

# FAIR: guide des bonnes pratiques

(prototype. version 3: 14/11/2021)

2021

Réalisé par: *Grunin A.*

License: CC BY-NC-ND 4.0



## Table des matières

1 Présentation des principes FAIR .....	3
1.1 Qu'est-ce que FAIR? .....	3
1.2 Ce que FAIR n'est pas .....	3
1.3 Pourquoi FAIRiser les données? .....	4
1.4 Qui FAIRise quoi? .....	6
1.5 Des données au dépôt de données: quelques notions principales ..	9
2 Étapes de la FAIRisation .....	19
Bibliographie .....	23
Annexe 1. Principes FAIR .....	25
Annexe 2. Principes FAIR étendus et les différents acteurs .....	27
Annexe 3. Choix d'un dépôt de données .....	30
Annexe 4. Modèle (provisionnel) des métadonnées .....	31

# 1 Présentation des principes FAIR

## 1.1 Qu'est-ce que FAIR?

FAIR est une abréviation de "**Findable (= Facile à trouver), Accessible, Interoperable (= Interopérable), Reusable (= Réutilisable)**". Il s'agit de principes directeurs ou de recommandations (et non pas d'un standard) qui doivent être appliqués aux données de la recherche. Le but de la mise en place des principes FAIR est de rendre les données scientifiques plus accessibles et plus compréhensibles et leurs sauvegarde et réutilisation plus pérennes.

Les principes FAIR ont une visée multidisciplinaire; ils doivent être applicables à toutes les données de la recherche, de la physique nucléaire aux études littéraires. Cela a imposé certaines limites dans la formulation de ces principes. De fait, pour pouvoir être utilisables dans des domaines très divers, ces principes doivent rester les plus généraux possible. Ainsi, l'interprétation de ces principes et surtout leur application concrète restent le plus souvent à l'appréciation du responsable de chaque projet et peuvent donc varier, de façon considérable, d'un projet à l'autre, d'une discipline à l'autre.

Un autre écueil dans la mise en oeuvre des principes FAIR est leur relative nouveauté; FAIR et ses principes n'ont été définis qu'en 2016 [18]. De ce fait, le travail sur l'amélioration et le développement de ces principes, ainsi que sur de différents outils de FAIRisation, se poursuit toujours. À l'heure actuelle, il peut parfois s'avérer délicat de trouver des exemples concrets de données qui suivent fidèlement tous les principes FAIR (quel que soit le domaine de la recherche auquel ces données appartiennent).

## 1.2 Ce que FAIR n'est pas

**FAIR data ne signifie pas Open Data ou Free Data.** Les données peuvent être protégées contre la réutilisation et même contre un simple accès (s'il s'agit, par exemple, de données sensibles, classifiées

ou privées). Un embargo (accessible dans X années) peut également être imposé aux données. Dans le cadre de la mise en oeuvre des principes FAIR, les données doivent être *“ouvertes autant que possible et fermées autant que nécessaire”* [11].

**FAIR data ne signifie pas “archivage à vie”.** La sauvegarde et l'archivage des données sont étroitement liés au lieu où les données sont déposées (dépôts institutionnels ou publics, centres de données privés, serveurs locaux, etc.). Cela peut induire des coûts supplémentaires et le financement de l'archivage des données à très long terme n'est jamais certain. En outre, les dépôts ou les centres de données eux-mêmes peuvent disparaître avec le temps. Ces différentes raisons peuvent conduire à la disparition des données. Ce problème peut être pallié, à certains égards, par l'indexation uniquement des métadonnées (cf. *infra* Annexe 1, principe FAIR:A2. Ce principe ne précise pas toutefois où ces seules métadonnées doivent être stockées).

### 1.3 Pourquoi FAIRiser les données?

Le travail avec les données numériques n'est pas toujours aisé et plusieurs obstacles et problématiques peuvent apparaître. Chacun des principes FAIR vise à répondre à un de ces problématiques.

**Compréhension des données.** Il n'est pas rare qu'une fois que l'article ou le projet est terminé, les données utilisées ou produites se retrouvent encodées dans un format dont on a déjà oublié le nom qui ne s'ouvre qu'avec un logiciel spécifique qui ne fonctionne plus que sur des ordinateurs datant d'il y a dix ans et où toutes les données sont dans des colonnes dont les noms sont des abréviations dont on a déjà oublié la signification exacte... Ces données sont donc souvent définitivement perdues et ne peuvent plus être utilisées par d'autres et, très souvent, par l'auteur des données lui-même.

Pour pallier ce problème, l'initiative FAIR introduit **le principe “Interopérable (Interopérabilité)”**. Ce principe se résume aux consignes suivantes: pour vos données privilégiez les formats de fichiers et les logiciels non-propriétaires, largement connus et répandus (c'est-à-dire qu'ils ont plus de chances de survivre pendant longtemps que des formats propriétaires, obscurs et/ou peu connus); si vous utilisez les

abréviations, les balises, la mise en page ou tout autre type de marquage pour vos données, fournissez des explications détaillées et complètes afin que vos données puissent être comprises et interprétées facilement par toute personne (ou machine).

**Sauvegarde / Archivage des données.** Il n'est pas rare qu'une fois que l'article ou le projet est terminé, les données utilisées ou produites se retrouvent éparpillées dans différents fichiers, sur différentes clés USB ou sur les ordinateurs de différents collaborateurs. Par exemple, le texte de la lettre de Léon III à Charlemagne sera dans un fichier "*LettreLéonCh.doc*" sur l'ordinateur du bureau, les images du manuscrit de cette lettre dans un dossier "*Images février 2015 et autres documents*" sur une clé USB rouge qui vous avez prêtée à un ami et l'analyse de la lettre dans un fichier "*A-l-III-à-Ch.txt*" sur votre ordinateur portable personnel qui ne s'allume plus depuis mercredi dernier... Le plus souvent il est extrêmement délicat, voire impossible, de restituer ces éléments dans un ensemble cohérent; les données sont donc perdues pour les autres et souvent pour l'auteur des données lui-même.

Pour pallier ce problème, l'initiative FAIR introduit **les principes "Findable (Facile à trouver)" et "Accessible"**. Ces principes se résument aux consignes suivantes: attribuez à chaque jeu de données un identifiant unique et pérenne (par exemple DOI); sauvegardez vos données dans un dépôt institutionnel fiable; grâce aux métadonnées (données sur des données) explicitez ce que vos fichiers contiennent (titre exact, auteur de données, date de production, résumé du contenu, mots-clés, etc.); si votre jeu de données est constitué de plusieurs sous-éléments (textes, images, etc.) reliez ces sous-éléments entre eux (grâce au référencement croisé dans les métadonnées). Suivre ces consignes permet à toute personne et à l'auteur des données lui-même de trouver facilement ces données, comprendre/se rappeler ce que contient chaque fichier et de pouvoir reconstituer l'ensemble des sous-éléments d'un jeu complexe de données.

**Droits de la re/utilisation / Autorité des données.** Il n'est pas rare qu'une fois que l'article ou le projet est terminé, les données utilisées ou produites ont été transmises par un chercheur à un autre qui les a légèrement modifiées (en corrigeant les erreurs ou en le mettant à jour) puis retransmises à un autre chercheur qui les modifie à son tour

... et qu'entre-temps la législation a changé en rendant une partie des données sensibles ou interdites à la diffusion. L'autorité de ce type de données (qui est auteur de ces données? comment les citer?) et leur droit de l'utilisation (données ouvertes et publiques? données privées? classifiées? sensibles?) ne sont pas toujours très clairs.

