalement Réutilisables.

Pour en savoir plus sur l'historique de l'élaboration des principes FAIR, consulter la section du [groupe de publication des données FAIR sur le site Force11](#).

3.2. Principes FAIR

La meilleure façon de rendre ses données FAIR est d'utiliser un entrepôt de données aligné sur les principes FAIR et conforme aux standards internationaux d'entrepôts de données (certification [CoreTrustSeal](#)). Si on fait cela, une grande partie du travail nécessaire pour rendre ses données FAIR est prise en charge par le service de dépôt.

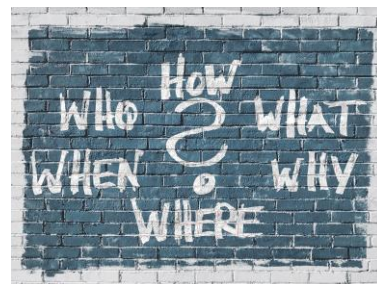
Pour en savoir plus sur ce que les entrepôts doivent respecter pour obtenir la certification CoreTrustSeal, consulter les [conseils donnés par OpenAIRE](#) (voir la section "What are these repository certifications based on?").

Ci-dessous, nous allons explorer chacun des principes FAIR et examiner comment l'utilisation d'un entrepôt aligné sur ces principes aide à respecter chacun d'eux.

Les informations contenues dans cette partie ont été synthétisées à partir des conseils produits par la [GO FAIR Initiative](#), le projet [Enabling FAIR Data Project](#) et [OpenAIRE](#).

3.2.1. Facile à trouver

La première étape consiste à s'assurer que vos données peuvent être retrouvées, tant par les humains que par les machines.



- La meilleure façon d'y parvenir est de déposer vos données de recherche dans un entrepôt respectant les principes FAIR, reconnu dans votre domaine. De cette façon, vos données pourront être consultées et découvertes en ligne. Les catalogues de services et annuaires d'entrepôts proposent des listes d'entrepôts thématiques (voir le sous-chapitre 8 "Quand et comment partager vos données"). Un nombre croissant d'universités et d'autres organismes de recherche mettent également en place des entrepôts de données de recherche spécialisés, pensez à vérifier ce qui pourrait être disponible localement.
- Fournir une description de vos données sous forme de métadonnées lisibles à la fois par les humains et les machines. Pour ce faire, utiliser un système standard généraliste ([Dublin Core](#) ou [DataCite](#)). Pour décrire de manière plus complète vos données, afin de fournir un contexte riche qui aide la communauté à comprendre vos données, vous pouvez utiliser un standard de métadonnées spécifique à la discipline ([DDI](#), [EML](#)...). Pour savoir ce qui est pertinent pour votre discipline, consultez le site web du [Digital Curation Centre](#), le [Metadata Standards Catalog](#) de la RDA ou les [standards de FAIRsharing](#).

Si vous déposez dans un entrepôt de données, il propose très probablement un standard de métadonnées reconnu. Il suffira de saisir les informations dans les champs prédéfinis fournis lors du dépôt de vos données. Comme les métadonnées et les données sont généralement des fichiers séparés, il est impératif que de mentionner l'identifiant pérenne dans la description des métadonnées.

- Un identifiant pérenne est un élément très important pour rendre les données FAIR. La plupart des entrepôts de données attribueront un identifiant unique et pérenne à vos données, tel qu'un Digital Object Identifier (DOI). Grâce à un DOI, les données sont plus faciles à trouver et à citer - et vous pouvez ainsi obtenir des crédits ! Pour des informations plus détaillées sur les identifiants, consulter les informations fournies par l'[initiative GO FAIR](#) ou la [thématique "Identifiants pérennes" de DoRANum](#).

3.2.2. Accessible

Une fois que quelqu'un - ou une machine ! - a trouvé vos données, il doit savoir comment y accéder.



- Si vous utilisez un entrepôt de données disciplinaire, institutionnel ou généraliste, il sera doté de protocoles standards pour faciliter la recherche en ligne.
- « Données FAIR » ne signifient pas nécessairement données ouvertes. Il peut y avoir de nombreuses bonnes raisons de ne pas partager les données, comme la protection des données sensibles personnelles ou commerciales. Dans ce cas, il s'agit d'indiquer clairement dans les métadonnées la restriction d'accès aux données connexes et d'expliquer comment la communauté scientifique peut en demander l'accès. Cela peut impliquer

de passer par un processus d'authentification automatisé ou nécessiter l'utilisation de zones de sécurité où seuls les utilisateurs autorisés ont un accès contrôlé.

- Il est recommandé d'attribuer une licence appropriée pour vos données si ce n'est pas déjà fait. Les licences indiquent clairement ce que les utilisateurs peuvent et ne peuvent pas faire avec les données. Pour plus d'informations, voir la leçon sur [les licences ouvertes](#).

Si l'accès aux données est limité, il est conseillé de s'adresser à votre équipe locale de soutien à la recherche qui vous accompagnera pour trouver un entrepôt de données adapté à vos besoins. Certaines universités mettent actuellement en place des coffres-forts électroniques locaux pour permettre un accès contrôlé aux données sensibles.

Toutes les données ne sont pas conservées à perpétuité. Certaines ne sont conservées que pendant une période déterminée, puis éliminées. Cependant, même si les données n'existent plus, il est bon de s'assurer que les métadonnées décrivant les données restent accessibles. De cette façon, les réutilisateurs à long terme de votre travail sauront exactement ce qu'il est advenu des données et ne se retrouveront pas dans une impasse où le statut des données est inconnu.

3.2.3. Interopérable

Pour rendre ses données interopérables, il faut s'assurer qu'elles peuvent être intégrées à d'autres données et qu'elles peuvent être utilisées par des applications ou des flux de travail pour l'analyse, le stockage, le traitement et l'utilisation à long terme.

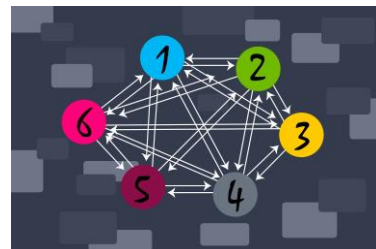


- Si possible, il vaut mieux éviter d'utiliser des logiciels propriétaires lors de la création des données. Opter pour des formats ouverts permettra aux réutilisateurs d'accéder plus facilement à vos données et de les réutiliser. Les entrepôts sérieux disposent généralement d'une liste de formats de fichiers privilégiés qu'ils acceptent. Ces formats ont été sélectionnés pour aider l'entrepôt à conserver les données utilisables à long terme.
- Pour faciliter la recherche automatique et l'interopérabilité des jeux de données, il est important d'utiliser des vocabulaires contrôlés couramment utilisés et de s'assurer que vos données sont bien structurées à l'aide de modèles de données bien définis. L'initiative GO FAIR fournit des [exemples](#) dans ce sens.

Dans [cette ressource de DoRANum](#) on peut accéder à une liste non exhaustive de formats fermés et leurs équivalents ouverts.

3.2.4. Réutilisable

Pour maximiser la réutilisation potentielle de vos données, assurez-vous que les données et leurs métadonnées associées sont bien décrites. Elles pourront ainsi être reproduites et/ou combinées dans différents contextes.



- Il est recommandé d'attribuer une licence ou une « renonciation aux droits » aux données qui sont entièrement ouvertes. Lorsque vous traitez des données fermées ou partagées, il est judicieux de lever toute ambiguïté sur ce que les utilisateurs peuvent et ne peuvent pas faire avec vos données en leur attribuant une licence. Les [Creative Commons](#) offrent une gamme de différents types de licences qui répondront aux besoins de la plupart des chercheurs.

