
FAIR serait bien, bien FAIR”e” serait mieux

Isabelle Charpentier^{*1,2}, Cécile Pignol^{3,4}, Wilfried Heintz^{5,6}, Mathias Rouan^{7,8}, and David Sarramia^{9,10}

¹LTSER-FR Zone Atelier Environnementale Urbaine (ZAEU) – Centre National de la Recherche Scientifique – France

²Laboratoire des sciences de l’ingénieur, de l’informatique et de l’imagerie – université de Strasbourg, Centre National de la Recherche Scientifique : UMR7357 – France

³LTSER-FR Zone Atelier Alpes (ZAA) – CNRS-INEE – France

⁴Environnements, Dynamiques et Territoires de la Montagne – Centre National de la Recherche Scientifique : UMR5204 / FRE 2641, Université Savoie Mont Blanc – France

⁵UMR Dynafor – Institut national de recherche pour l’agriculture, l’alimentation et l’environnement (INRAE), Institut National Polytechnique de Toulouse - INPT – France

⁶Zone Atelier Pyrénées-Garonne – Institut Ecologie et Environnement – France

⁷Littoral, Environnement, Télédétection, Géomatique (LETG - Géomer) – Université de Rennes II - Haute Bretagne, Université de Nantes, Université de Caen Basse-Normandie, CNRS : UMR6554, Université d’Angers, Institut Universitaire Européen de la Mer (IUEM), Université de Bretagne Occidentale (UBO), Université de Bretagne Occidentale [UBO] – Place Nicolas Copernic 29280 PLOUZANE, France

⁸LTSER-FR Zone Atelier Brest Iroise – CNRS INEE – France

⁹Laboratoire de Physique Corpusculaire (LPC) – CNRS/IN2P3 UMR6533 – France

¹⁰LTSER-FR Zone Atelier Territoires Uranifères – CNRS INEE – France

Résumé

Pour répondre à des ambitions institutionnelles, gouvernementales et européennes d’une science plus ouverte, la mise à disposition des données FAIR et des métadonnées les documentant s’organise à tous les niveaux. Nombreux sont les outils et les services (entrepôts de données, méthodes de gestion, bonnes pratiques, formations...) actuellement développés et déployés pour supporter la nécessaire ”FAIRisation” des données, c’est à dire pour les rendre trouvables (F) par une indexation correcte, Accessibles, Interopérables et Réutilisables. Si les propriétés ”F” et ”A” des jeux de données sont principalement gérées par les outils de collecte de fiches de métadonnées mis à la disposition des chercheurs, leurs propriétés ”I” mais surtout ”R” sont bien plus délicates à assurer par le biais d’outils, demandant alors un investissement important en temps et en formation pour le chercheur, avec souvent le besoin d’un accompagnement ciblé par un spécialiste de la donnée.

Comme illustration, nous avons analysé automatiquement les fiches de métadonnées décrivant les sites LTER décrits dans les systèmes de gestion de l’information DEIMS-SDR (1) et DIP-Iter. Si d’un point de vue ”FAIR”, les fiches de métadonnées sont a priori ”FAI” de par le modèle proposé par le système d’information lui-même, les informations qu’elles contiennent

*Intervenant

sont en qualité et en quantité très variables. Les premières constatations montrent que certaines rubriques sont mal interprétées ou non complétées par une portion non négligeable de gestionnaires de sites. A l’opposé, les contours de sites (shapefiles) fournis par les gestionnaires de site sont parfois ”non-simplifiés”, impliquant alors un espace mémoire et des flux de données exagérés. Enfin, la description du dictionnaire des données est limitée au nom du paramètre mesuré aligné par exemple avec EnvThes, sans information claire sur le protocole d’acquisition. Un modèle de fiche de métadonnées FAIR et la bonne volonté des gestionnaires ne suffisent pas toujours à bien FAIR.

Cette analyse conduit aux questionnements suivants. Quelle modération et quel enrichissement mettre œuvre pour accompagner une ouverture ”de qualité” des données ? Par qui ? A quel niveau du cycle de vie (collecte et/ou dépôt) ? Comment (manuelle ou intelligence artificielle) ? Avec quels moyens ?

Enfin, comme soulignée dans la contribution (2), la dimension”e” (économique, écologique et énergivore) est aujourd’hui absente du FAIR alors même que toutes les étapes du cycle de la donnée ont un coût énergétique. Du jeu de données brutes, potentiellement lourd, aux données simplifiées ou agrégées, anticiper le bien FAIR”e” ne serait-il pas indispensable dès à présent ?

(1) Wohner, Peterseil, Genazzio, Guru, Hugo, Klug (2020). Towards interoperable research site documentation – Recommendations for information models and data provision, *Ecological Informatics*, 60, 101158, <https://doi.org/10.1016/j.ecoinf.2020.101158>.

(2) Mallarino (2022). Les impacts du numérique face à l’urgence environnementale, <https://prospectives21.sciencesco>

Mots-Clés: Métadonnées, analyse, qualité, impact environnementale, modération, enrichissement