Pour pallier ce problème, l'initiative FAIR introduit **le principe “Reusable (Réutilisable)”**. Ce principe se résume aux consignes suivantes: les différents droits/ licences (droit d'auteur, droit d'utilisation, etc.) liés à chaque jeu de données ainsi que la provenance et l'historique des données (qui est l'auteur des données, qui les a collectées, qui les a modifiées, etc.) doivent être clairement mentionnés dans les métadonnées. Suivre ces consignes permet à toute personne et à l'auteur des données lui-même de savoir s'il est possible ou non, du point de vue légal, de re/utiliser ces données et si elles sont re/utilisables de savoir qui sont ses auteurs et comment les citer.

## 1.4 Qui FAIRise quoi?

*La présentation qui suit fait appel à un écosystème FAIR basé sur un dépôt de données institutionnel (sur les différents écosystèmes FAIR, voir infra “1.5 Des données au dépôt de données: quelques notions principales / Écosystème FAIR”).*

Les principes FAIR s'appliquent aussi bien aux **métadonnées** qu'aux **données** elles-mêmes (voir *infra* Annexe 1). Le respect des principes qui s'appliquent aux données doit être assuré par **l'auteur des données**, tandis que les principes qui s'appliquent aux métadonnées sont à la charge **du dépôt où les données sont entreposées**.

**Le principe “Findable (Facile à trouver)”** est atteint avant tout par le biais de la création des **métadonnées** associées au jeu de données et par la mise de ce jeu de données dans un **dépôt de données**. De fait, lors de l'ajout d'un jeu de données dans un dépôt spécialisé, c'est souvent ce dernier qui se charge de la création d'un identifiant unique (PID) et qui propose des formulaires de renseignements à compléter afin de créer des métadonnées associées. Le choix d'un dépôt de données est donc crucial: le dépôt doit fournir un PID valide et unique,

le dépôt doit fournir un formulaire le plus complet possible des métadonnées et, enfin, c'est le dépôt qui doit créer les métadonnées facilement reconnaissables et indexables par les moteurs de recherche (en utilisant les langages et les standards largement connus pour les métadonnées, par exemple RDF, OWL, JSON, etc.).

*Dès lors, il est possible de dire que c'est avant tout le dépôt où les données sont déposées qui doit répondre au principe "Findable (Facile à trouver)".*

**Le principe "Accessible"** est atteint avant tout grâce au **dépôt où les données sont déposées**. En effet, le dépôt de données doit fournir un moyen (ouvert, gratuit et largement utilisé) d'accéder aux données. En pratique cela signifie qu'une fois que les données sont mises dans un dépôt, ce dernier doit vous fournir un lien (URL) qui permet d'accéder aux données. Le choix de dépôt de données est donc également crucial pour ce principe; il est indispensable de prendre en compte les moyens que chaque dépôt vous propose pour accéder à vos données. (À noter que le protocole HTTP (=adresse web), bien que le plus connu n'est pas le seul moyen d'accéder aux données digitales. D'autres protocoles existent également, par exemple FTP, SMTP, Microsoft Exchange Server, etc.).

*Dès lors, il est possible de dire que c'est avant tout le dépôt où les données sont déposées qui doit répondre au principe "Accessible".*

**Le principe "Interoperable (Interopérabilité)"** est atteint aussi bien par des **métadonnées** que par des **données elles-mêmes**.

En ce qui concerne **les données**, il est indispensable qu'elles soient encodées dans un format largement connu et répandu et qu'elles soient compréhensibles (c'est-à-dire que toutes les abréviations, balises, etc. doivent être explicités). En outre, les données doivent s'appuyer sur un standard/vocabulaire déjà bien documenté (voir par exemple ce recensement des standards répandus: <http://rd-alliance.github.io/metadata-directory>). Ces tâches sont **à la charge de l'auteur des données**.

En ce qui concerne **les métadonnées**, elles doivent également être dans un format largement connu et utilisé (par exemple RDF, OWL, JSON, etc.) et elles aussi doivent être compréhensibles (c'est-à-dire que les métadonnées doivent s'appuyer sur un des standards prévus

pour les métadonnées, par exemple DublinCore, Metadata Object Description Schema, etc. ). Les métadonnées doivent ainsi être compréhensibles aussi bien aux humains (qui, grâce à ces métadonnées, pourront avoir un premier aperçu des données) qu'aux machines (notamment aux moteurs de recherche qui pourront, grâce à ces métadonnées, indexer les données correctement). Cette tâche de l'interopérabilité des métadonnées **est à la charge de celui qui encode les métadonnées, c'est-à-dire le dépôt de données.**

*Dès lors, il est possible de dire que le principe "Interopérabilité" est assuré par l'auteur des données pour les données et par le dépôt où les données sont déposées pour les métadonnées.*

**Le principe "Reusable (Réutilisable)"** est atteint avant tout grâce aux **métadonnées**. Ce principe stipule que certains types d'informations (provenance des données, historiques des changements et des modifications, les droits d'auteurs, les droits d'utilisation, etc.) doivent nécessairement figurer dans les métadonnées associées au jeu de données. Il est indispensable de souligner toutefois que les métadonnées ne se résument pas aux informations demandées par le principe "Reusable (Réutilisable)", les métadonnées doivent être "complètes" (voir *supra* le principe "Findable") et contenir également un ensemble des autres informations (voir *infra* Annexe 4).

*Dès lors, il est possible de dire que ce principe doit être assuré **aussi bien par le dépôt de données** qui doit fournir un formulaire des métadonnées adéquates (c'est-à-dire le formulaire qui contient des champs sur l'historique des données, sur leur provenance, sur leurs droits d'auteurs, etc.) **que par l'auteur des données** qui a l'obligation de remplir ce formulaire.*

**En conclusion**, on peut dire que le processus de la FAIRisation est assuré par **le dépôt de données** (qui se charge du respect des principes "Findable", "Accessible", "Interopérable [pour les métadonnées]" et "Reusable") et par **l'auteur des données** (qui doit assurer le respect du principe "Interopérable [pour les données]" et remplir tous les champs du formulaire des métadonnées fournies par le dépôt de données).

Il est nécessaire de souligner que le rôle que joue **le dépôt de données** peut être rempli de façon différente. (Sur ce point voir *infra* "1.5 De données au dépôt de données...". Sur les différentes tâches



accomplies par l'auteur des données et le dépôt de données, voir *infra* Annexe 2).

## 1.5 Des données au dépôt de données: quelques notions principales

Les brèves présentations qui suivent ne sont en aucun cas des définitions arrêtées et elles ne donnent que quelques lectures possibles de notions complexes. L'objectif ici est d'offrir un premier aperçu de ces concepts clés et de proposer quelques repères pour mieux comprendre les motivations et les applications possibles de certains principes FAIR.

### Données / Métadonnées

Il est impossible et inutile de dresser dans cette présentation la liste de toutes les définitions des "données" ou des "métadonnées" qui existent. Cependant, dans le cadre de la FAIRisation, il est indispensable de faire la différence entre les données, les métadonnées et leurs différents types, le modèle de données et l'objet d'étude.