Le site data.gouv.fr recense les licences de réutilisation autorisées dans le cadre de la loi pour une République numérique pour les "informations publiques (données, documents...)" et codes sources, ainsi que les licences spéciales homologuées.

Inclure la licence que sélectionnée pour vos données dans la description des métadonnées associées.

N'oubliez pas d'attribuer une licence à tout logiciel (ou code source) utile pour interpréter ou traiter vos données. Ceux-ci doivent également être produits et conservés de manière FAIR.

Des outils pratiques comme [License selector](#), [Choose an open source license](#) et [Licentia by inria](#) peuvent vous aider à attribuer, en fonction de vos besoins, la meilleure licence pour vos produits de recherche.

- Il est important de fournir une bonne documentation, telle qu'une description de la méthodologie de recherche utilisée pour créer vos données.

Il doit également y avoir une documentation suffisante pour le répertoire de fichiers, les éléments et les variables, afin de permettre à un tiers de les comprendre. Il peut s'agir de s'assurer que les feuilles de calcul Excel comportent des descriptions de lignes et de colonnes sensées ou qu'un document expliquant correctement les abréviations utilisées est joint à l'ensemble de données.

Tout logiciel ou outil nécessaire pour interpréter ou traiter les données - ceux créés ou utilisés - doit être inclus dans la documentation.

- Relier vos données à des résultats connexes fournira un cadre contextuel permettant aux réutilisateurs de comprendre vos données et la manière dont elles ont été obtenues. Il est souvent impossible de donner un sens à des données qui se trouvent isolées.

N'oubliez pas de relier tous les résultats à votre [identifiant ORCID](#) !

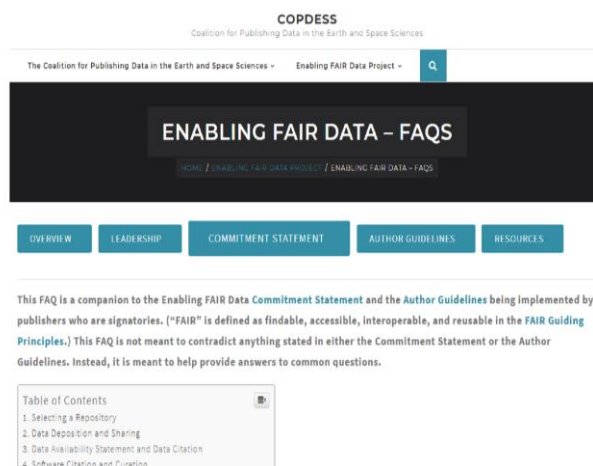
- Inclure une citation préférentielle pour vos données permet aux réutilisateurs de les citer plus facilement. Cela est également utile pour vous-même et votre organisation pour suivre l'impact de votre recherche.

Tenez-vous en aux standards de votre discipline, s'ils existent. S'il n'existe pas de standards, fournissez une méthode recommandée pour citer vos données. La citation doit inclure le ou les auteurs, le titre, l'éditeur, la date et l'identifiant.

Ce [guide](#) du Digital Curation Centre fournit quelques conseils utiles sur la citation des données.

3.2.5. Mettre en pratique les principes FAIR

Les [FAQs développées par le Enabling FAIR Data Project](#) fournissent des conseils pour démarrer de manière très pratique. Elles ont été élaborées pour les chercheurs en sciences de la terre et de l'espace, mais la plupart des conseils partagés sont applicables à toutes les disciplines.



<http://www.copdess.org/enabling-fair-data-project/enabling-fair-data-faqs/>

[Cette ressource](#) permet d'identifier les moyens à mobiliser pour faciliter la découverte, l'accès, l'interopérabilité et la réutilisation des données. Elle explique comment appliquer les principes FAIR.

3.3. Outils d'auto-évaluation des données FAIR

Dans cette partie, nous vous présentons quelques outils gratuits qui peuvent vous aider à évaluer le caractère FAIR des données. Ces outils peuvent être utilisés aussi bien par les chercheurs que par le personnel de soutien à la recherche.

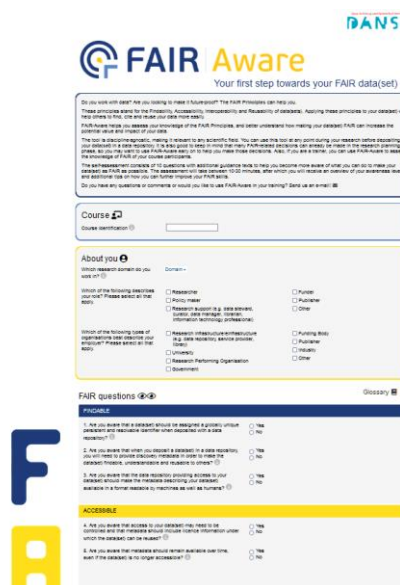
3.3.1. FAIR-Aware

[FAIR-Aware](#) est un outil en ligne visant à aider chercheurs et gestionnaires de données à évaluer leur connaissance des bonnes pratiques pour rendre leurs jeux de données FAIR, avant de les déposer dans un entrepôt de données. FAIR-Aware a été développé par les partenaires du projet FAIRsFAIR DANS, DCC et UniHB.

La [version française](#) a été rendue aussi simple et efficace que possible. Afin de l'améliorer, certaines questions s'adressent directement aux utilisateurs.

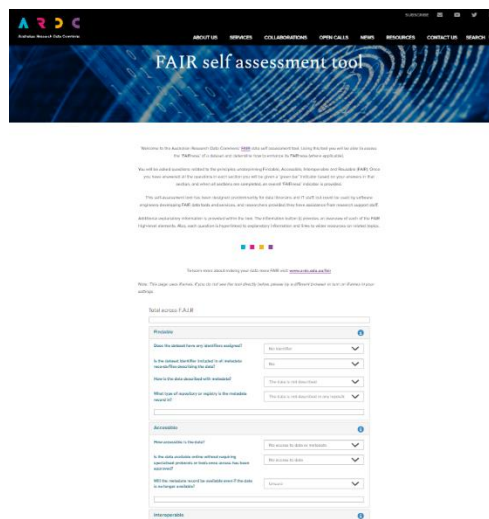
Comment fonctionne l'outil

Il comporte 10 questions enrichies d'informations supplémentaires et de conseils pratiques. Les utilisateurs qui parcourent le questionnaire en ayant en tête un exemple de jeu de données peuvent ainsi mieux comprendre les principes FAIR.



3.3.2. Outil d'auto-évaluation des données FAIR de l'ARDC

L'[outil d'auto-évaluation des données FAIR](#) de l'Australian Research Data Commons (ARDC) aide à évaluer le caractère « FAIR » d'un jeu de données et, si nécessaire, aide à voir comment on peut améliorer son caractère FAIR. Cet outil d'auto-évaluation a été conçu principalement pour être utilisé par les data librarians et le personnel informatique. Il peut également être utilisé par les ingénieurs en informatique qui développent des outils et des services FAIR. Les chercheurs souhaitant l'utiliser peuvent demander l'aide du personnel de soutien à la recherche. L'ARDC propose une large [gamme de ressources](#) pour aider à rendre ses données FAIR.



Comment fonctionne l'outil

Il s'agit d'un outil d'auto-évaluation qui aide à vérifier le caractère FAIR des données en guidant à travers une série de questions relatives à chacun des quatre groupes de principes (Facile à trouver, Accessible, Interopérable et Réutilisable). Il suffit de sélectionner une réponse dans la liste déroulante fournie pour chacune des questions posées. L'outil dispose d'une aide intégrée pour s'assurer que les utilisateurs comprennent ce que signifie chacun des quatre principes FAIR.

