

Le Minimum à FAIR

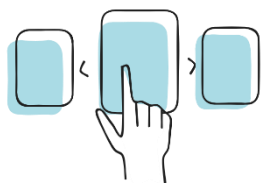
<https://view.genially.com/665dcf34ed494b00145845ec>

21 novembre 2024

Table des matières

1. Choisir un entrepôt de données adapté dans lequel déposer ses données.....	2
2. Vérifier que l'entrepôt dans lequel les données sont déposées leur attribue un identifiant pérenne	3
3. Documenter un maximum les données et utiliser des vocabulaires contrôlés de sa communauté disciplinaire.....	4
4. Définir clairement les droits d'accès aux données au moment du partage	5
5. Privilégier des formats de fichiers ouverts, non-propriétaires.....	6
6. Attribuer une licence de réutilisation à ses jeux de données	7
7. Mentionner les productions liées aux données (publications, thèses, autres jeux de données, codes et logiciels.....)	8

1. Choisir un entrepôt de données adapté dans lequel déposer ses données



Tous les entrepôts ne permettent pas d'atteindre un niveau suffisant de « FAIRisation » des données. C'est pourquoi le choix d'un entrepôt de données est **déterminant pour rendre les données de recherche Faciles à trouver, Accessibles, Interopérables, Réutilisables (FAIR)**.

[Accéder au logigramme « Où publier vos données ? »](#)

Étape 1 : Planification

Dès la planification, je me renseigne sur les entrepôts de données existants dans mon domaine, dans mon institut... Je peux effectuer des recherches dans des annuaires et catalogues de référence (**CatOPIDoR, re3data, FAIRsharing...**).

N.B. : un entrepôt de données spécialisé dans mon domaine ou proposé par mon institut aura plus de chances d'être adapté à mes données.

Étape 4 : Préparation du dépôt

Je m'assure de choisir **un entrepôt de confiance** et adapté à mes données. Cela me permettra de les rendre plus **Faciles à trouver, Accessibles, Interopérables et Réutilisables**.

2. Vérifier que l'entrepôt dans lequel les données sont déposées leur attribue un identifiant pérenne



L'attribution d'un identifiant pérenne est **indispensable pour rendre des données FAIR.**

Je dois donc vérifier que mes jeux de données ont un identifiant pérenne. Si l'entrepôt que je choisis n'attribue pas automatiquement d'identifiant pérenne, je contacte **l'atelier de la donnée** le plus proche de mon établissement pour me faire conseiller.

[Accéder à la liste des ateliers de la donnée](#)

Étape 1 : Planification

Dans un premier temps, je vérifie **si les entrepôts pressentis attribuent un identifiant pérenne aux jeux de données** (DOI, ARK...). Si ce n'est pas le cas, je recherche un autre moyen d'attribuer un identifiant pérenne à mes données avant le dépôt (ex : DataCite, référencement sans dépôt dans Recherche Data Gouv...). Je peux me faire conseiller par un **atelier de la donnée**.

N.B. : s'il n'existe aucun atelier de la donnée à proximité de mon établissement, je peux contacter directement **[le support de Recherche Data Gouv.](#)**

Étape 5 : Dépôt

Au moment du dépôt, je vérifie qu'un identifiant pérenne est bien **associé à mes jeux de données**.

Étape 6 : Publication pour ouverture ou partage

Lorsque mon jeu de données est publié, je vérifie que son identifiant pérenne est **visible dans les métadonnées** et qu'il est **actif**.

3. Documenter un maximum les données et utiliser des vocabulaires contrôlés de sa communauté disciplinaire



Pour être FAIR, je donne le plus d'**informations contextuelles** possible sur les données. Par exemple, je documente le processus de production ou de réutilisation des données. Je note la version du logiciel utilisé, les paramétrages des instruments de mesure, les unités utilisées, le nom des variables... Pour décrire les données, je me réfère à des **normes** (ex : normes ISO), **standards**, **vocabulaires communs**. J'indique des **mots-clés** à partir de vocabulaires contrôlés (ex : taxonomie, thesaurus, glossaire, etc.). Toutes ces informations sont des métadonnées.

[Accéder à une fiche synthétique sur les métadonnées](#)

Étape 1 : Planification

Dès la planification, je réfléchis à la façon dont je vais décrire mes données : quelles **informations** je vais renseigner, quel **standard de métadonnées** je vais utiliser, sur quels **vocabulaires contrôlés** (liste de mots-clés, taxonomie, thesaurus, glossaire...) je vais m'appuyer.