Données vs Métadonnées. Bien que le mot "métadonnées" ne soit apparu que vers la fin des années 1960 dans le contexte de l'arrivée de l'informatique moderne, l'origine du concept est étroitement liée à l'histoire des bibliothèques (pour plus de détail sur ce point, voir *infra* la bibliographie [14]). De fait, le catalogue d'une bibliothèque est la première réalisation pratique de la notion "métadonnée". La bibliothèque est constituée des livres, le catalogue est constitué des fiches qui décrivent les livres présents dans la bibliothèque. Chaque fiche contient, *a minima*, le titre, l'auteur, la date, le lieu de la publication, ainsi qu'un court résumé du livre. L'objectif d'une fiche est de donner au lecteur des indications nécessaires pour trouver un livre dont il a besoin sans qu'il soit nécessaire de consulter le livre lui-même. La distinction entre un livre et une fiche, ainsi que l'objectif de ladite fiche sont clairs. Dans le contexte des notions "données" et

“métadonnées”, le livre représente les données (objet à décrire) et la fiche représente les métadonnées (description de l’objet). Si on transpose ces concepts dans un monde digital, la distinction entre les données et les métadonnées doit également être aisée. De fait, un document numérique qui contient un texte ou une image représente une donnée, et une “fiche digitale” qui décrit le contenu de ce document (titre, auteur, date de production, résumé...) représente les métadonnées. *Les données sont dès lors les objets que l’on décrit et les métadonnées sont les descriptions de ces objets.* L’objectif des métadonnées est de permettre aux utilisateurs (humains) ainsi qu’aux ordinateurs et moteurs de recherche (machines) d’identifier et de trouver un document numérique en fonction des critères spécifiques attribués à ce document (titre, auteur, date, etc.) sans qu’il soit nécessaire de consulter le document lui-même.

*Différents types de métadonnées.* Si le titre, l’auteur ou la date de publication d’un document, qu’il soit numérique ou non, peuvent être utiles jusqu’à un certain point, ces renseignements ne seront pas suffisants pour quelqu’un qui cherche un contenu spécifique, par exemple un document qui ne contient que du texte en latin. Et si on ne recherche que des documents écrits à Rome? Il se pose alors la question de savoir quels renseignements doivent contenir les métadonnées. Doivent-elles être exhaustives? Faut-il y mettre toutes les informations potentiellement utiles?

Il est rarement possible de véhiculer toute la complexité et la richesse d’un objet à travers sa description. Chaque description est résolument incomplète et ne peut représenter qu’un des aspects de l’objet qu’elle décrit. Il sera, par exemple, peu judicieux d’essayer de mettre sur une seule carte de France toutes les informations que l’on dispose sur ce pays; il est plus commode de recourir à des cartes différentes pour chaque type d’information que l’on veut représenter (carte administrative, topographique, routière, etc.).

Le même principe s’applique aux métadonnées. De fait, les métadonnées peuvent être de différents types où chaque type n’exprime qu’un des aspects de l’objet que l’on décrit (pour plus de détail sur ce point, voir *infra* la bibliographie [14]). Par exemple, pour la même photo numérique un informaticien voudra savoir la taille de la photo en mégaoctets et sa résolution en DPI, un photographe souhaitera plutôt connaître l’appareil et les filtres utilisés pour l’acquisition de la photo et un historien de l’art s’intéressera avant tout

à l'histoire de la peinture représentée sur la photo. Ou bien encore, un fichier xml qui contient une charte du XII<sup>e</sup> siècle peut être accompagné des métadonnées générales (titre, auteur, date de production, langue, etc.), des métadonnées davantage centrées sur l'aspect informatique (taille de fichier, format d'encodage, standard, etc.) et des métadonnées propres à la recherche historique (sources manuscrites, bibliographie, éditions, type, état et qualité du support, etc.).

Les métadonnées de différents types offrent alors des renseignements différents sur le même objet. Ces types peuvent varier aussi bien en fonction de notre problématique que de l'objet lui-même que l'on décrit. On utilisera, par exemple, les différentes métadonnées pour décrire un texte latin du Moyen Âge et une peinture rupestre du Paléolithique supérieur. Les principes FAIR eux-mêmes incitent à introduire des métadonnées spécifiques (voir *infra* Annexe 4). Dès lors, *par un type de métadonnées on entend des renseignements sur l'objet que l'on décrit (=donnée) qui peuvent être regroupés dans un ensemble spécifique.*

De nombreux types de métadonnées ont été étudiés et ont fait l'objet d'une standardisation. Par exemple, les métadonnées que l'on peut appeler "essentiels" (titre, auteur, date de la publication, langue, etc.) ont donné lieu à un standard de métadonnées "Dublin Core" ("core" est traduit alors comme "essentiel") qui comprend 15 éléments basiques (qui peuvent, si nécessaire, être enrichis par quelque 40 "termes"). Les métadonnées plus spécifiques ont été standardisées au sein de leurs domaines respectifs, par exemple le standard "Darwin Core" pour les métadonnées biologiques ou bien encore "Metadata Object Description Schema" (MODS) pour des métadonnées bibliographiques.

Cependant, il ne sera guère envisageable de prévoir un standard pour chaque type de métadonnées. Il existe dès lors la possibilité de mettre en place ses propres standards (appelés alors "schéma") pour décrire un type de métadonnées personnelles spécifiques. Les formats sous lesquels les métadonnées peuvent être encodées sont également très variés. À l'heure actuelle, il existe plusieurs dizaines de formats et de standards différents pour encoder les métadonnées et leur choix est souvent un point important dans l'organisation des données d'un projet.

*(En théorie, rien n'empêche d'associer plusieurs fichiers de métadonnées au même fichier de données. Dans ce cas, chaque fichier de métadonnées peut s'appuyer sur des formats et des standards des métadonnées différents. Or, il est indispensable de se rappeler qu'une telle configuration nécessite que tous ces fichiers soient reliés entre eux, par exemple par un référencement interne croisé de leurs identifiants uniques. Quelle que soit la configuration mise en place, elle dépendra nécessairement des besoins et des ressources disponibles de chaque projet. Sur l'emplacement des métadonnées voir également infra "Données et Métadonnées, ensembles ou séparées?")*

*(Pour une présentation plus détaillée des différents langages, formats ou standards des métadonnées voir la bibliographie infra [6], [14]).*

Modèle des données vs Objet d'étude. Malgré le fait que la distinction entre les données et les métadonnées décrites plus haut soit d'apparence claire, dans certains contextes la séparation entre les deux peut être plus délicate. C'est le cas, par exemple, d'un acte médiéval quand son contenu (=donnée) est accompagné par des renseignements supplémentaires (=métadonnées) aussi bien sur le langage et la forme de discours que sur le support du document, l'encre et l'écriture. Toutes ces précisions peuvent être aussi une source riche d'informations pour un historien en devenant donc des "données".

Ce même processus peut être observé à l'exemple, déjà cité, des livres et des fiches. De fait, pour quelqu'un qui s'intéresse aux livres, ces derniers sont des "données" et les fiches de catalogue qui les décrivent sont des "métadonnées". En revanche, pour quelqu'un qui s'intéresse à l'histoire des catalogues des bibliothèques, ce sont les fiches elles-mêmes qui deviennent des "données" à étudier... Il est donc indispensable de se rappeler que les métadonnées des uns peuvent être les données des autres.

Pour mieux saisir ce processus d'interchangement entre les données et les métadonnées, il est nécessaire de faire la différence entre un modèle conceptuel et un modèle logique de données d'un côté et l'objet d'étude de l'autre. **Un modèle conceptuel (ou formel) de données** désigne le concept (ou objet formel) à décrire (ici, par exemple, le texte d'un acte) et tous les renseignements supplémentaires (ici, auteur de l'acte, date de l'écriture, lieu

d'écriture, support de l'écriture, etc.) qui peuvent accompagner le concept que l'on décrit. **Un modèle logique de données**, de son côté, décrit comment (entités, attributs, relations, etc.) il est prévu de mettre en pratique le modèle conceptuel. Par exemple, cela peut être un fichier (ou une table, s'il s'agit de la base de données) avec le texte de l'acte, accompagné des informations auxiliaires sur cet acte (auteur de l'acte, date de l'écriture, lieu d'écriture, support de l'écriture, etc.) écrites soit dans le même fichier (ou la même table) soit dans un fichier (ou une table) joint. Enfin, toute information, qu'elle soit auxiliaire ou non, contenue dans les modèles de données peut être un **objet d'étude** qui est défini alors par la problématique de la recherche. "L'objet d'étude" d'une recherche, dès lors, n'est pas la même chose que "l'objet formel" qui est la base d'un modèle des données. Si les modèles sont assez fixes (il est peu probable que la base de données soit reconstruite ou les fichiers re-encodés au cours d'un projet), l'objet d'étude est plus "mouvant" et peut changer au gré des questions posées.

