

Historique du document			
Date	Version	Auteur	Commentaires
23/02/2021	1.0	GT science ouverte	Création du document

Des principes ? Pour quoi FAIR ?

Les principes FAIR

Les principes FAIR ont été établis pour « Accroître la capacité des ordinateurs à trouver et utiliser des données, et faciliter leur réutilisation par des individus »¹. En somme, ils vous permettent de trouver et réutiliser facilement des données utiles à vos recherches, tout en étant certain de la fiabilité de ces données. Il s'agit d'éviter une situation aussi caricaturale que celle-ci :

Vous voudriez réutiliser des données télescopiques spectrales, mais les fichiers sont des tables en texte ASCII, sans entête. Vous arrivez à décoder la première colonne (qui ressemble à une longueur d'onde, sans certitude sur l'unité), mais impossible de reconstruire l'unité de la mesure, ni les conditions d'observation ou de pointage.

Conclusion : c'est une impasse, les données sont inexploitable !

Vous avez dit FAIR ?

FAIR est l'acronyme de :

- Findable (Facile à trouver)
- Accessible
- Interoperable
- Re-usable (Ré-utilisable)

Ces quatre principes sont applicables à vos jeux de données, **ils doivent en garantir une bonne gestion et une réutilisation facile.**

¹ « enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. » Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. « The FAIR Guiding Principles for Scientific Data Management and Stewardship ». *Scientific Data* 3, n° 1 (15 mars 2016), p.1. <https://doi.org/10.1038/sdata.2016.18>.

🔗 Mais comment FAIR ?

Le tableau ci-dessous explique en quelques lignes les bonnes pratiques à mettre en place afin de respecter ces quatre principes.

Principes FAIR	En pratique, cela signifie :
Findable (Facile à Trouver)	Accompagner ses données de métadonnées riches et structurées . Des données décrites correctement seront plus faciles à trouver.
	Attribuer un identifiant pérenne (DOI) à ses données et mentionner cet identifiant dans ses métadonnées.
	Quand les données sont ouvertes, les déposer dans un entrepôt de données fiable (PADC, CDS, Zenodo, etc) où celles-ci seront indexées et où on pourra les retrouver grâce à leurs métadonnées.
Accessible	Choisir un protocole libre et ouvert (par exemple: HTTP, ou le protocole TAP de l'IVOA). Cela évite toute dépendance à un protocole propriétaire (par exemple : Dropbox), ou obsolète (par exemple : FTP)
	Si besoin, identifier les données nécessitant un accès restreint , et les rendre accessibles avec authentification (à voir avec l'entrepôt de données)
	Maintenir à jour les métadonnées , tout au long de la vie des données, y compris si celles-ci sont supprimées.
	Privilégier des standards de métadonnées reconnus par votre communauté scientifique (exemple : FITS/WCS)
Interopérable	Utiliser un vocabulaire contrôlé, structuré, accessible et largement partagé pour décrire ses données et métadonnées. Ces données doivent pouvoir être lues/exploitées par n'importe qui et n'importe qui (humain, machine, workflow...).
	Utiliser un format reconnu, dont les spécifications sont publiques (par exemple PDF, FITS, etc), et qui permet d'associer directement les métadonnées aux données.
Reusable - Réutilisable	Penser à attribuer une licence à ses données . L'utilisateur saura quel usage il peut en faire ou ne pas en faire tout en respectant la licence choisie. Cette licence doit être claire et accessible.
	Définir la durée de vie des données , avec au besoin, une période d'embargo.
	Détailler la provenance de ses données (Où/quand/comment ont-elles été collectées? Qui en est à l'origine ?...). Il faut que les utilisateurs puissent remonter à la source et estimer la qualité de la donnée.
Référence :	https://doranum.fr/enjeux-benefices/principes-fair/

Faut-il ouvrir toutes ses données ?

L'objectif de la science ouverte est d'ouvrir les données, ce qui est généralement requis pour des données de recherche financée sur fonds publics. Les principes FAIR permettent de les rendre visibles et de les diffuser. Certaines données ne pourront pas être librement diffusées - cas des données sensibles, des données utilisées pour un brevet... Respecter les principes FAIR ne veut pas forcément dire tout ouvrir, mais signifie avoir une bonne gestion de ses données dès le début du processus de recherche jusqu'à la conservation et la réutilisation des données.

On ne le répètera jamais assez les données doivent être « Aussi ouvertes que possible, aussi fermées que nécessaire »

Pour quoi FAIR ?

Pour plus de visibilité et une meilleure valorisation de vos travaux. Si vos publications sont reliées à vos données et que vous donnez accès à celles-ci, cela renforce la confiance en vos publications et leur apporte davantage de crédit. Cela permet aussi d'éviter de dupliquer des données ou des résultats déjà publiés.

Parce que des données mal documentées ou mal conservées sont des données perdues ! Les principes FAIR facilitent la bonne gestion des données, pendant et après le travail de recherche. L'objectif d'un travail de recherche est d'être lu, vérifié, cité, réutilisé. Ceci est impossible si personne n'a accès aux informations, si le format choisi pour le stockage des données n'est plus lisible, si l'information est perdue, ou si l'adresse url d'accès n'est plus disponible !

Pour une science citoyenne : vos publications, vos données gagnent à être mise à la disposition de tous, mais pour ce faire, il faut qu'elles soient compréhensibles et manipulables, et accompagnées de métadonnées.

Pour fertiliser la recherche : vos données peuvent être utiles à tous et être réutilisées pour d'autres recherches.

Pour se conformer aux pratiques d'intégrité scientifique.

Être FAIR c'est faire les bons choix !

Et en pratique, les principes FAIR à l'Observatoire ça donne quoi ?

- L'utilisation des interfaces de l'Observatoire Virtuel sur les collections de données distribuées par PADC.
- La mise en place de DOI sur les collections de données et les documentations associées.
- Le plan de gestion de données, qui aide à valoriser les données.
- Un fichier texte « ReadMe » ne suffit pas. Il faut des informations structurées, par exemple, pour être insérée dans Vizier au CDS.

Pour aller plus loin

<https://dorum.fr/enjeux-benefices/principes-fair/>

<https://www.go-fair.org/fair-principles/>

Hodson, Jones et al. (2018) Turning FAIR data into reality. Interim report of the European Commission Expert Group on FAIR data. <https://doi.org/10.5281/zenodo.1285272>

Pour être convaincu de l'intérêt des principes FAIR tout en se divertissant :

<https://www.youtube.com/watch?v=N2zK3sAtr-4>

Pour toute question, vous pouvez envoyer un mail à l'adresse
la.bibliotheque@obspm.fr