

---

## OSDN, une plateforme pour la recherche interdisciplinaire en Science Ouverte

Vincent-Nam Dang <sup>1</sup>, Nathalie Aussenac-Gilles <sup>2</sup>, Imen Megdiche<sup>3</sup>  
Franck Ravat <sup>1</sup>

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

[Vincent-nam.Dang@irit.fr](mailto:Vincent-nam.Dang@irit.fr), [Franck.Ravat@irit.fr](mailto:Franck.Ravat@irit.fr)

2. IRIT, CNRS (UMR 5505), France

[Nathalie.aussenac-gilles@irit.fr](mailto:Nathalie.aussenac-gilles@irit.fr)

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

[Imen.Megdiche@irit.fr](mailto:Imen.Megdiche@irit.fr)

---

REFERENCE DE L'ARTICLE INTERNATIONAL Vincent-Nam Dang, Nathalie Aussenac-Gilles, Imen Megdiche, Franck Ravat. *Enabling Interdisciplinary Research in Open Science: Open Science Data Network*. In: Araújo, J., de la Vara, J.L., Santos, M.Y., Assar, S. (eds) *Research Challenges in Information Science*. RCIS 2024. LNBIP, vol 513. Springer. 19-34.

---

### 1. Introduction

La science ouverte est un mouvement qui vise à améliorer le processus de création de connaissances pour permettre la collaboration entre différentes communautés de recherche, qu'elles soient inter- ou intra-disciplinaires. La science ouverte est adossée aux principes FAIR, qui décrivent quatre piliers pour améliorer le partage et l'accès aux données. Parmi ces principes, la "trouvabilité" est définie comme la facilité à trouver des données pour les humains et les machines (Jacobsen A. et al, 2020). Les chercheurs de différents domaines et communautés soulignent d'importants freins lorsqu'ils recherchent des données brutes ou des jeux de données résultant de travaux de recherche. Ces freins sont essentiellement liés aux : (i) modèles de métadonnées des sources de données peu interopérables, (ii) manque de coordination entre les initiatives disciplinaires conduisant à un nombre très important de plateformes de gestion des données de recherche. L'un des moyens de résoudre ces problèmes consiste à mettre en place une plateforme centralisée de gestion des données pour la Science Ouverte. Mais la centralisation pose plusieurs problèmes : (i) impossibilité de répondre aux besoins spécifiques des chercheurs des

différentes communautés ; (ii) volume trop important pour une seule plateforme ; (iii) coût de déploiement trop élevé d'une plateforme qui deviendrait un point de fragilité ; (iv) aucune garantie de sécurité et de confidentialité.

## 2. La plateforme OSDN

La plateforme Open Science Data Network (OSDN) que nous proposons consiste en un réseau décentralisé, fédéré et distribué de plateformes de gestion de données de la Science Ouverte. Cette solution est basée sur une API RESTful utilisant un registre partagé par toutes les plateformes, qui contient des informations relatives aux plateformes (nom, URL, plateformes interconnectées, etc.), aux modèles de métadonnées utilisés (nom, contenu, etc.) et aux correspondances entre les modèles de métadonnées. Un mécanisme de propagation des modifications les diffuse à tous les voisins jusqu'à atteindre l'ensemble du réseau. Ce module prend la forme d'un conteneur Docker à déploiement automatique dont l'intégration ne nécessite que l'installation de la fonction d'interopération entre ce module et le mécanisme de recherche d'information de la plateforme. Choisir la bonne topologie du réseau est déterminant pour bien gérer les suppressions volontaires et involontaires. Une solution qui minimise la vulnérabilité à ces deux événements est la topologie de réseau sans échelle avec un seul hub, les autres nœuds ayant tous un degré de cinq. Pour évaluer la faisabilité de l'OSDN, nous avons développé une preuve de concept qui intègre 3 plateformes de Science ouverte avec différentes technologies de gestion des métadonnées. Nous avons exécuté une requête sur une plateforme et vérifié la propagation de cette requête aux plateformes voisines, puis aux voisins indirects. Pour garantir la gestion des modèles de métadonnées, nous avons intégré et mis en relation 19 modèles de métadonnées dans le registre de l'OSDN. Nous avons créé manuellement des correspondances entre ces modèles structurels afin d'assurer leur interopérabilité. Pour l'expérimentation de OSDN, un chercheur spécialiste en agronomie a exploité la plateforme pour chercher des jeux de données décrivant une activité de biocontrôle destinée aux agriculteurs. Plus précisément, il cherchait des données relatives à la souche « *trichoderma harzianum* T-22 ». Grâce à l'utilisation de l'OSDN, le temps de recherche est réduit de 80 % ; il accède à un plus grand nombre de jeux de données (+ 7%). Dans le cadre des travaux futurs, nous envisageons de réduire le coût d'adoption de cette solution et de mieux exploiter l'interopérabilité sémantique. Enfin, nous voulons mieux assurer le passage à l'échelle, en particulier en optimisant la consommation de ressources dans l'OSDN.

## 3. Références

Jacobsen A. et al. (2020). Fair principles: interpretations and implementation considerations. *Data intelligence*, vol. 2, no 1-2, p. 10–29