La "métadonnée" n'est donc qu'un rôle que n'importe quelle donnée peut jouer à l'égard des autres données. De ce point de vue, la question de savoir où sont les données et où sont les métadonnées peut paraître erronée. Il est plus opportun de se demander: "Dans le contexte de ces données prises comme un objet à décrire, quelles sont les autres données qui joueront le rôle des métadonnées?". La réponse dépend inévitablement du contexte dans lequel cette question est posée (s'agit-il du modèle conceptuel ou logique ou de l'objet d'étude?) et de quel est l'auteur de cette question (administrateur, informaticien, scientifique...).

*(Pour plus de détails, voir infra la bibliographie [6], [8], [9], [10], [14]).*

Données et Métadonnées, ensemble ou séparées? Un des derniers points concerne l'emplacement des métadonnées. Deux solutions sont possibles: stocker les métadonnées dans le même endroit que les données (par exemple le même fichier ou la même table de la base de données) ou séparément. Le choix dépend de l'architecture et de l'organisation des données que vous souhaitez ou pouvez mettre en place.

Il faut toutefois garder à l'esprit que l'objectif des métadonnées est de permettre de trouver, rapidement et facilement, un objet spécifique (voir *supra* "Données vs Métadonnées"). Placer les métadonnées dans

un fichier séparé donne la possibilité - aussi bien à l'utilisateur (humain) qu'à l'ordinateur ou au moteur de recherche (machine) - de parcourir des fichiers courts et légers avec des informations essentielles (=métadonnées) au lieu de traiter des documents volumineux et complexes (=données) pour y chercher des renseignements particuliers. Enfin, il faut se rappeler que si les métadonnées sont stockées séparément des données, elles [données et métadonnées] doivent être reliées par un référencement croisé (par exemple un PID).  
(Sur ce point voir *infra* la bibliographie [6], [8], [14], cf. également Annexe 1, le principe F1 et surtout le principe F3)

## **Jeu de données**

Les données auxquelles les principes FAIR sont appliqués peuvent être désignées comme *Objet Digital FAIR* ou FDO (FAIR Digital Object) [7], [15], [17]. Cela signifie que dans le cadre de la FAIRisation des données, l'identifiant pérenne et les métadonnées sont associés à un FDO [1].

Les données scientifiques, notamment en histoire, se présentent le plus souvent sous la forme d'un corpus constitué de plusieurs documents. Il se pose alors la question de savoir si le FDO correspond à un corpus entier ou à un seul document. Les recommandations FAIR sont muettes à ce sujet. La bibliographie précise seulement que le FDO doit être "*une entité significative*" pour la FAIRisation [1], [7], [16]. À la fin, le choix dépend de l'écosystème FAIR choisi, soit une infrastructure personnelle soit un dépôt de données (sur les différents types d'écosystème FAIR voir *infra* "*Écosystème FAIR*").

Si c'est l'infrastructure personnelle qui a été retenue, il appartient à chaque projet de choisir le niveau de la graduation préférable. Il est possible de proposer l'accès aux données aussi bien document par document qu'au corpus entier (en tant qu'un fichier unique) ou bien les deux.

En revanche, dans le cas du choix d'un dépôt institutionnel comme une infrastructure principale de la FAIRisation, il semble plus judicieux d'opter pour le jeu de toutes les données d'un corpus et non pour une seule valeur atomique (c'est à dire notice, acte, charte, lettre, etc.) comme un FDO. Ce choix est dicté aussi bien par les pratiques déjà existantes (les dépôts Zenodo et Harvard Dataverse, par exemple, ne semblent héberger que les "*dataset* / jeu de données") que par une

plus grande simplicité de cette approche (déposer un seul fichier du corpus plutôt que plusieurs centaines de documents un par un).

Quel que soit le procédé utilisé, il est important de préciser ce qui est entendu par le “jeu de données” dans le contexte de la FAIRisation. Un “jeu de données” désigne l'ensemble de toutes les données du corpus réuni dans un seul fichier (ou archive numérique, par exemple sous format zip). Il est important de se rappeler qu'il peut exister plusieurs jeux de données où chaque “jeu” réunit des données analogues, par exemple les images, les textes, les notes, etc. Par la suite, ces différents ensembles pourront être reliés grâce aux référencements croisés des identifiants uniques dans les métadonnées (principe FAIR I3).

(Pour plus d'informations sur le FDO voir infra la bibliographie, notamment [7], [15]).

