

# FAIR par la pratique

## FAIRiser quoi?

Avant tout, il faut savoir quelles sont les données que l'on souhaite et que l'on peut FAIRiser. De fait, une "unité de données" à laquelle les principes FAIR sont appliqués est communément désignée comme un "Objet Digital FAIR" (FAIR Digital Object, FDO). C'est à ce FDO que les identifiants pérennes et les métadonnées sont associés. Mais, en pratique, comment choisir ce FDO?

Les principes FAIR indiquent seulement que le FDO doit être "une entité significative" pour la FAIRisation. Il appartient donc à chaque projet de faire le choix de "l'unité" de la FAIRisation. Dans les faits, **il est possible de FAIRiser soit un "jeu de données" (dataset) qui regroupe alors toutes les données d'un corpus, soit une "unité atomique" d'un corpus** (par exemple, une notice, un document, une image...). Il faut, toutefois, garder à l'esprit que la FAIRisation des unités atomiques, une par une, signifiera de déposer dans un dépôt des données (si cette approche est choisie) des centaines/milliers de fichiers, un par un. Et cela ne semble pas être raisonnable à l'échelle d'un corpus.

À noter également que des corpus constitués de données d'une nature différente (par exemple, données textuelles, visuelles, chiffrées, etc.) peuvent regrouper chaque type de données dans un jeu de données séparé. Autrement dit, **au sein d'un même corpus il peut exister plusieurs jeux de données où chaque "jeu" réunit des données analogues, par exemple les images, les textes, les notes, etc.**

## FAIRiser comment?

Bien qu'il revienne à chaque projet de choisir les moyens par lesquels il souhaite atteindre les différents principes FAIR, **il est généralement possible d'envisager deux approches: la mise des données dans un dépôt de données institutionnel ou l'élaboration d'une infrastructure personnelle de gestion des données.** Si, pour une raison ou une autre, vous voulez garder la main sur la totalité du processus et si vous avez la possibilité de faire accomplir les tâches techniques lourdes par du personnel compétent vous pouvez opter pour l'infrastructure personnelle, dans tous les autres cas, il vaut mieux se tourner vers un dépôt des données.

Pour plus de détails sur les différentes approches de la FAIRisation, voir: [Écosystème FAIR](#)

Exemple comment un dépôt peut remplir les critères FAIR: [Zenodo: FAIR Principles](#)

Exemple de modèle des données pour un dépôt institutionnel: [Dataset upload example](#)

Exemple d'un dépôt réalisé: [CartulR. Répertoire des cartulaires médiévaux et modernes](#)

Pour plus de détails sur le programme de la FAIRisation du consortium COSME<sup>2</sup>, voir: [COSME<sup>2</sup>{FAIR}](#)

# Éléments-clés des principes FAIR

Les recommandations FAIR sont composées de 15 principes (ou 27 dans la version étendue, voir: [Principes FAIR étendus](#)) qui doivent être satisfaits afin que les données soient considérées comme "FAIRisées". Il est cependant possible de résumer l'ensemble de ces principes en quelques éléments-clés.

## Identifiant pérenne

### Principes FAIR concernés:

F1, F3, A1

### Points à retenir:

☞ Il existe différents systèmes d'identifiants pérennes (*Persistent Identifier, PID*). Parmi les plus connus, on peut citer DOI ([Digital Object Identifier](#)), ARK ([Archival Resource Key](#)) ou bien encore VIAF ([Virtual International Authority File](#)). Par exemple, un DOI se présente comme "10.1000/182", un ARK comme "ark:/53355/cl010066723" et un VIAF comme "106965171". Il est possible de choisir un type d'identifiant qui convient le mieux à vos données. À noter toutefois qu'un DOI reste l'identifiant générique le plus utilisé et le mieux connu et qu'il peut être aisément associé à une plus grande variété de données numériques.

