
Interopérabilité des métadonnées de la Science Ouverte : qu'en est-il de la réalité ?

Vincent-Nam Dang¹, Nathalie Aussenac-Gilles², Imen Megdiche³, Franck Ravat¹

1. IRIT, CNRS (UMR 5505), Université Toulouse Capitole, France

vincent-nam.dang@irit.fr, franck.ravat@irit.fr

2. IRIT, CNRS (UMR 5505), France

nathalie.aussenac-gilles@irit.fr

3. IRIT, CNRS (UMR 5505), INU Champollion, ISIS Castres, Université de Toulouse, France

imen.megdiche@univ-jfc.fr

REFERENCE DE L'ARTICLE INTERNATIONAL Cet article est un résumé de l'article : Dang, V. N., Aussenac-Gilles, N., Megdiche, I., & Ravat, F. (2023, May). Interoperability of Open Science Metadata: What About the Reality?. In International Conference on Research Challenges in Information Science (pp. 467-482). Cham: Springer Nature Switzerland.

La Science Ouverte est un mouvement de la recherche scientifique visant à mettre en place des échanges d'informations inter-communautaires et inter-domaines. Ces échanges ont pour objectif d'enrichir le processus de création de connaissance de la recherche. Ce mouvement promeut en particulier la mise à disposition auprès de tous des jeux de données, algorithmes, codes informatiques et processus méthodologiques utilisés et produits par les recherches présentées dans des articles. Pour en faciliter la recherche et la réutilisation, il est prévu que ces jeux de données soient décrits par des métadonnées et rendus accessibles via des plateformes de partage sur le Web. Cependant, l'utilisation conjointe de plusieurs jeux de données venant de différentes sources suppose qu'ils sont interopérables. La grande variété de modèles de métadonnées utilisés dans la Science Ouverte ne permet pas une interopérabilité native de ces métadonnées. Une solution est apportée par les outils de mappings automatiques de modèles de métadonnées. Cependant, on observe un écart entre les performances observées de ces outils dans les campagnes d'évaluations et leur utilisation sur des données réelles. Pour comprendre précisément les besoins liés à la mise en place de la Science Ouverte, nous avons exploré l'état des lieux de l'interopérabilité des métadonnées dans la Science Ouverte.

Dans un premier temps, nous avons mené une étude comparative des différentes définitions en nous basant sur les critères relatifs à une théorie formelle de l'interopérabilité. Aucune des définitions ne permet de répondre à tous les critères.

Nous avons donc proposé une définition de l'interopérabilité répondant à ces critères et permettant de répondre aux problématiques soulevées par les autres définitions. Nous définissons l'interopérabilité comme « la capacité de 2 entités communicantes à travailler en collaboration à travers un échange d'informations pour réaliser un objectif ». Cette définition repose sur 3 aspects : le contexte de la communication, les entités communicantes et l'objectif à réaliser. Nous avons proposé un modèle en 7 couches, nécessaires à la mise en place d'une interopérabilité. Chaque couche possède un objectif précis ainsi qu'une liste de mécanismes d'interopérabilité associés, dépendants du contexte, des entités communicantes et de l'objectif de la communication. On observe 2 catégories de mécanismes d'interopérabilité : les mécanismes de standardisation et les mécanismes de mise en place de passerelles.

Nous avons exploré la dernière couche de ce modèle, la couche de croisement (mise en correspondance ou mapping). L'objectif de cette couche est la mise en place d'un croisement des informations entre les 2 entités communicantes. Dans le contexte de la Science Ouverte, il s'agit de mettre en place un croisement des modèles de métadonnées pour permettre une recherche sur plusieurs plateformes de gestion de données. Nous avons évalué plusieurs outils, comme COMA provenant de la campagne d'évaluation contrôlée OAEL, afin de voir si les outils de mappings automatiques sont capables de mettre en place une interopérabilité automatique sur des métadonnées provenant de plateformes de gestion de données de la Science Ouverte. Les expérimentations ont montré leur inadéquation au monde de la Science Ouverte, avec un F1-score maximum de 0.21, malgré les bons résultats observés dans leur évaluation initiale. Nous avons réalisé une étude statistique de ces résultats. Il ressort une corrélation négative entre la taille du modèle et les performances des outils. De plus, nous avons observé que seules les métadonnées générales et techniques ont réussi à être correctement appariées.

Nous en avons retenu 3 pistes potentielles pour une amélioration de l'interopérabilité des métadonnées : (i) la classification des métadonnées (générales, techniques et spécifiques à un domaine) permettant une sélection de sous-modèles pour effectuer le mapping et améliorer l'alignement de ces modèles, (ii) une utilisation croisée d'outils génériques et spécifiques à un domaine permettant de palier les problèmes de spécifications des vocabulaires liés à certains domaines, (iii) la conception d'une solution architecturale décentralisée et communautaire pour permettre le passage à l'échelle des solutions de mappings manuels, qui mettent à profit des connaissances d'experts déjà présentes dans la recherche.