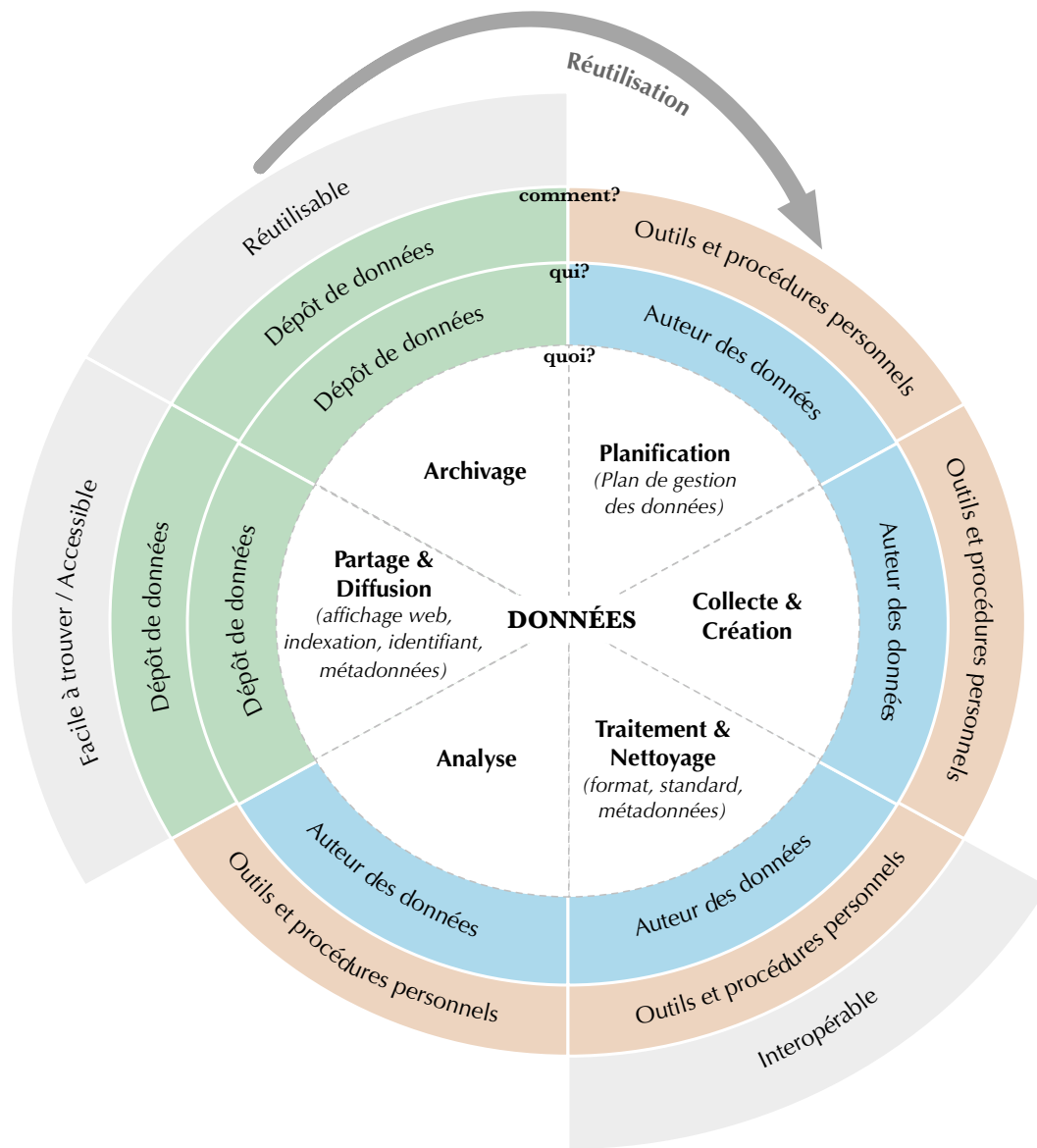
## Écosystème FAIR

La mise en place des principes FAIR passe par plusieurs procédures et peut s'appuyer sur des outils différents. On parle alors d'un écosystème FAIR qui comprend des acteurs, services et concepts très variés (identifiant pérenne, FAIR Digital Object, Plan de Gestion des Données, dépôt des données, etc.). Bien qu'il revienne à chaque projet de choisir les moyens par lesquels il souhaite atteindre les différents principes FAIR, il est possible d'envisager deux types majeurs d'écosystème.

Le premier type s'appuie sur un dépôt de données institutionnel<sup>1</sup> (**Figure 1**). Dans le cas où il est préférable de se concentrer sur la récolte et l'analyse des données, il est judicieux de confier toutes les autres opérations, comme le stockage, l'archivage, la diffusion et, même, l'affichage des données brutes, à un dépôt déjà existant.

---

<sup>1</sup> Le terme “repository”, utilisé en anglais pour désigner un “dépôt” de données, peut, parfois, être traduit en français comme “référentiel”. Par exemple: [https://fr.wikipedia.org/wiki/D%C3%A9p%C3%BAt\\_\(informatique\)](https://fr.wikipedia.org/wiki/D%C3%A9p%C3%BAt_(informatique)). Cependant un “référentiel” est une notion polysémique qui désigne, dans la langue française, avant tout un système de référence plutôt qu'un dépôt. L'utilisation de ce vocable peut donc prêter à confusion et il est indispensable d'y prendre garde.

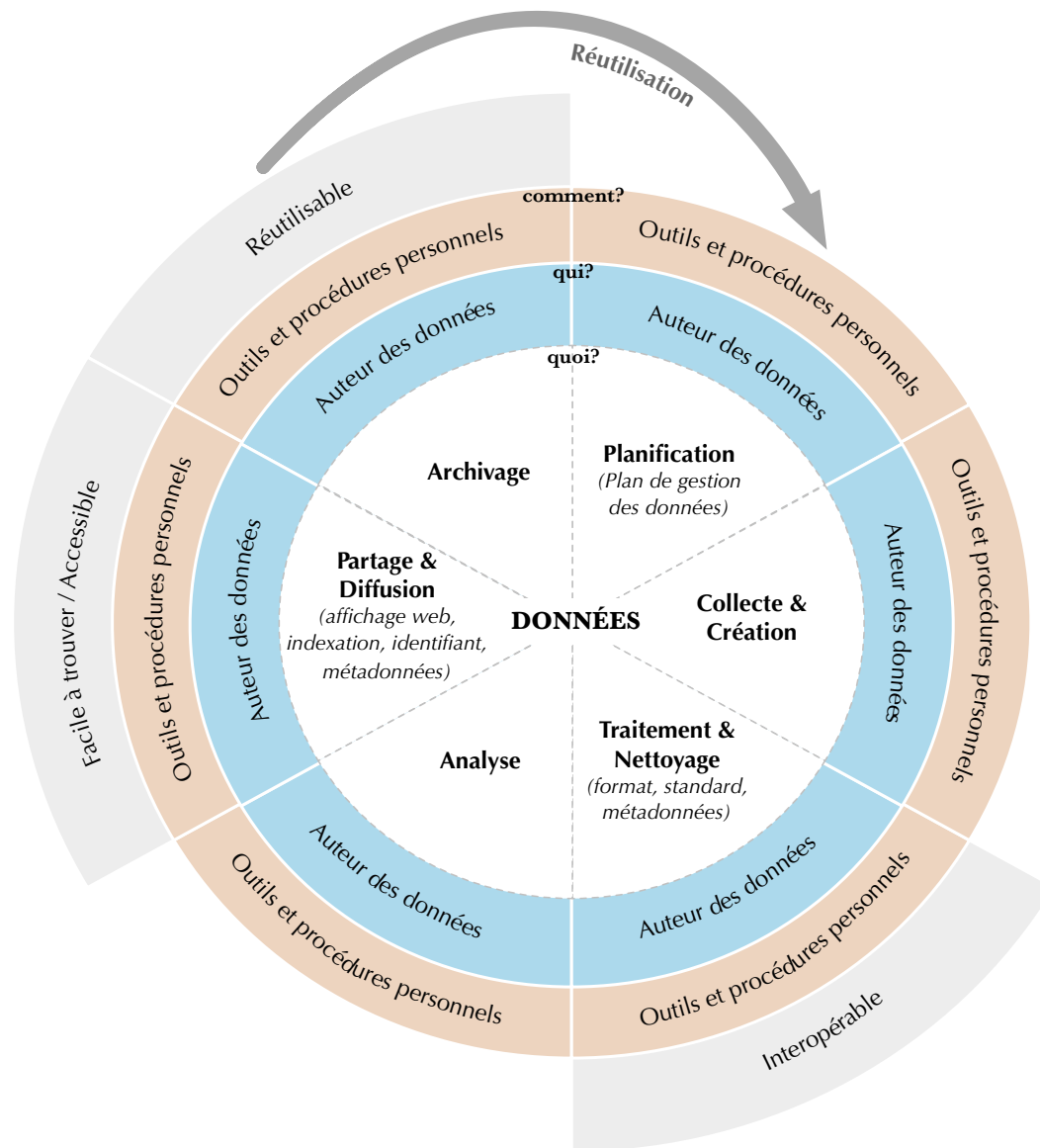


**Figure 1. Principes FAIR et écosystème basé sur un dépôt institutionnel dans le contexte du cycle de vie des données.**

De fait, il est indispensable de se rappeler que le dépôt de données ne remplit pas uniquement la fonction de l'archivage des données. La majorité des dépôts institutionnels s'occupe de tâches très diverses comme l'attribution des identifiants, la gestion des métadonnées, l'indexation et la diffusion des données auprès des moteurs de recherche, le versioning, la mise en place de l'accès aux données, etc. (À noter toutefois que d'ordinaire c'est un jeu de toutes les



données d'un corpus qui est versé dans un dépôt. Sur ce point, voir *supra* "Jeu de donnée"). Dans ce type d'écosystème, l'auteur des données ne s'occupe que de l'encodage des données dans un format et avec un standard connus et répandus (principes FAIR 11 et 12). (Le programme de la FAIRisation proposé ci-dessus s'appuie sur cette approche, voir *infra* "2 Étapes de la FAIRisation").