Étape 2 : Collecte, création, stockage

Je récupère les métadonnées qui sont produites automatiquement par les différents appareils de mesure, aussi appelées **métadonnées embarquées** (ex : date et heure générées automatiquement lors de la prise de photos, coordonnées GPS...).

Je note au fur et à mesure toutes les informations utiles à la compréhension de mes données et qui ne sont pas générées automatiquement (méthodes, paramétrages des instruments, unités, noms des variables...). Il s'agit des **métadonnées enrichies**.

Étape 3 : Analyse, traitement, calcul

Pour cette étape, se référer aux informations se trouvant à l'**étape 2 : collecte, création, stockage**.

Étape 4 : Préparation du dépôt

Je rédige un [fichier README](#) pour centraliser toutes ces informations.

Je prépare les **métadonnées manquantes** : **liées au dépôt**, **liées au standard de métadonnées** proposé par l'entrepôt...

Étape 5 : Dépôt

Je **complète** les champs de **métadonnées de l'entrepôt**. Je renseigne au minimum les métadonnées obligatoires et un maximum de métadonnées facultatives.

N.B. : plus les métadonnées seront riches, précises et en conformité avec les standards établis au sein de ma communauté, plus les données seront FAIR.

4. Définir clairement les droits d'accès aux données au moment du partage



Dans une démarche FAIR, je dois définir **qui** a le droit d'accéder aux données, **comment** (authentification, demande d'autorisation...), **sous quel délai** (embargo).

Étape 1 : Planification

Dès la planification, je réfléchis aux types de données (produites ou réutilisées) qui pourraient faire l'objet d'une **restriction d'accès** (ex : exploitation commerciale, partenariat public-privé, confidentialité...).

Étape 4 : Préparation du dépôt

Je choisis un entrepôt qui propose des **conditions d'accès conformes à mes besoins** (accès ouvert, fermé, restreint, embargo).

Étape 5 : Dépôt

Je dépose mes jeux de données en définissant les **conditions d'accès adéquates**.

Étape 6 : Publication pour ouverture ou partage

Je vérifie que les **modalités d'accès** aux données sont **telles que je les ai définies** et qu'elles sont **claires et explicites** pour les utilisateurs (ex : savoir qui contacter, comment...).

5. Privilégier des formats de fichiers ouverts, non-propriétaires



Voici quelques exemples de **formats de fichiers fermés** et d'équivalents **ouverts** :

Fermé	Ouvert
.xls	.csv ; .netcdf
.doc	.odt
.wma	.mp3

[Accéder à une ressource DoRANum sur les formats de fichiers ouverts et fermés.](#)

Étape 2 : Collecte, création, stockage

Dès cette étape, je **privilégie** l'utilisation de **logiciels** et de **formats de fichiers ouverts** et **interopérables** pour que mes fichiers soient lisibles et modifiables indépendamment de l'usage d'un logiciel unique.

Étape 3 : Analyse, traitement, calcul

Pour cette étape, se référer aux informations se trouvant à l'**étape 2 : collecte, création, stockage**.

Étape 4 : Préparation du dépôt

Si j'ai utilisé des formats de fichiers fermés et que cela est possible, **je les convertis dans des formats ouverts** pour permettre, au moment du partage, la réutilisation et l'interopérabilité de mes fichiers de données.

6. Attribuer une licence de réutilisation à ses jeux de données

Pour choisir une licence appropriée, je prends en compte les **contraintes légales et éthiques** liées à mes données.

[Accéder au guide des licences ouvertes.](#)

[Accéder à la ressource « Les licences de réutilisation dans le cadre de l'Open data et de la loi pour une République numérique ».](#)

Étape 1 : Planification

Dès la planification, je réfléchis au partage de mes données en me renseignant sur les **différentes licences** et en croisant les exigences du financeur du projet avec les contraintes légales et éthiques.

Étape 5 : Dépôt

J'indique dans les métadonnées la **licence de réutilisation choisie pour mes données**.

La licence indique aux potentiels réutilisateurs **ce qu'ils ont le droit et n'ont pas le droit** de faire avec les données.

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



Si d'autres productions liées à mes données ont des **identifiants pérennes**, je pense à **les mentionner dans les métadonnées**.

Étape 5 : Dépôt

Je fais **référence dans les métadonnées** aux entités, activités, personnes et productions qui sont en relation avec mes données via leurs identifiants pérennes pour permettre l'interconnexion des données.

Par exemple, je cite l'identifiant pérenne de : l'institution propriétaire des données, les producteurs des données, le data paper décrivant les données, la publication liée aux données, les jeux de données, codes sources et logiciels associés...