Une fois que vous avez répondu à toutes les questions de chaque section, une « barre verte » s'affiche pour indiquer si vos données sont FAIR pour le principe en question. Une fois toutes les sections complétées, on obtient une évaluation globale du caractère FAIR de ses données.

3.3.3. Checklist pour des données FAIR d'EUDAT

La checklist d'[EUDAT](#), facile à utiliser, permet de connaître les mesures qui pourraient être prises pour améliorer le caractère FAIR de ses données.

Comment fonctionne l'outil

La checklist EUDAT n'est pas un outil en soi, mais plutôt une fiche de référence pratique qu'on peut imprimer pour vérifier rapidement le caractère FAIR de ses données. Il suffit de parcourir chaque section et de cocher toutes les cases correspondantes. S'il y a des déclarations qui ne sont pas cochées, réfléchir à la manière dont on peut les remplir. Il est possible de demander de l'aide à une équipe de soutien ou à sa bibliothèque locale.

La [Checklist EUDAT](#) peut être téléchargée et imprimée dans différents formats à partir de Zenodo.

How FAIR are your data?

Findable
It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Accessible
It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

Interoperable
Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Reusable
Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

Findable Accessible Interoperable Reusable

How FAIR are your data? checklist, CC-BY by Sarah Jones & Margen Groothuis, EUDAT, image CC-BY-SA by bangafundit

4. Plan de Gestion des Données

Décider quelles données doivent être ouvertes, fermées ou partagées nécessite une planification avancée. Dans cette section, vous apprendrez ce qu'est un plan de gestion des données et comment il peut vous aider à prendre des décisions importantes avant le début de votre recherche.

4.1. Qu'est-ce qu'un Plan de Gestion des Données (PGD) ?

Un plan de gestion des données (PGD) ou Data Management Plan (DMP) est un document qui décrit le volume et le(s) format(s) des données générées, collectées ou réutilisées pendant la durée d'un projet et qui décrit comment elles seront traitées et partagées pendant le projet et à plus long terme.

De nombreux organismes de financement exigent qu'un PGD soit soumis dans le cadre des nouvelles demandes de subvention. Cependant, même si votre recherche n'est pas soutenue par des fonds externes, l'élaboration d'un PGD est un exercice utile lorsque vous travaillez avec des données.

En outre, le PGD peut servir de plaque tournante entre vos données et d'autres résultats et contribuer à une communication efficace entre vous et le personnel de soutien de votre institution en ce qui concerne vos exigences en matière de traitement, de stockage et de partage des données.

Dans la plupart des cas, il est beaucoup plus difficile et coûteux de rendre les données FAIR rétrospectivement que si vous aviez prévu des actions spécifiques dès le début de votre projet de recherche. De nombreux organismes de financement ont fourni des modèles de PGD avec des conseils intégrés pour vous aider à rédiger un PGD en gardant à l'esprit les principes FAIR.

4.2. Que doit couvrir votre PGD ?

Cette [courte vidéo](#) synthétise ce qu'est un PGD et les principales questions posées.

Les modèles de PGD de la plupart des organismes de financement couvrent cinq thèmes principaux, détaillés ci-dessous :

Types de données, formats, standards et méthodes de saisie

Exposer et justifier ses choix.

- Détailler les données qui seront créées et expliquer les choix de formats, standards et méthodologies spécifiques. Ces choix peuvent rendre plus facile ou plus difficile le partage et la conservation des données.
L'utilisation de formats standards ou largement adoptés rendra les données interoperables. Les formats ouverts ou non propriétaires sont préférables, car il sera plus facile de les traiter par la suite. Si les données doivent être déposées dans une archive, des formats particuliers peuvent être demandés.
- La documentation et les métadonnées permettent aux données d'être comprises et découvertes par d'autres. Il est fondamental de saisir les détails contextuels sur la façon dont les données ont été créées et sur les raisons pour lesquelles elles l'ont été. Les métadonnées sont un sous-ensemble de cette vaste documentation, qui décrit les données en détail.
Il existe plusieurs standards de métadonnées qui facilitent la description des données, de manière cohérente. Les documentalistes, les entrepôts de données ou vos collègues peuvent vous conseiller sur les standards pertinents.

Ethique et propriété intellectuelle

Présenter des arguments solides en faveur de toute restriction au partage.

- Expliquer les contraintes éventuelles, telles que les périodes d'embargo ou les restrictions d'accès, et s'assurer qu'elles sont correctement justifiées. On s'attend généralement à ce que les données de recherche financées par des fonds publics soient ouvertement disponibles dès que possible.
- Toute recherche impliquant des données ou du matériel d'origine humaine est soumise à un examen éthique formel. Dans ce cas, il faut décrire les mesures prises pour protéger les participants à la recherche, par exemple en rendant les données anonymes.
- Il est possible de négocier le consentement éclairé pour le partage des données. De nombreux comités d'éthique universitaires fournissent des exemples de formulaires de consentement. Il faut se référer à une législation pertinente, comme le règlement général sur la protection des données (RGPD).
Voir pour cela le [site de la CNIL](#).
- La propriété des données doit être clarifiée et, si nécessaire, des accords de consortium doivent être rédigés pour négocier les licences au début du processus de recherche. Si on accepte ou qu'on achète des licences pour réutiliser les données de tiers, il faut être conscient des restrictions que cela impose pour le dépôt et le partage ultérieur des données. Un soutien institutionnel est généralement disponible auprès d'experts de services juridiques ou de services partenariats et valorisation.

Accès, partage des données et réutilisation

Anticiper et planifier la réutilisation des données.

- Lors de la prise de décision du partage des données, il faut envisager à quels utilisateurs vos données seraient utiles et comment répondre à leurs besoins.
- Les centres de données peuvent demander de respecter des normes de qualité minimales afin de s'assurer que vos données peuvent être comprises et réutilisées par d'autres chercheurs.
- Pour rassurer les financeurs, il est nécessaire de fournir des détails spécifiques sur l'accès : lieu, date et manière dont vos données seront mises à disposition.
Les financeurs indiquent souvent des délais pour la publication, par exemple la mise à disposition des données dès leur publication.

Si vous ne pouvez pas répondre à ces attentes ou si vous devez imposer des restrictions, démontrez que vous avez envisagé différents moyens de surmonter ces difficultés.

- Si possible, choisir une base de données disciplinaire, un centre de données ou un entrepôt institutionnel approprié.

Pour vous aider, consultez la liste des entrepôts référencés par [Cat OPIDoR](#) ou [re3data](#). Si l'accès à vos données doit être restreint, recherchez des services (ou des espaces) de données sécurisés.

Vérifiez également si le financeur demande d'utiliser un entrepôt certifié (par exemple, les entrepôts certifiés [CoreTrustSeal](#)).

Stockage à court terme et gestion des données

Définir le soutien à la gestion des données dont on peut disposer.