**Figure 2. Principes FAIR et écosystème basé sur des outils personnels dans le contexte du cycle de vie des données.**

Le deuxième type d'écosystème s'articule principalement autour d'une infrastructure personnelle (**Figure 2**). Cette approche est judicieuse s'il est nécessaire d'avoir la main mise sur chaque étape du cycle de vie des données, de leur récolte à leur diffusion. L'infrastructure personnelle peut se matérialiser par la création de ses propres outils de gestion des métadonnées (par exemple à l'aide de *FAIR Data Point*, voir [9]), des sites web ou des plateformes personnels pour l'affichage des données, des serveurs locaux pour leur stockage, etc. (Voir par exemple l'infrastructure FAIR mis en place par le consortium MASA: <https://masa.hypotheses.org/files/2021/02/MASAecosystem2020-scaled.jpg>). Dans le cadre de cette deuxième approche, le recours à un dépôt de données peut également trouver sa place en tant qu'un des éléments de l'infrastructure (par exemple uniquement pour le partage et l'exposition des données).

Bien que, dans une grande partie des cas, l'écosystème, qui s'appuie grandement sur l'utilisation de dépôt de données, soit le moyen le plus simple et le plus rapide pour satisfaire les principes FAIR, il revient à chaque projet de choisir l'infrastructure qu'il estime le plus adaptée à ses besoins et à ses ressources. À la fin, quelle que soit l'infrastructure choisie, il est indispensable de s'assurer qu'elle réponde à tous les principes FAIR attendus (voir *infra* Annexe 2 et Annexe 3).

*(Pour plus de détails sur la notion de l'écosystème FAIR voir la bibliographie [10], [12], [16], [17], [18]).*

## 2 Étapes de la FAIRisation

Les étapes suivantes décrivent le processus complet de la FAIRisation qui peut être appliqué à un jeu de données d'un corpus afin que ce dernier réponde à tous les critères FAIR. Ce programme de la FAIRisation fait appel à un écosystème FAIR basé sur un dépôt des données institutionnel (sur les différents écosystèmes FAIR, voir *supra* "1.5 Des données au dépôt de données: quelques notions principales / Écosystème FAIR").

### 1. Vérification des données

1.1. S'assurer que les langages et les formats/extensions utilisés pour l'encodage des données sont connus, largement utilisés et, de préférence, libres. (Il est impossible de définir in fine ce qui est un format/langage "connu" ou "largement utilisé". On s'appuie donc sur le bon sens... Privilégiez toujours les formats libres, par exemple csv (format libre) au lieu de xls (format propriétaire de Microsoft), txt (format libre) au lieu de doc (format propriétaire de Microsoft), etc. Pour plus d'informations sur le choix des formats à privilégier pour la pérennité des données de la recherche voir ce recensement: <https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>). Si cela n'est pas le cas, expliquer clairement ce qui a dicté le choix d'un langage ou d'un format particulier et indiquer tous les logiciels (nom, version, système d'exploitation) et outils nécessaires pour la lecture des données. (Ces informations seront par la suite utilisées dans les métadonnées). **FAIR: I1**

1.2. S'assurer que les données sont explicites et compréhensibles en l'état. Cela signifie que toutes les abréviations, tous les termes ambigus ou toutes les balises utilisées dans la description des données doivent être clairement expliqués ou avoir déjà une documentation existante. (Pour plus d'informations sur les différents standards voir ce recensement: <http://rd-alliance.github.io/metadata-directory/standards/>).

Trois options sont possibles: i) les données sont explicites en l'état; ii) les données ne sont pas explicites et sont accompagnées d'une documentation détaillée personnelle; iii) les données ne sont pas explicites et sont accompagnées d'un renvoi vers une documentation détaillée déjà existante (par exemple TEI P5 TEI Guidelines). (Ces informations seront par la suite utilisées dans les métadonnées). **FAIR I2**

## 2. Préparation des métadonnées

### 2.1. Préparation des métadonnées complètes.

2.1.1. S'assurer que les métadonnées contiennent le plus de détails possible. (Se référer au modèle des métadonnées détaillées, voir Annexe 4) **FAIR: F2**

2.1.2. S'assurer que les données disposent une licence et qu'elle est clairement mentionnée dans les métadonnées. (Les données peuvent avoir tous types de licence, y compris les licences "non libres de droits"). **FAIR: R1.1**

2.1.3. S'assurer que les métadonnées contiennent un historique détaillé de la provenance des données. (Par exemple, les données peuvent être produites durant un autre projet et ensuite réutilisées, corrigées et modifiées par le présent projet). **FAIR: R1.2**

2.1.4. S'assurer que les données utilisent des standards (s'il en existe) de présentation, d'organisation ou d'archivage communément utilisés dans le domaine scientifique auquel ces données appartiennent. (Ces informations seront par la suite utilisées dans les métadonnées).

(Bien qu'un "standard" commun ne semble pas exister en sciences humaines et sociales, certains projets ou programmes de recherche peuvent s'appuyer sur des conventions similaires. Par exemple, l'organisation des données dans le même type de tableau, l'utilisation des mêmes principes de balisage, etc. Si c'est le cas, ces conventions doivent être mentionnées, par exemple dans les métadonnées. Ne pas confondre avec I2. Le I2 vise à rendre les données compréhensibles / interopérables grâce à la documentation associée

à ces données. Le R1.3 vise à faciliter la réutilisation des données grâce aux indications générales sur les standards et les conventions sur lesquels les données s'appuient. Pour plus d'informations sur les différents standards voir ce recensement: <http://rd-alliance.github.io/metadata-directory/standards/>) **FAIR: R1.3**

## 2.2. Relier les données.

S'assurer que dans les cas où les présentes données peuvent être reliées à d'autres données (par exemple, les données textuelles peuvent être reliées aux images des manuscrits), les métadonnées contiennent des références à ces données reliées (par exemple le PID des données reliées, voir F1). Le type de lien entre les données doit être explicitement mentionné. (Par exemple, "ce texte est la transcription de ce manuscrit" ou "ce texte a été écrit par cette personne"). **FAIR: I3.**

## 3. Mise des données dans un dépôt

### 3.1. Enregistrement des données et des métadonnées.

3.1.1. Choisir une ressource consultable (par exemple, un dépôt de données) qui répond à tous les critères FAIR applicables aux métadonnées (F1, F3, F4, A1.1, A1.2, A2, I1, I2, I3, R1.1, R1.2, R1.3). La plupart des grands dépôts de données affichent de façon claire leur conformité aux critères FAIR. (Pour plus de détails sur le choix de dépôt et sur les critères FAIR auxquels il doit répondre voir infra Annexe 3)

3.1.2. Enregistrer les données et les métadonnées y associées dans une ressource consultable choisie précédemment. **FAIR: F4**

### 3.2. Vérification de l'accessibilité et de l'exactitude des (méta)données.

3.2.1. Vérifier que les données possèdent un identifiant persistant et unique (PID) (fourni par la ressource où les données sont enregistrées). **FAIR: F1**

3.2.2. Vérifier que le PID des données est mentionné dans les métadonnées. **FAIR: F3**

3.2.3. Vérifier que les données sont accessibles via un protocole de communication ouvert, gratuit et largement utilisé. Si les données sont enregistrées dans une ressource qui est accessible via un site internet, ce critère est automatiquement rempli. *(Par un protocole de communication "ouvert, gratuit et largement utilisé", on entend avant tout un protocole informatique libre et connu comme HTTP/HTTPS (=site web), FTP etc. Les autres "protocoles de communications" qui s'appuient sur des logiciels propriétaires et non universellement implantés (par exemple Skype, Microsoft Exchange Server, etc.) ne répondent pas à ce critère).* **FAIR: A1.1**

