
FAIRiser des données : état des lieux, barrières et choix .Une réflexion à partir des données des corpus d'auteurs ?

Ioana Galleron*¹ and Fatiha Idmhand*²

¹LATTICE – Université Sorbonne Nouvelle - Paris 3, CNRS : UMR8094 – France

²CRLA-Archivos ITEM (ITEM) – CNRS : UMRUMR8132, ITEM – France

Résumé

Depuis leur publication en 2016 dans *Scientific Data* (1), les principes FAIR ont été présentés comme des "lignes directrices" destinées à faciliter l'accessibilité et la réutilisation des données et ressources scientifiques numériques. En tant que "principes", ils orientent, suggèrent et encouragent leur mise en œuvre en vue d'une recherche plus ouverte, plus partagée et transparente. Toutefois, depuis que certains États, organismes (2) et la Commission Européenne exigent des projets financés de les suivre de façon impérative, leur application est devenue un enjeu financier et technique qui élude trop souvent la nécessaire réflexion scientifique qui doit accompagner la circulation des objets d'études en milieu numérique. Du point de vue technique, respecter l'ensemble des principes requiert des efforts conséquents et onéreux que seules de très grandes infrastructures (comme Huma-Num en France) peuvent assumer. Ainsi, l'existence d'un écosystème numérique est essentielle pour stocker les données et leurs métadonnées, pour leur délivrer un identifiant pérenne unique (ex. DOI ou URL) de sorte qu'elles soient faciles à trouver (F), pour accéder à long terme aux données et les télécharger (A), pour pouvoir les utiliser et les croiser avec d'autres jeux de données (I) et, enfin, pour les réutiliser dans de nouvelles recherches scientifiques (R).

Si, à première vue, les principes FAIR semblent dessiner un horizon technique à destination des systèmes informatiques d'abord, et des humains ensuite, créer de nouvelles connaissances à partir de ces données nécessite de réfléchir à la façon dont les pratiques actuelles des humanistes numériques encouragent ou, à l'inverse, bloquent, la recherche à partir de leurs données et métadonnées. Dans un travail antérieur, nous avons engagé la réflexion en interrogeant la notion de "Réutilisabilité" (R) dans l'édition électronique (3), au terme de nos observations, nous avons suggéré que pour un même projet, différents états du fichier numérique soient proposés de façon à faciliter la réutilisation de celui-ci. Cette contribution propose de prolonger la réflexion initiée sur la "Réutilisabilité" et de questionner les autres dimensions des principes FAIR : pourquoi "facile à trouver" et interface (un site web le plus souvent) peuvent entrer en contradiction ? quels sont, dans le contexte des Humanités numériques, les freins à l'accessibilité ? enfin, on se demandera pourquoi l'interopérabilité n'est pas qu'une question de langage et de technique, mais un problème de positionnement de la recherche scientifique sur les textes dans les nouvelles frontières des disciplines ?

Pour exemplifier notre propos, nous nous appuyerons sur le dépouillement récemment réalisé

*Intervenant

des éditions et données produites par les membres du consortium CAHIER. Avec une soixantaine de projets, plus de 327000 fichiers texte et près de 500.000 images, CAHIER représente un terrain d'étude éclairant au sujet des obstacles et difficultés qui jalonnent le chemin de la FAIRisation dans une communauté ayant développé antérieurement des pratiques et des standards autour de l'édition papier (4). L'acculturation nécessaire pour passer d'un régime de publication à un autre, ainsi que l'accompagnement technique mis en œuvre dans le cadre de CAHIER seront abordés tour à tour dans notre communication.

Bibliographie

(1) <https://www.nature.com/articles/sdata201618>

(2) <https://www.science-ouverte.cnrs.fr/>

(3) **Ioana Galleron and Fatiha Idmhand**, " De l'interopérabilité à la réutilisabilité des éditions électroniques ", *Humanités numériques* [Online], 1 — 2020. URL: <http://journals.openedition.org/revuehn> ; DOI : <https://doi.org/10.4000/revuehn.350>

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