

Les données FAIR : Qu'est-ce que c'est ?

Un des objectifs de la gestion des données est de faciliter la découverte et la réutilisation des connaissances scientifiques par les humains et les systèmes informatiques. Les principes FAIR (Findable, Accessible, Interoperable, Reusable) fonctionnent comme une ligne directrice pour ceux qui veulent atteindre cet objectif. La Commission Européenne a repris ces principes et demande à tous les projets financés par le programme H2020 de les suivre.

Findable / Facile à trouver

Exigence : Les données doivent être faciles à trouver à la fois par les humains et les systèmes informatiques.

Points d'attention :

- les données doivent avoir un **identifiant unique et pérenne** ;
- les données sont décrites par des **métadonnées** riches ;
- les données sont enregistrées ou indexées dans une **source interrogeable**.

Accessible

Exigence : les données doivent être stockées à long terme de façon à ce qu'elles puissent être facilement accessibles et/ou téléchargées. Des données peuvent être FAIR sans être directement accessibles (accès restreint, embargo...) mais les métadonnées doivent être accessibles.

Points d'attention :

- Déposer les données dans un **entrepôt de confiance**, fondé sur un modèle économique connu et **pérenne**, et de préférence **certifié** ;
- Bien définir les **conditions d'accès** et la **licence de diffusion** à la fois pour les métadonnées et pour les données ;
- Si possible, les données et métadonnées doivent être en **Open Access** ;
- Si les données sont en accès restreint ou sous embargo, les **métadonnées doivent être accessibles** pour signaler l'existence des données tout en préservant leur protection.



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Interoperable

Exigence : les données doivent être faciles à combiner avec d'autres jeux de données, à la fois par les humains et par les systèmes informatiques.

Points d'attention :

- Privilégier des **formats libres et ouverts** (au moins pour la diffusion et l'archivage des données) ;
- Si des logiciels de traitement et d'analyse des données sont mis au point, mettre à disposition le **code source du logiciel** et le documenter pour faciliter sa réutilisation ;
- Privilégier les **standards de métadonnées** et les **vocabulaires standardisés** ;
- Les données doivent si possible **faire référence à d'autres données**

Reusable / Réutilisable

Exigence : les données doivent être prêtes à être réutilisées pour une future recherche et à être traitées en utilisant des méthodes informatiques.

Une méthode d'évaluation des jeux de données : Mon jeu de données est-il FAIR ?

Cette méthode, proposée par l'institut néerlandais DANS (Data Archiving and Networked Services) a pour objectif d'évaluer la qualité des jeux de données au regard des principes FAIR. Elle consiste à établir un score de 1 à 5 pour les trois premiers critères : trouvable, accessible et interopérable. A partir des scores obtenus pour ces trois critères, vous pourrez déterminer par calcul si votre jeu de données est réutilisable. La qualité de votre jeu de données pourra donc être représentée schématiquement comme sur l'exemple à droite. Le DANS propose d'apposer ce schéma à côté de chaque jeu de données dans les entrepôts de données. Ce n'est pas encore le cas, mais cela vous permet visuellement de savoir si votre jeu de données est FAIR.



Nous vous proposons ci-dessous quelques tableaux¹ pour évaluer vos jeux de données selon cette méthode. Vous pouvez également utiliser [l'outil d'évaluation des données FAIR](#) en ligne (ainsi que le [guide](#) pour vous aider à répondre aux questions), proposé par [UK Data Service](#).

¹ **Source** : ces tableaux ont été construits à partir du webinaire de l'EUDAT : [FAIR Data in Trustworthy Data Repositories Webinar](#)

Le jeu de données est trouvable

	Identifiant pérenne	Métadonnées
☆	✗	✗
☆☆	✓	✗
☆☆☆	✗	Suffisantes*
☆☆☆☆	✓	Suffisantes*
☆☆☆☆☆	✓	Riches**

* Les métadonnées sont suffisantes pour comprendre le jeu de données.

** Les métadonnées indiquent comment les données peuvent être réutilisées.

Le jeu de données est accessible

	Licence de diffusion	Métadonnées accessibles	Données publiques	Restriction d'accès
☆	✗	✗	✗	
☆☆	✗	✓	✗	
☆☆☆	✓	✓	✗	Accès fortement restreint*
☆☆☆☆	✓	✓	✓	Inscription
☆☆☆☆☆	✓	✓	✓	Aucune**

* Exemples de restrictions :

- Confidentialité : des questions éthiques limitent l'accès
- Des logiciels sont nécessaires pour accéder aux données
- Intérêt commercial (obligation de payer un journal pour accéder aux données)
- Période d'embargo (accès pendant seulement 24h par exemple)
- Accès uniquement à un groupe spécifique

** Données en Open Access (licence CC0 ou CC-BY par exemple)

Le jeu de données est interopérable

	Format ouvert	Appartenance à la liste de formats privilégiés*	Vocabulaire standardisé**	Lien avec d'autres données
☆	✗	✗	✗	✗
☆☆	✓	✗	✗	✗
☆☆☆	✓	✓	✗	✗
☆☆☆☆	✓	✓	✓	✗
☆☆☆☆☆	✓	✓	✓	✓

* Les formats privilégiés sont des formats de fichiers dans lesquels l'institut néerlandais DANS (Data Archiving and Networked Services) est confiant et qui offriront les meilleures

garanties à long terme en termes d'utilisabilité, d'accessibilité et de durabilité : [voir la liste des formats privilégiés](#)

** Afin de connecter les jeux de données de différents projets de manière significative, il est important qu'ils utilisent des méthodes communes de codage des valeurs de données. De nombreuses communautés de recherche ont défini des vocabulaires standard, des ontologies, des schémas de codage / marquage / balisage, des thésaurus, des systèmes de classification et d'autres graphiques de connaissances sémantiques pour y parvenir.

Le jeu de données est réutilisable

La réutilisabilité étant une notion très subjective, il est difficile d'établir un score. Vous pouvez calculer la réutilisabilité de votre jeu de données à partir du score obtenu aux 3 premières conditions : $(F+A+I) / 3$