3.2.4. Vérifier que les données sont accessibles via un protocole de communication qui permet, si nécessaire, de mettre en place une procédure d'authentification et d'autorisation. *(FAIR data ne signifie pas OpenData. Le protocole via lequel les données et les métadonnées sont accessibles, doit avoir la possibilité, en cas de besoin, de restreindre l'accès aux données, par exemple uniquement aux utilisateurs possédant un login et un mot de passe).* **FAIR: A1.2**

3.2.5. Vérifier que les métadonnées seront accessibles même lorsque les données elles-mêmes ne seront plus disponibles. *(De façon ordinaire ce critère est considéré comme rempli, si la condition F4 est satisfaite).* **FAIR: A2**

## Bibliographie

- [1] *A Persistent Identifier (PID) policy for the European Open Science Cloud*, European Commission, Directorate-General for Research and Innovation, 2020. doi: 10.2777/926037
- [2] Bloemers M, Montesanti A. "The FAIR funding model: Providing a framework for research funders to drive the transition toward FAIR data management and stewardship practices", *Data Intelligence* 2(2020), 171–180. doi: 10.1162/dint\_a\_00039.
- [3] David R, *et al.* "FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles", *Data Science Journal*. 2020;19(1):32. doi: 10.5334/dsj-2020-032
- [4] *FAIR Data Maturity Model: specification and guidelines*, RDA FAIR Data Maturity Model Working Group, 2020. doi: 10.15497/rda00045
- [5] *Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0, 2016. En ligne: [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- [6] *Handbook of Metadata, Semantics and Ontologies*, éd. Sicilia M.-A., World Scientific Publishing, 2014
- [7] Harjes J. *et al.* "FAIR digital objects in environmental and life sciences should comprise workflow operation design data and method information for repeatability of study setups and reproducibility of results", *Database: the journal of biological databases and curation* (2020). doi: 10.1093/database/baaa059
- [8] Hay D. C., *Data model patterns: a metadata map*, Elsevier, 2006
- [9] Jacobsen A. *et al.* "A Generic Workflow for the Data FAIRification Process." *Data Intelligence* 2 (2020): 56-65. doi: 10.1162/dint\_a\_00028
- [10] Jacobsen A. *et al.* "FAIR principles: Interpretations and implementation considerations", *Data Intelligence*. 2020;2 (1-2) :10-29. doi: 10.1162/dint\_r\_00024

- [11] Landi A. et al. "The "A" of FAIR – as open as possible, as closed as necessary", *Data Intelligence* 2(2020), 47–55. doi: 10.1162/dint\_a\_00027.
- [12] de Miranda Azevedo R., Dumontier M., "Considerations for the Conduction and Interpretation of FAIRness Evaluations", *Data Intelligence*. 2020 Jan 1;2(1-2):285-292. doi: 10.1162/dint\_a\_00051
- [13] Pergl R., et al. 2019. "Data Stewardship Wizard": A Tool Bringing Together Researchers, Data Stewards, and Data Experts around Data Management Planning", *Data Science Journal*, 18: 59, pp. 1–8. doi: 10.5334/dsj-2019-059
- [14] Pomerantz J., *Metadata*, MIT Press, 2015
- [15] Schwardmann U., "Digital Objects – FAIR Digital Objects: Which Services Are Required?", *Data Science Journal*, 19(1) 2020. doi: 10.5334/dsj-2020-015
- [16] Thompson M., et al. "Making FAIR Easy with FAIR Tools: From Creolization to Convergence." *Data Intelligence* 2 (2020): 87-95. doi: 10.1162/dint\_a\_00031
- [17] *Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data*. Luxembourg: Publications Office of the European Union, 2018 doi: 10.2777/1524
- [18] Wilkinson M., et al. "The FAIR Guiding Principles for scientific data management and stewardship", *Science Data* 3, 160018 (2016). doi: 10.1038/sdata.2016.18
- [19] Wilkinson M., et al. "A design framework and exemplar metrics for FAIRness", *Science Data* 5, (2018), 1–4. doi: 10.1038/sdata.2018.118
- [20] Wise J., et al. "Implementation and relevance of FAIR data principles in biopharmaceutical R&D", *Drug Discovery Today*, Volume 24/4 April 2019. doi: 10.1016/j.drudis.2019.01.008



# Annexe 1

## Principes FAIR

D'après "GO FAIR: FAIR Principles". En ligne: <https://www.go-fair.org/fair-principles>. Le nombre total des principes est de 15.

**Rappel:** Tous les principes (sauf F2, F3, A2) s'appliquent aussi bien aux données qu'aux métadonnées; cela explique l'utilisation du vocable "(méta)données" (pour les principes FAIR étendus voir *infra* Annexe 2). La partie "données" est garantie par l'auteur des données, la partie "métadonnées" est garantie par le dépôt de données (voir *supra* "1.4 Qui FAIRise quoi?" et "1.5 De données au dépôt de données: quelques notions principales / Écosystème FAIR", voir également *infra* Annexe 2 et Annexe 3).

### **Findable (= Facile à trouver)**

- F1.** Les (méta)données possèdent un identifiant persistant et unique au monde (*persistent identifier*, PID).
- F2.** Les données sont décrites avec des métadonnées riches.
- F3.** Les métadonnées incluent, d'une façon claire et explicite, l'identifiant des données qu'elles décrivent.
- F4.** Les (méta)données sont enregistrées ou indexées dans une ressource consultable.

### **Accessible**

- A1.** Les (méta)données sont récupérables par leur identifiant en utilisant un protocole standard de communication.
  - A1.1.** Le protocole est ouvert, gratuit et largement utilisé.
  - A1.2.** Le protocole permet, si nécessaire, une procédure d'authentification et d'autorisation.
- A2.** Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles.

**Interoperable (= Interopérable)**

- I1.** Les (méta)données utilisent un langage formel, accessible, partagé et largement répandu pour la représentation des connaissances.
- I2.** Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR.
- I3.** Les (méta)données incluent des références qualifiées à d'autres (méta)données.

**Reusable (= Réutilisable)**

- R1.** Les méta(données) sont décrites d'une façon complète avec plusieurs attributs précis et pertinents.
  - R1.1.** Les (méta)données sont publiées avec une licence d'utilisation des données claire et accessible.
  - R1.2.** Les (méta)données disposent d'une provenance détaillée.
  - R1.3.** Les (méta)données répondent aux standards communautaires de leur domaine.

## Annexe 2

### Principes FAIR étendus et les différents acteurs

**Rappel:** Les principes FAIR étendus désignent chaque principe FAIR doublé dans le cas s'il s'applique aussi bien aux données qu'aux métadonnées. Le nombre total des principes FAIR est de 15 (voir *supra* Annexe 1). Le nombre des principes FAIR étendus est de 27.

Chaque principe FAIR peut être accompli par des acteurs différents (*sur les différents écosystèmes FAIR voir supra "1.5 Des données au dépôt de données: quelques notions principales / Écosystème FAIR"*). Voir, par exemple, comment les différents dépôts remplissent les principes FAIR: le dépôt Zenodo <https://about.zenodo.org/principles/>, le dépôt Dataverse <https://dataverse.org/software-features>.