☞ Une URL (*Uniform Resource Locator*, communément appelée "adresse web") peut, à certains égards, être considérée comme un PID, mais son utilisation doit être déconseillée à cause de son caractère instable. En pratique: Si vos données sont, par exemple, stockées et affichées par le biais de votre propre site internet, l'adresse web (URL) qui permet d'accéder à la page où vos données sont affichées peut, à certains égards, être considérée comme un identifiant pérenne pour les données en question. Or, cette approche est contre-indiquée car la nature même d'un site internet ne permet pas de garantir la persistance d'une adresse web à moyen et, surtout, à long terme.

☞ Toutes les données que vous souhaitez FAIRiser doivent posséder un identifiant pérenne. Si vous optez pour la FAIRisation de votre corpus en tant que seul jeu de données c'est *ce jeu de données* qui doit avoir un PID. Si vous choisissez de FAIRiser les notices de votre corpus une par une alors c'est *chaque notice* qui doit détenir son propre PID.

### Comment procéder?

**dépôt de données:** Si vous entreposez les données dans un dépôt de données, le dépôt se charge automatiquement de la création et de l'attribution d'un PID pour vos données.

**infrastructure personnelle:** Si c'est l'infrastructure personnelle qui a été choisie pour la FAIRisation, il vous revient de créer et d'attribuer par vous-même un PID pour vos données. (Pour la création manuelle d'un identifiant pérenne, vous pouvez vous appuyer, par exemple, sur les services de [OPIDoR-CNRS](#)).

## Métadonnées

### Principes FAIR concernés:

F1, F2, F3, F4, A1, I1, I2

### Points à retenir:

- ☞ Toutes les données que vous souhaitez FAIRiser doivent être accompagnées par des métadonnées.
- ☞ Dans la plupart des cas, les métadonnées sont stockées dans un fichier spécifique. Ce fichier doit être encodé sous un format et selon les standards recommandés pour la représentation des métadonnées. Parmi ces formats et standards, on retient avant tout les recommandations RDF et les formats RDF/XML et JSON-LD.
- ☞ Chaque fichier de métadonnées doit contenir l'identifiant pérenne (PID) des données auxquelles elles [métadonnées] sont associées.
- ☞ Les métadonnées doivent contenir des informations suffisamment détaillées (auteurs, intitulé, date de création, description, mots-clés, etc.) sur les données qu'elles accompagnent. Il est possible, par exemple, de s'appuyer sur les éléments du vocabulaire *Dublin Core* pour la construction des métadonnées.

### Comment procéder?

**dépôt de données:** Si vous entreposez les données dans un dépôt de données, le dépôt se charge automatiquement de la création et de la conservation des fichiers des métadonnées avec un format et selon les standards en vigueur. Il ne vous reste qu'à remplir les champs (auteurs de données, contenu du dépôt, etc.) fournis par le dépôt lors de l'ajout de vos données.

**infrastructure personnelle:** Si c'est l'infrastructure personnelle qui a été choisie pour la FAIRisation, il vous revient de vous assurer que vos métadonnées contiennent tous les éléments nécessaires, qu'elles sont conservées par vos soins sous un format et selon les standards recommandés et qu'elles [métadonnées] sont reliées (par le biais d'un PID) à vos données.

## Stockage et affichage des données et des métadonnées

### Principes FAIR concernés:

A1, A2

### Points à retenir:

☞ Les données et les métadonnées doivent être accessibles par un moyen répandu et connu du plus grand nombre. En pratique, on s'attend, dans la majorité des cas, que les données et les métadonnées soient accessibles sur Internet par l'intermédiaire d'un site web.

### Comment procéder?

**dépôt de données:** Si vous entreposez les données dans un dépôt de données, le dépôt met automatiquement à votre disposition un espace de stockage (un serveur) et un espace d'affichage (une page web) qui contient toutes les informations nécessaires pour accéder à vos données et métadonnées.

**infrastructure personnelle:** Si c'est l'infrastructure personnelle qui a été choisie pour la FAIRisation, il vous revient de créer et de maintenir un espace de stockage et d'affichage pour vos données et métadonnées.

## Format et langage des données

### Principes FAIR concernés:

I1

### Points à retenir:

☞ Les données doivent être encodées sous un format et avec un langage connu et largement répandu. En pratique, c'est le critère le plus facile à satisfaire; il suffit d'éviter de représenter les données sous un format qui n'a été utilisé que par trois personnes dans le monde au milieu des années 80 du siècle passé...