- Décrire les services disponibles au sein de votre institution et les compétences ou ressources supplémentaires que vous devez obtenir. Si vous avez besoin d'un soutien externe, justifiez les choix effectués et le budget demandé.
- Indiquer clairement qui sera responsable des différentes tâches.
- Aurez-vous besoin d'une infrastructure permettant un accès à distance sécurisé ? Comment la qualité des données sera-t-elle contrôlée si vous travaillez dans un réseau réparti sur plusieurs sites ?
- De solides conventions de nommage des fichiers et des applications de gestion des versions peuvent être utiles pour suivre le processus de développement, en particulier lorsque plusieurs personnes travaillent ensemble.
- Appliquer des niveaux appropriés de gestion des données. Les financeurs veulent être rassurés sur le fait que la gestion quotidienne des données est adaptée à leur objectif. Vous pouvez appliquer différents niveaux de service ou adopter une combinaison d'approches :
 - La sécurité peut être plus robuste pour toutes les données sensibles que vous collectez que pour les données secondaires que vous détenez sous licence. Réfléchissez à la manière dont vous allez transférer les données en toute sécurité, par exemple en les cryptant ou en utilisant un stockage en ligne sécurisé. Si vous utilisez des services en ligne, vous devez savoir où vos données sont hébergées et être certain que cela est légalement autorisé.
 - La sauvegarde des données uniques est plus importante que les copies de données secondaires. Plus les données sont importantes et plus elles sont utilisées, plus elles doivent être sauvegardées régulièrement. Les services de stockage avec sauvegarde automatisée, tels que ceux offerts par les services informatiques des universités, sont très robustes. Ces services peuvent être utilisés en combinaison avec le stockage portable ou l'informatique dans le cloud pour répondre à des besoins particuliers.

Dépôt et conservation à long terme

Sélectionner les données à conserver au-delà de la durée de vie du projet.

- Le partage et la conservation des données peuvent ne pas être applicables dans tous les cas. Le Digital Curation Center (DCC) propose un guide d'évaluation intitulé « [How to...](#) » qui présente des stratégies pratiques pour aider à sélectionner les données importantes.
- Décider de ce qui a une valeur à long terme et préparer ces données selon les normes prévues pour le dépôt sont des processus qui prennent du temps et pour lesquels il faut allouer des ressources importantes.

Cette [checklist](#) aide à préparer vos données et à vérifier que vous n'avez rien oublié avant de déposer vos données.
- Ne pas oublier de sauvegarder les données associées à un graphique (dans une publication). Les financeurs s'attendent généralement à ce que les résultats publiés comprennent des informations sur la manière d'accéder aux données justificatives. Les données qui sous-tendent les publications doivent être extraites, saisies sous une forme lisible par une machine et déposées quelque part pour qu'elles restent accessibles.

- S'assurer que vos données resteront accessibles. Les universités fournissent de plus en plus d'infrastructures d'accompagnement sur la gestion des données et certains services disciplinaires peuvent être utiles.
- En ce qui concerne l'archivage à long terme, le [CINES](#) (Centre Informatique National de l'Enseignement Supérieur) a été mandaté par le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation pour mettre en place une solution d'archivage électronique du patrimoine scientifique pour toute la communauté ESR (Enseignement Supérieur et Recherche) française.

Pour remplir cette mission, le CINES propose plusieurs solutions d'archivage, dont la plateforme PAC (Plateforme d'Archivage du CINES).

Ces solutions présentent plusieurs caractéristiques : elles sont mutualisées, personnalisées, sécurisées et certifiées.

En pratique, le projet d'archivage doit être analysé et préparé en amont avec les personnes ressources de l'établissement ou du laboratoire (archiviste, documentalistes, services supports etc.). L'équipe de recherche doit contacter le CINES afin de présenter le projet, de décrire le type de données, les formats, la volumétrie etc. Le CINES procède ensuite à l'évaluation de la faisabilité du projet. Le format des données doit être publié, libre et recevable par le CINES.

L'outil [FACILE](#) offre la possibilité de tester la validité du format des fichiers conformément à la liste des formats archivables au CINES. Des tutoriels d'aide à la correction de PDF y sont également proposés.

4.3. Rédiger un PGD permet d'éviter les cauchemars de la gestion des données !

Cette [courte vidéo](#) (créée par Karen Hanson, Kevin Read et Alisa Surkis, de la bibliothèque des sciences de la santé de l'Université de New York) fournit quelques conseils sur les éléments à inclure dans un PGD.

5. Exemples disciplinaires de PGD publics

Vous trouverez ci-dessous quelques liens vers des plans de gestion des données (PGD) publics, qui ont été produits par des projets dans les domaines des arts et des sciences humaines et des sciences de la vie.

Pour d'autres exemples de PGD, consulter

- la rubrique « [DMPs publics](#) » de DMP OPIDoR,
- le site web du [Digital Curation Centre \(DCC\)](#),
- l'outil [DMPonline du DCC](#),
- la collection de [plus de 800 DMPs publics d'OpenAIRE](#),
- La fiche pratique « [Trouver des plans de gestion de données \(PGD\) pour s'en inspirer, en 4 points](#) » du Cirad.

Exemple de PGD dans le domaine des arts et des sciences humaines

Le projet [FREME](#), soutenu par H2020, est un cadre ouvert de services électroniques pour un projet d'enrichissement multilingue et sémantique du contenu numérique. Le projet a partagé publiquement [son PGD](#).

Exemple de PGD dans le domaine des sciences de la vie

Ce [PGD public](#) a été créé par Laurent Gatto, de l'Université de Cambridge, pour le projet « Understanding protein multi- and trans-localisation at the full proteome level ». Le PGD a été publié dans la revue [Research Ideas and Outcomes \(RIO\)](#) qui publie des résultats sur l'ensemble du cycle de recherche, y compris des propositions de projet, des données, des méthodes, des flux de travail, des logiciels, des rapports de projet et des articles de recherche. La publication de votre PGD est

un excellent moyen de rendre votre activité de gestion des données transparente et cela génère une autre production digne d'être citée.

Exemple de PGD dans le domaine de l'agronomie

Ce [PGD public](#) sur les gènes du raisin pour la rareté de l'eau a été rédigé par Cédric Goby et Laurent Torregrosa qui ont utilisé le modèle proposé par l'INRAE et bénéficié d'un accompagnement de cet organisme. Les données sont décrites très précisément et très bien documentées. Elles sont stockées de manière sécurisée sur un serveur dédié, avec un accès examiné suivant les règles établies par l'accord de consortium. Elles seront partagées le plus largement possible via [l'entrepôt de l'INRAE](#), même si l'un des partenaires a un statut privé.

Exemple de PGD en sciences médicales

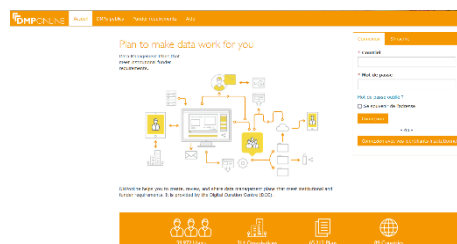
Ce [PGD](#) a été rédigé en français et partagé par Joseph Larmarange et Nicolas Rouveau de l'IRD, pour son projet [ATLAS](#) visant à décrire, analyser et comprendre les effets de l'introduction de l'autodépistage du VIH en Afrique de l'Ouest à travers l'exemple du programme ATLAS en Côte d'Ivoire, au Mali et au Sénégal. Ce PGD décrit très précisément les données d'enquêtes qualitatives et quantitatives. Le protocole de recherche a été évalué et autorisé par 4 comités d'éthique. Des formulaires de consentement ont été recueillis systématiquement et les données personnelles anonymisées.

6. Outils de planification de la gestion des données

6.1. DMPonline

[DMPonline](#)

C'est un outil gratuit qui aide les équipes de recherche à rédiger des plans de gestion des données qui répondent aux exigences des organismes de financement. DMPonline a été développé conjointement par le Digital Curation Centre (DCC) et l'University of California Curation Centre (UC3). L'outil contient un certain nombre de modèles qui représentent les exigences des différents organismes de financement à travers l'Europe. Trois questions sont posées aux utilisateurs dès le départ pour déterminer le modèle approprié à afficher. L'utilisation d'outils tels que DMPonline vous permet d'éliminer les aléas dans la rédaction de votre plan de gestion des données en vous fournissant l'ensemble des questions spécifiques auxquelles les différents organismes de financement souhaitent que vous répondiez. L'outil fournit également aux utilisateurs des conseils généraux - et, le cas échéant, des conseils institutionnels - pour s'assurer que vos réponses sont réalistes et applicables.