Principes FAIR étendus	Auteur des données	Dépôt de données ou Infrastructure personnelle
<b>F1.</b> Les <b>données</b> possèdent un identifiant persistant et unique au monde (persistant identifier, PID).		X
<b>F1.</b> Les <b>métadonnées</b> possèdent un identifiant persistant et unique au monde (persistant identifier, PID).		X
<b>F2.</b> Les <b>données</b> sont décrites avec des métadonnées riches.	X (remplir le formulaire des métadonnées)	X (fournir le formulaire des métadonnées)
<b>F3.</b> Les <b>métadonnées</b> incluent, d'une façon claire et explicite, l'identifiant des données qu'elles décrivent.		X
<b>F4.</b> Les <b>données</b> sont enregistrées ou indexées dans une ressource consultable.		X
<b>F4.</b> Les <b>métadonnées</b> sont enregistrées ou indexées dans une ressource consultable.		X

<b>A1.</b> Les <b>données</b> sont récupérables par leur identifiant en utilisant un protocole standard de communication.		<b>X</b>
<b>A1.</b> Les <b>métadonnées</b> sont récupérables par leur identifiant en utilisant un protocole standard de communication.		<b>X</b>
<b>A1.1.</b> Le protocole est ouvert, gratuit et largement utilisé. <b>[pour les données]</b>		<b>X</b>
<b>A1.1.</b> Le protocole est ouvert, gratuit et largement utilisé. <b>[pour les métadonnées]</b>		<b>X</b>
<b>A1.2.</b> Le protocole permet, si nécessaire, une procédure d'authentification et d'autorisation. <b>[pour les données]</b>		<b>X</b>
<b>A1.2.</b> Le protocole permet, si nécessaire, une procédure d'authentification et d'autorisation. <b>[pour les métadonnées]</b>		<b>X</b>
<b>A2.</b> Les <b>métadonnées</b> sont accessibles, même lorsque les données ne sont plus disponibles.		<b>X</b>
<b>I1.</b> Les <b>données</b> utilisent un langage formel, accessible, partagé et largement répandu pour la représentation des connaissances.	<b>X</b>	
<b>I1.</b> Les <b>métadonnées</b> utilisent un langage formel, accessible, partagé et largement répandu pour la représentation des connaissances.		<b>X</b>
<b>I2.</b> Les <b>données</b> utilisent des vocabulaires qui suivent les principes FAIR.	<b>X</b>	
<b>I2.</b> Les <b>métadonnées</b> utilisent des vocabulaires qui suivent les principes FAIR.		<b>X</b>
<b>I3.</b> Les <b>données</b> incluent des références qualifiées à d'autres <b>données</b> .	<b>X</b> (remplir le champ approprié des métadonnées)	<b>X</b> (fournir le champ approprié des métadonnées)
<b>I3.</b> Les <b>métadonnées</b> incluent des références qualifiées à d'autres <b>métadonnées</b> .	<b>X</b> (remplir le champ approprié des métadonnées)	<b>X</b> (fournir le champ approprié des métadonnées)

<b>R1.</b> Les <b>données</b> sont décrites d'une façon complète avec plusieurs attributs précis et pertinents.	<b>X</b> (remplir le formulaire des métadonnées)	<b>X</b> (fournir le formulaire des métadonnées)
<b>R1.</b> Les <b>métadonnées</b> sont décrites d'une façon complète avec plusieurs attributs précis et pertinents.		<b>X</b>
<b>R1.1.</b> Les <b>données</b> sont publiées avec une licence d'utilisation des données claire et accessible.	<b>X</b> (remplir le champ approprié des métadonnées)	<b>X</b> (fournir le champ approprié des métadonnées)
<b>R1.1.</b> Les <b>métadonnées</b> sont publiées avec une licence d'utilisation des données claire et accessible.		<b>X</b>
<b>R1.2.</b> Les <b>données</b> disposent d'une provenance détaillée.	<b>X</b> (remplir le champ approprié des métadonnées)	<b>X</b> (fournir le champ approprié des métadonnées)
<b>R1.2.</b> Les <b>métadonnées</b> disposent d'une provenance détaillée.		<b>X</b>
<b>R1.3.</b> Les <b>données</b> répondent aux standards communautaires de leur domaine.	<b>X</b> (remplir le champ approprié des métadonnées)	<b>X</b> (fournir le champ approprié des métadonnées)
<b>R1.3.</b> Les <b>métadonnées</b> répondent aux standards communautaires de leur domaine.		<b>X</b>

## Annexe 3

### Choix d'un dépôt de données

**Rappel:** Un dépôt de données doit garantir tous les principes FAIR applicables aux métadonnées (c'est-à-dire, tous les principes sauf F2, voir *supra* "1.4 Qui FAIRise quoi?" et Annexe 1).

#### Comment choisir un dépôt de données?

- choisir le dépôt le plus utilisé dans votre discipline;
- ou sinon, choisir le dépôt le plus utilisé dans votre institution;
- ou sinon, choisir un dépôt "généraliste".

#### Où chercher un dépôt de données pour votre discipline?

- *Registry of Research Data Repositories*: <https://www.re3data.org>
- *OpenDOAR*: <https://v2.sherpa.ac.uk/opensdoar/search.html>
- *DATAACC. Déposer ses données en ligne: où et comment*: <https://www.dataacc.org/vos-besoins/valoriser-ses-donnees/deposer-ses-donnees-en-ligne-ou-et-comment>

#### Les dépôts des données "généralistes" les plus utilisés:

- *Zenodo (CERN Data Centre / EU Horizon 2020 / OpenAIRE)*: <https://zenodo.org>
- *Harvard Dataverse*: <https://dataverse.harvard.edu>
- *Figshare*: <https://figshare.com>
- *Mendeley Data*: <https://data.mendeley.com>
- *Open Science Framework (OSF)*: <https://osf.io>
- *Dryad*: <https://datadryad.org/stash>

## Annexe 4

### Modèle (prévisionnel) des métadonnées

Rubriques principales	Correspondance éléments <i>Dublin Core</i>
Identifiant (PID):	<i>DC.Identifier</i>
Intitulé (des données):	<i>DC.Title</i>
Auteur(s):	<i>DC.Creator</i>
Contributeur(s):	<i>DC.Contributeur</i>
Éditeur: (par exemple équipe ou institution)	<i>DC.Publisher</i>
Financement:	
Date de la publication (des données):	<i>DC.Date</i>
Dernière modification (des données):	
Version (des données):	
Propriétaire (des données):	
Langue(s):	<i>DC.Language</i>
Description (des données):	<i>DC.Description</i>
Mots-clés:	<i>DC.Subject</i>
Rubriques complémentaires (selon principes FAIR)	
Format de fichier (FAIR: I1): (voir: <a href="https://en.wikipedia.org/wiki/List_of_file_formats">https://en.wikipedia.org/wiki/List_of_file_formats</a> )	<i>DC.Format</i>
Logiciels nécessaires pour la lecture des données: (nom, version, système d'exploitation) (FAIR: I1):	
Si format ou logiciel particulier, expliquer clairement pourquoi ils ont été choisis (FAIR: I1):	

Si les données ne sont pas explicites en l'état, renvoyer vers les explications/documentation/vocabulaire pour expliciter les données (FAIR: I2):	
Licence des données (FAIR: R1.1):	<i>DC.Rights</i>
Provenance et historique des données (FAIR R.1.2):	
Standards, conventions des données utilisées (FAIR: R1.3):	
Données liées (FAIR: I3): (mentionner, le cas échéant, le PID, l'intitulé des données liées et le type de lien. Par exemple "ces textes (PID, intitulé du jeu de données avec les textes) <u>est la transcription</u> de ces photos de manuscrits (PID, intitulé du jeu de données avec les images) "	<i>DC.Relation</i>