### Comment procéder?

**auteur des données:** Quelle que soit l'approche que vous choisirez pour la FAIRisation de vos données (par l'intermédiaire d'un dépôt institutionnel ou par le biais d'une infrastructure personnelle), c'est à l'auteur de données qu'il appartient de s'assurer que ces données soient encodées sous un format et avec un langage appropriés.

## Documentation des données

### Principes FAIR concernés:

I2, R1.3

### Points à retenir:

- ☞ Toutes les données doivent être explicites et compréhensibles en l'état. En pratique, cela signifie que le modèle et la structure des données, les abréviations, les vocabulaires et les conventions utilisés, les éventuels écueils et les difficultés du traitement des données, tous ces éléments doivent être décrits et annotés. En résumé, cette documentation doit être suffisamment complète pour permettre à une personne tierce de comprendre et d'analyser les données présentes.
- ☞ Si vous vous appuyez sur les modèles des données déjà décrits et documentés, par exemple les recommandations TEI P5 pour le langage XML, il suffit, dans la plupart des cas, de renvoyer vers cette documentation existante.

### Comment procéder?

**auteur des données:** Quelle que soit l'approche que vous choisissez pour la FAIRisation de vos données (par l'intermédiaire d'un dépôt institutionnel ou par le biais d'une infrastructure personnelle), c'est à l'auteur de données qu'il appartient de mettre en place une documentation appropriée.

## Licence d'utilisation

### Principes FAIR concernés:

R1.1

### Points à retenir:

- ☞ Les métadonnées doivent contenir de façon claire et explicite la licence d'utilisation associée aux données qu'elles [métadonnées] accompagnent.

### Comment procéder?

**dépôt de données:** Si vous entreposez les données dans un dépôt de données, vous pouvez choisir la licence lors de la mise en ligne; ce choix sera retenu par le dépôt et ajouté dans les métadonnées générées automatiquement.

**infrastructure personnelle:** Si c'est l'infrastructure personnelle qui a été choisie pour la FAIRisation, il vous revient de vous assurer que vos métadonnées contiennent, de façon claire et explicite, la licence d'utilisation appropriée.

## Provenance des données

### Principes FAIR concernés:

R1.2

### Points à retenir:

☞ Les métadonnées doivent contenir les informations sur la provenance des données. En pratique, il s'agit, le plus souvent, d'indiquer la provenance de données en tant qu'objet digital. Par exemple, les données peuvent être produites lors d'un autre projet et ensuite réutilisées, corrigées et modifiées par le présent projet. Le contenu de ce champ peut, par exemple, être le suivant: *"Les données qui concernent les VIII<sup>e</sup> et IX<sup>e</sup> siècles ont été extraites d'un jeu de donnée sous format SQL produit lors du projet "X" et corrigées ensuite par Monsieur Untel. Les données qui concernent le XII<sup>e</sup> siècle ont été extraites d'un jeu de donnée sous format XML produit lors du projet "Y" sans aucune correction supplémentaire. Toutes les autres données ont été saisies manuellement par Monsieur Unautre et Madame Uneautre sous format des documents JSON dans la base MongoDB"*. Pour plus de détail sur la notion de la provenance des données, voir par exemple: [RD-Alliance/Issues: Indicators for R1.2: \(meta\)data are associated with detailed provenance](#).

### Comment procéder?

**dépôt de données:** Si vous entreposez les données dans un dépôt de données, vous pouvez soit inclure les informations sur la provenance des données dans le champ "Description" lors de la mise en ligne, soit accompagner vos données avec un document explicatif (par exemple un fichier "Readme") qui contiendra ces informations.

**infrastructure personnelle:** Si c'est l'infrastructure personnelle qui a été choisie pour la FAIRisation, vous pouvez soit inclure les informations sur la provenance des données dans vos métadonnées, soit accompagner vos données avec un document explicatif (par exemple un fichier "Readme") qui contiendra ces informations.

Pour plus de détails sur le programme de la FAIRisation du consortium COSME<sup>2</sup>, voir: [COSME<sup>2</sup>{FAIR}](#)