6.2. DMP OPIDoR

[DMP OPIDoR](#)

C'est un outil gratuit d'aide à la rédaction d'un DMP développé par l'Inist-CNRS sur la base du code source DMP Roadmap (fusion du code DMP Online et DMPTool, outils britannique et américain). Il est accessible à l'ensemble de la communauté scientifique de l'Enseignement Supérieur et de la Recherche et à ses partenaires français ou étrangers. Il permet les échanges collaboratifs entre les partenaires d'un même projet et les services d'accompagnement. Il est sécurisé sur un serveur hébergé à l'Inist-



CNRS, en France. C'est un outil qui évolue grâce aux retours des utilisateurs (chercheurs et personnels d'accompagnement).

De nouveaux développements permettent à l'outil de se connecter avec d'autres systèmes d'informations pour récupérer les informations, éviter des doubles saisies, proposer des référentiels...

Pour plus d'informations sur les plans de gestion des données et des conseils sur leur rédaction, consulter la [thématique « Plan de gestion de données » de DoRANum](#).

7. Attentes des financeurs concernant la planification de la gestion et du partage des données

En France, en suivant les mesures du Plan national pour la science ouverte, la diffusion ouverte des données de recherche est obligatoire pour les projets financés sur fonds publics.

La plupart des financeurs s'attendent à ce que toutes les données nécessaires à la validation des résultats publiés soient rendues accessibles, dans la mesure du possible, et que les données sensibles soient protégées de manière adaptée.

En ce qui concerne les plans de gestion des données (PGD), les financeurs veulent avoir la preuve, au stade de la demande de subvention, que le traitement et le partage des données ont été suffisamment pris en compte avant le début de toute recherche. De nombreux organismes de financement exigent désormais cette preuve en soumettant un plan de gestion des données durant le projet et à la fin. Il est important de souligner que les financeurs ne s'attendent pas à quelque chose de gravé dans le marbre. Ils s'attendent à ce que votre PGD soit mis à jour au fur et à mesure de l'avancement du projet. Vous devez être en mesure de prouver que vous avez réfléchi aux données que vous allez générer et à la manière dont elles seront gérées et partagées de manière responsable.

Vous pouvez trouver une sélection de modèles de PGD des organismes de financement ou d'établissements de recherche dans [DMP OPIDoR](#).

7.1. Plans de gestion des données dans le cadre d'Horizon 2020 et d'Horizon Europe

Dans le cadre d'Horizon 2020 (H2020), la Commission européenne a introduit le projet pilote pour des données de recherche ouvertes (ORD) qui vise à rendre les données de recherche générées par les projets Horizon 2020 accessibles avec le moins de restrictions possibles tout en garantissant la protection des données sensibles. Le projet OpenAire donne un excellent aperçu du [projet pilote ORD](#).

Dans la version révisée du programme de travail 2017, le projet pilote pour des données de recherche ouvertes a été étendu pour couvrir tous les domaines thématiques. Les projets H2020 sont tenus de fournir un plan de gestion des données dans les six premiers mois de leur projet.

Afin de rendre les données de recherche [Faciles à trouver, Accessibles, Interopérables et Réutilisables \(FAIR\)](#), un PGD doit inclure des informations sur :

- le traitement des données de recherche pendant et après la fin du projet
- quelles données seront collectées, traitées et/ou générées
- la méthodologie et les standards qui seront appliqués
- si les données seront partagées/en accès libre
- comment les données seront conservées et préservées (pendant et après la durée de vie du projet).

Pour plus d'informations, veuillez consulter le [guide de la CE sur la gestion FAIR des données à l'horizon 2020, version 3.2 du 21 mars 2017](#).

Le projet OpenAire fournit également des [conseils étape par étape sur la manière de rédiger un plan de gestion des données](#).

Dans le cadre d'Horizon Europe, il y a deux nouveautés :

- L'obligation de partage des données de recherche dans des entrepôts de référence
- L'accès ouvert au plan de gestion de données (PGD), les coûts de gestion des données étant éligibles au remboursement.

Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. Présentation du programme Horizon Europe. Septembre 2019.

<https://www.horizon2020.gouv.fr/cid145004/presentation-du-programme-horizon-europe-09-2019.html>

7.2. Plans de gestion des données dans le cadre des projets financés par l'Agence Nationale de la Recherche (ANR)

[L'ANR demande l'élaboration d'un PGD](#) pour les projets financés à partir de 2019. Ce document vise à décrire la façon dont les données de recherche seront produites, collectées, réutilisées, stockées, protégées, disséminées ou encore conservées à long terme. L'objectif est une meilleure anticipation et un suivi rigoureux tout au long des travaux, pour préparer le partage, la réutilisation et la pérennisation des données dans le respect du principe « aussi ouvert que possible, aussi fermé que nécessaire ». L'ANR suit les recommandations du Comité pour la science ouverte qu'elle a sollicité à ce sujet, en privilégiant le modèle de PGD proposé par Science Europe. L'Agence a travaillé en concertation avec l'Inist-CNRS pour que [son modèle de PGD](#) soit intégré dans DMP OPIDoR.

8. Quand et comment partager vos données ?

8.1. Quand partager vos données ?

Idéalement, il faudrait partager ses données de recherche dès que possible. Cela peut se faire dès la publication des résultats de la recherche, mais aussi beaucoup plus tôt dans le cycle de vie de la recherche.

De nombreux chercheurs sont réticents à partager leurs données, car ils craignent qu'elles ne soient pas parfaites. Bien qu'il soit préférable de produire des données de bonne qualité, la réalité est que la plupart des données présentent de légères bizarreries et anomalies. Vous n'avez pas besoin d'attendre que vos données soient parfaites pour les partager. Elles peuvent toujours être très utiles à d'autres. En outre, le fait de partager vos données à un stade précoce peut aider à mettre en évidence les erreurs ou les incohérences qu'il convient de corriger avant de publier vos résultats.



8.2. Comment partager vos données ?

Vous pouvez partager vos données de plusieurs manières. Voici quelques options à envisager.

Pour plus d'informations sur les entrepôts, consulter la thématique ["Dépôt et entrepôts" de DoRANum](#).

8.2.1. Entrepôts disciplinaires

Il est toujours préférable de déposer ses données de recherche dans un entrepôt disciplinaire reconnu dans sa communauté. La plupart d'entre eux offriront une certaine valeur ajoutée en termes de contrôle de la qualité ou de contexte supplémentaire pour soutenir une réutilisation future.

De nombreuses revues et financeurs précisent les entrepôts de données qu'ils souhaitent voir utiliser par les chercheurs. Il est donc conseillé de vérifier les conditions de votre subvention ou de votre accord de publication pour plus de détails.

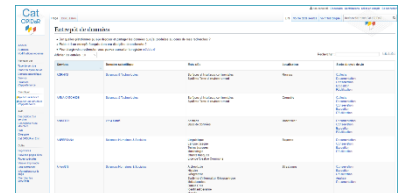
S'il n'y a pas de recommandation, il est possible de choisir un entrepôt dans un annuaire, en fonction de ses besoins.

Exemples de catalogues de services ou d'annuaires d'entrepôts

[Cat OPIDoR](#)

Catalogue pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche

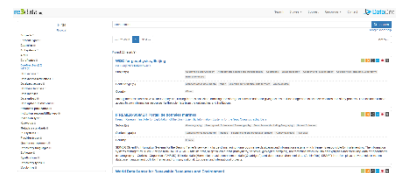
Il recense et décrit les services français dédiés aux données scientifiques. Par domaine scientifique, Cat OPIDoR présente des sites d'information, de formation, des outils de gestion, des plateformes pour accompagner les chercheurs sur l'ensemble des étapes clés de la gestion, collecte, stockage, conservation et ouverture des données. Il permet notamment de rechercher des entrepôts de données français.



Domaine	Nom de l'entrepôt	Description	URL
Biologie	Biobase	Base de données de séquences génétiques	http://biobase.org
Chimie	Chemical Abstracts	Base de données de réactions chimiques	http://chemabst.org
Physique	Physica	Base de données de résultats expérimentaux	http://physica.org
Mathématiques	Mathematics	Base de données de publications	http://maths.org

[re3data](#)

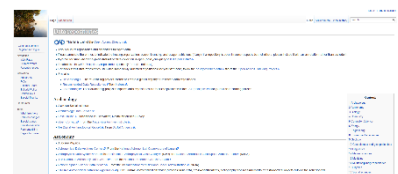
Il donne accès à des entrepôts de données sur un très large éventail de sujets. En 2021, plus de 900 entrepôts sont répertoriés pour les sciences humaines et sociales et près de 1500 pour les sciences de la vie - chacun d'entre eux permettant aux utilisateurs d'effectuer des recherches plus spécifiques par sous-domaine.



Repository Name	URL
re3data	http://re3data.org
re3data	http://re3data.org
re3data	http://re3data.org

[Open Access Directory - Data repositories](#)

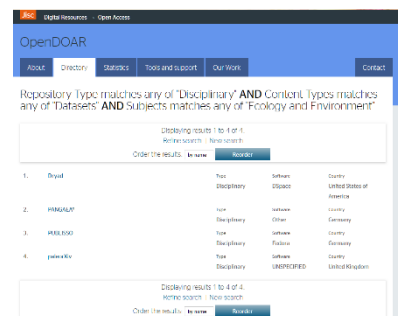
Il s'agit d'une liste d'entrepôts et de bases de données pour les données ouvertes.



Repository Name	URL
Open Access Directory	http://openaccess.org
Open Access Directory	http://openaccess.org
Open Access Directory	http://openaccess.org

[OpenDOAR](#)

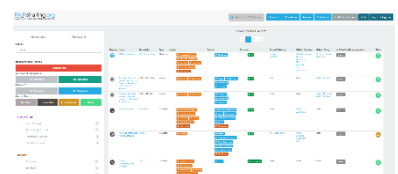
C'est un répertoire international d'entrepôts en libre accès. Chaque enregistrement d'entrepôt dans OpenDOAR a été soigneusement examiné et traité par un membre de l'équipe éditoriale, ce qui permet d'offrir un service de confiance à la communauté.



Repository Name	URL
OpenDOAR	http://opendoar.org
OpenDOAR	http://opendoar.org
OpenDOAR	http://opendoar.org

[FAIRsharing](#)

Les chercheurs peuvent utiliser FAIRsharing comme une ressource de consultation pour identifier et citer les standards, les bases de données ou les entrepôts qui existent pour leurs données et leur discipline.



Repository Name	URL
FAIRsharing	http://fairsharing.org
FAIRsharing	http://fairsharing.org
FAIRsharing	http://fairsharing.org

Cette [checklist](#) fournit des critères utiles pour trouver l'entrepôt de données approprié pour votre recherche.

8.2.2. Entrepôts de données institutionnels

Un nombre croissant d'universités et d'autres organismes de recherche mettent en place des entrepôts de données de recherche spécialisés (exemple : [Data INRAE](#)). Même si vous avez déposé vos données dans un entrepôt disciplinaire, vous pouvez également ajouter une description de métadonnées comprenant un lien vers les données dans votre entrepôt de données institutionnel. C'est un autre moyen de rendre vos données plus visibles et de soutenir l'évaluation de la recherche institutionnelle. Si vous n'êtes pas sûr que votre établissement dispose d'un entrepôt de données, adressez-vous à l'équipe de votre bibliothèque qui pourra vous aider.

8.2.3. Entrepôt national

Un nouvel entrepôt national, [Recherche Data Gouv](#), permettra de stocker les données orphelines, dites de longue traîne, dont le poids en octet est faible mais dont le poids scientifique peut s'avérer majeur. Il offrira une solution souveraine aux chercheurs pour déposer leurs données de recherche. Ce service sera hébergé sur un datacenter national labélisé et sera complété d'un espace de modération.

L'entrepôt Recherche Data Gouv est organisé en espaces institutionnels de publication et de signalement des données des établissements qui souhaitent participer au projet.

8.2.4. Entrepôts de données généralistes

Vous pouvez également utiliser un entrepôt de données généraliste tel que [Dryad](#) ou [Zenodo](#).

8.2.5. Coffres-forts numériques sécurisés

Si vous travaillez avec des données sensibles qui ne peuvent pas être partagées à grande échelle, vous pouvez envisager d'utiliser un coffre-fort sécurisé pour partager vos données de manière responsable avec des chercheurs spécifiques. Il existe des coffres-forts sécurisés disciplinaires et certaines universités mettent en place des coffres-forts locaux pour fournir un accès contrôlé aux données sensibles.

8.2.6. Data journals

Il existe une liste croissante de data journals émergeant dans toutes les disciplines qui permettent aux chercheurs de publier leurs données. Comme les data journals sont évalués par des pairs, les chercheurs bénéficient d'une autre production sur leurs travaux, digne de citation.

Cette [courte vidéo](#) vous explique comment publier vos données dans un data journal.

Pour plus d'informations sur les data journals, consulter [cette thématique dans DoRANum](#).

9. Conseils pour obtenir le meilleur impact du partage de vos données de recherche

Voici quelques idées pour obtenir un bon retour sur investissement lorsque vous partagez vos données de recherche :

Attribuer un DOI aux données

Attribuez aux données un identifiant unique et pérenne tel qu'un identifiant d'objet numérique (DOI). Les DOI sont des chaînes alphanumériques qui sont liées à l'emplacement numérique de votre objet et à la description des métadonnées correspondantes.

Pour en savoir plus sur l'attribution des DOI, consulter le site web de [DataCite](#).

La plupart des entrepôts de données attribuent un DOI aux jeux de données qu'on a déposés. En ayant un DOI, vos données sont plus faciles à trouver et à citer. Cela favorise l'évaluation.



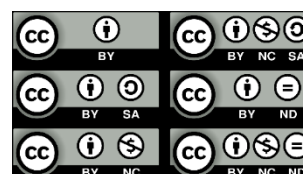
Attribuer une licence aux données

Si vous voulez rendre vos données ouvertes, elles doivent être accompagnées d'une licence. Pour les données fermées ou partagées, il est préférable de lever toute ambiguïté sur ce que les autres peuvent et ne peuvent pas faire avec vos données en leur attribuant une licence.

[Creative Commons](#) propose différents types de licences qui répondront aux besoins de la plupart des chercheurs.

Le site [data.gouv.fr](#) recense les licences de réutilisation autorisées dans le cadre de la loi pour une République numérique pour les "informations publiques (données, documents...)" et codes sources, ainsi que les licences spéciales homologuées.

Des outils comme [License selector](#), [Choose an open source license](#) et [Licentia by inria](#) peuvent également vous aider.



Fournir une citation recommandée

Respectez les normes de votre discipline, si elles existent. S'il n'y a pas de normes, fournissez un moyen recommandé de citer vos données. La citation doit inclure le(s) auteur(s), le titre, l'éditeur, la date et l'identifiant.

Ce [guide du Digital Curation Centre](#) fournit des conseils sur la citation des données.



Fournir une documentation de qualité

Il faut fournir une documentation de qualité, telle qu'une description de la méthodologie de recherche utilisée pour créer les données. Il doit également y avoir une documentation suffisante au niveau du dossier, de l'élément et de la variable pour permettre à un tiers de les comprendre. Il peut s'agir de s'assurer que les feuilles de calcul Excel comportent des descriptions de lignes et de colonnes pertinentes ou qu'un document est joint à l'ensemble des données pour expliquer correctement toute abréviation utilisée.



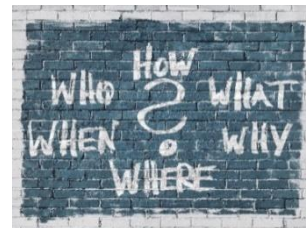
Ce [guide d'Etalab](#) sur la documentation des données peut s'avérer utile.

Ajouter des métadonnées enrichies

Les métadonnées sont un sous-ensemble d'informations documentaires qui utilisent des termes standardisés et sont présentées de manière structurée. Pour être conformes au modèle 5 étoiles des données ouvertes, les métadonnées doivent être lisibles par les machines comme par les humains. Le modèle 5 étoiles suggère également d'établir des liens avec d'autres données pour en augmenter la valeur. Les meilleures métadonnées attribuées à vos données sont les détails - et idéalement un lien permanent - de la publication qu'elles soutiennent !

Pour une liste des standards de métadonnées disciplinaires, consulter

- le site web du [Digital Curation Centre](#),
- le [Metadata Standards Catalog](#) de la RDA
- les [standards de FAIRsharing](#).



Fournir le contexte

Relier ses données à des résultats connexes fournira un cadre contextuel permettant aux réutilisateurs de comprendre les données et la manière dont elles sont produites.

Il est souvent impossible de donner un sens à des données isolées et il est probable qu'elles ne seront pas réutilisées (ou citées !) comme espéré. N'oubliez pas de relier toutes les productions à vous-même également par le biais de votre [identifiant ORCID](#) !



10. Testez vos connaissances

1/6. Les données FAIR sont celles qui sont :

Sélectionnez l'une des options ci-dessous

- ☐ Faciles à trouver, accessibles, interopérables, reproductibles
- ☐ Formatées, accessibles, interopérables, réutilisables
- ☐ Faciles à trouver, précises, interopérables, réutilisables
- ☒ Faciles à trouver, accessibles, interopérables, réutilisables

2/6. À mesure que j'approche de la fin de mon projet, je devrais commencer à penser à rendre mes données FAIR

Sélectionnez l'une des options ci-dessous

- ☐ Vrai
- ☒ Faux

Vous devez réfléchir à la manière dont vous rendrez FAIR tous les résultats que vous partagez dès le début de votre projet et aborder les principes FAIR dans votre plan de gestion des données. Il est également judicieux de procéder à des vérifications périodiques du caractère FAIR de vos données au fur et à mesure et vers la fin de votre projet.

3/6. Pour que les données soient FAIR, elles doivent être ouvertement accessibles à tous ceux qui le souhaitent.

Sélectionnez l'une des options ci-dessous

- ☐ Vrai
- ☒ Faux

Il n'est pas obligatoire de rendre les données ouvertes pour qu'elles soient FAIR. Il peut y avoir de nombreuses bonnes raisons pour lesquelles les données ne peuvent pas être partagées ouvertement.

Il existe des niveaux d'ouverture. Si vous souhaitez que certaines de vos données soient totalement ouvertes à tous, il peut y avoir d'autres données que vous ne pouvez partager qu'avec des groupes ou des individus spécifiques, sous certaines conditions.

Cependant, la description des métadonnées doit pouvoir être trouvée et expliquer clairement les restrictions et la manière de demander l'accès.

4/6. Un plan de gestion des données qui est soumis avec une demande de subvention ne peut pas être modifié.

Sélectionnez l'une des options ci-dessous

- ☐ Vrai
☒ Faux

Votre plan de gestion des données doit être considéré comme un document vivant et doit être mis à jour aussi souvent que nécessaire pour refléter ce qui se passe pendant la durée de votre projet. Le PGD final doit refléter ce qui s'est réellement passé plutôt que ce que vous espériez lorsque vous avez soumis votre proposition.

5/6. En tant que chercheur qui a mené le projet, je suis le seul à déterminer quelles données peuvent être partagées.

Sélectionnez l'une des options ci-dessous

- ☐ Vrai
☒ Faux

Les chercheurs ont un rôle clé à jouer en décidant quelles données peuvent être partagées, mais il est important de noter qu'ils ne sont pas les seules parties prenantes impliquées dans cette décision. Avant de décider du bon niveau d'ouverture de vos données, tenez compte de l'avis des partenaires concernés, tels que les collaborateurs de la recherche et les participants à l'étude.

6/6. L'utilisation d'un entrepôt de données aligné sur les principes FAIR peut aider à rendre vos données FAIR du fait de

Cocher toutes les cases qui s'appliquent

- ☒ L'utilisation de standards de métadonnées reconnus pour décrire vos données qui sont à la fois lisibles par l'homme et par la machine
☐ L'évaluation rigoureuse de la qualité des données
☒ La facilitation de récupération des données déposées (à moins qu'il n'y ait des restrictions d'accès)
☒ La fourniture d'un identifiant pérenne pour vos données

La plupart des entrepôts établis sont en mesure de fournir un identifiant pérenne tel qu'un DOI pour vos données.

La plupart des entrepôts généralistes adhèrent également à des standards de métadonnées tels que Dublin Core ou DataCite. Les entrepôts disciplinaires adhèrent généralement aussi aux standards de métadonnées pertinents pour le domaine. En utilisant un entrepôt établi, vous serez en mesure de fournir un accès en ligne à vos métadonnées et, le cas échéant, à vos données également. Si vous travaillez avec des données très volumineuses ou sensibles, il est conseillé de vérifier que l'entrepôt auprès duquel vous souhaitez déposer vos données peut les traiter. En cas de doute, contactez votre équipe locale de soutien à la recherche pour obtenir des conseils. Cependant, l'entrepôt ne peut pas garantir que des contrôles de qualité rigoureux ont été effectués sur les données. Bien que de nombreux entrepôts effectuent des contrôles de qualité de base, il incombe au chercheur de s'assurer que les données déposées ont fait l'objet de contrôles de qualité rigoureux. Cela devrait faire partie du processus normal de recherche.

Vous êtes maintenant sur la bonne voie pour gérer et partager vos données de recherche.

N'oubliez pas de :

1. bien réfléchir aux données que vous pouvez rendre publiques et à celles qui doivent être protégées et n'oubliez pas de consulter les parties prenantes concernées lorsque vous prenez des décisions
2. mettre à jour votre plan de gestion des données au fur et à mesure de l'avancement de votre projet ;
3. essayer au maximum de rendre vos données faciles à trouver, accessibles, interopérables et réutilisables (FAIR) ;
4. prendre le temps d'attribuer un DOI, de fournir une citation recommandée, de fournir une bonne description et d'attribuer une licence à vos données afin de tirer le meilleur parti du partage.

Vous voulez en savoir plus ? Veuillez consulter les ressources supplémentaires ci-dessous. Vous voulez apprendre quelque chose de nouveau ? Alors veuillez sélectionner votre prochain cours dans notre [menu principal](#).

11. Ressources supplémentaires

- DoRANum (Données de la Recherche : Apprentissage Numérique). <https://doranum.fr/>
- DoRANum. Parcours interactif sur la gestion des données de la recherche. <https://doranum.fr/enjeux-benefices/parcours-interactif-sur-la-gestion-des-donnees-de-la-recherche/>
- CoopIST. Gérer les données de la recherche. <https://coop-ist.cirad.fr/gerer-des-donnees>
- CNIL. Comprendre le RGPD. <https://www.cnil.fr/comprendre-le-rgpd>
- Force11. Guiding principles for Findable, Accessible, Interoperable, and Re-usable data publishing Version B1.0. <https://www.force11.org/fairprinciples>
- GO FAIR Initiative. <https://www.go-fair.org/>
- ANDS. FAIR data training. <https://www.ands.org.au/working-with-data/fairdata/training>
- DTL. FAIR Data Tools. <https://www.dtls.nl/fair-data/find-fair-data-tools/>
- FAIR Cookbook. <https://fairplus.github.io/the-fair-cookbook/content/home.html>
- FAIR Metrics. <http://fairmetrics.org/>
- 5 Star Open Data. <http://5stardata.info/en/>
- Krzysztof J. Gorgolewski, Daniel S. Margulies and Michael P. Milham. Making data sharing count: a publication-based solution. 6 février 2013. <https://www.frontiersin.org/articles/10.3389/fnins.2013.00009/full>
- Heather A. Piwowar and Todd J. Vision. Data reuse and the open data citation advantage. 1er octobre 2013. <https://peerj.com/articles/175/>
- John E. Kratz and Carly Strasser. Making Data Count. 4 août 2015. <https://www.nature.com/articles/sdata201539>
- Marie-Claude Deboin. Créer un identifiant chercheur ORCID ID. Montpellier (FRA) : CIRAD, 4 p. 26 novembre 2015. <https://doi.org/10.18167/coopist/0009>
- Mark D. Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship. 15 mars 2016. <https://www.nature.com/articles/sdata201618>
- Barend Mons et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. 7 mars 2017. <https://doi.org/10.3233/ISU-170824>
- Marie-Claude Deboin. Identifier et rechercher une publication ou un jeu de données par son DOI en 7 points. Montpellier (FRA) : CIRAD, 5 p. 24 novembre 2017. <https://doi.org/10.18167/coopist/0005>
- Karsten Kryger Hansen, Mareike Buss, & Lea Sztuk Haahr. A FAIRy tale. Zenodo. 2018. <http://doi.org/10.5281/zenodo.2248200>
- Data packaging guide. 17 février 2018. <https://github.com/saverkamp/beyond-open-data/blob/master/DataGuide.md>

- Anne-Laure Stérin. Le point sur les données à caractère personnel. 9 novembre 2018. <https://ethiquedroit.hypotheses.org/tag/donnees-personnelles>
- L'ANR met en place un plan de gestion des données pour les projets financés dès 2019. 5 septembre 2019. <https://anr.fr/fr/actualites-de-lanr/details/news/lanr-met-en-place-un-plan-de-gestion-des-donnees-pour-les-projets-finances-des-2019/>
- Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. Présentation du programme Horizon Europe. Septembre 2019. <https://www.horizon2020.gouv.fr/cid145004/presentation-du-programme-horizon-europe-09-2019.html>
- Valérie Gadrat, Yvette Lafosse, Claire Sowinski, Coralie Wysoczynski. Jeu GopenDoRe : le jeu coopératif pour acquérir des bonnes pratiques de gestion et de partage des données de la recherche. 8 novembre 2019. <https://doi.org/10.5281/zenodo.2654827>
- CESSDA Training. Data Management Expert Guide. 31 janvier 2020. <https://zenodo.org/record/3820473#.YDzA7twxk-U>
- Christine Hadrossek, Yvette Lafosse, Pierrette Paillassard, Claire Sowinski, Coralie Wysoczynski. Faut pas s'en FAIR : une extension du jeu coopératif GopenDoRe pour apprivoiser les principes FAIR. 16 décembre 2021. <https://doi.org/10.5281/zenodo.5785662>
- Julie Aucagne, Marguerite Bordry, Camille Desiles, Francine Filoche, Anne Garcia-Fernandez, et al. Vademecum pour la réutilisabilité des données : Groupe de travail Réutilisabilité, Consortium Cahier. [Rapport de recherche] Consortium CAHIER - Huma-Num. Janvier 2022. <https://hal.univ-grenoble-alpes.fr/hal-03630095v1>
- Luc Decker, Hanka Hensens, Caroline Doucouré, Pascal Aventurier. Guide de relecture d'un jeu de données avant publication (Qualité de l'entrepôt de données DataSuds). 20 janvier 2022. <https://doi.org/10.5281/zenodo.5884671>
- FAIRsFAIR et EOSC Synergy. Data Steward Training. 20 janvier 2022. <https://moodle.learn.eosc-synergy.eu/course/view.php?id=132>
- Marie-Claude Deboin. Trouver des plans de gestion de données (PGD) pour s'en inspirer. Montpellier (FRA) : CIRAD, 4 p. 11 février 2022. <https://doi.org/10.18167/coopist/0083>
- Grand Labo. Données de la recherche, quel partage ? 15 février 2022. <https://www.youtube.com/watch?v=xqtrE5D2ixY>
- Inist-CNRS. GopenDoRe : un jeu sur la gestion des données de recherche. Jeu en ligne. 21 février 2022. <https://doranum.fr/enjeux-benefices/gopendore-jeu-sur-la-gestion-des-donnees-de-recherche/>
- Alliance Sorbonne Université. MOOC La science ouverte. Mars 2022. <https://www.fun-mooc.fr/fr/cours/la-science-ouverte/>
- Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. Comité pour la science ouverte. Partager les données liées aux publications scientifiques. Guide pour les chercheurs. Mars 2022. <https://www.ouvrirlascience.fr/partager-les-donnees-liees-aux-publications-scientifiques-guide-pour-les-chercheurs/>
- Cyril Heude. Données de la Recherche : suivez le guide ! 13 avril 2022. <https://sciencespo.libguides.com/donnees-de-la-recherche>
- Université du Québec. Carrefour Gestion des Données de Recherche. 20 avril 2022. <https://uquebec.libguides.com/gdr/>

Remerciements

Ces cours ont été développés en 2018 dans le cadre du projet FOSTER Plus (Fostering the practical implementation of Open Science in Horizon 2020 and beyond) en réutilisant des contenus disponibles librement produits par une série de fournisseurs de contenus, notamment [DataOne](#), [Research Data Netherlands](#), [Open Data Institute](#), [European Data Portal](#), [Digital Curation Centre](#), [UK Data Service](#), [CESSDA ERIC](#), [DARIAH](#), [ELIXIR](#), [Software Sustainability Institute](#), [FOSTER](#) et bien d'autres qui développent activement des ressources éducatives libres liées à la science ouverte. Ils ont été mis à jour en 2021.

Ils ont été traduits et adaptés à la France par l'Inist-CNRS en 2022.

Les cours sont présentés dans un style similaire à celui utilisé par l'Open Data Institute (ODI) et l'European Data Portal, dans l'espoir que cela permettra à notre contenu d'accroître le corpus de ressources liées à la science ouverte déjà produites et de rendre leur réutilisation collective plus transparente. À cette fin, nous avons aussi fait usage de l'outil de création [Adapt](#), également utilisé par l'ODI et l'European Data Portal.

Nous avons utilisé une variante de l'approche des études de cas développée par [l'Open Science Monitor](#) de la Commission européenne pour aider à illustrer les outils et les initiatives utiles du point de vue des disciplines.

Le contenu de cette ressource pédagogique est sous licence CC-By, sauf indication contraire.