

Les différences entre plusieurs entrepôts de données généralistes : Recherche Data Gouv, Zenodo, Dryad et Figshare

 opencience.pasteur.fr/2025/02/03/les-differences-entre-plusieurs-entrepots-de-donnees-generalistes-recherche-data-gouv-zenodo-dryad-et-figshare/

CeRIS - Institut Pasteur

3 février 2025

Pour ouvrir ses données de recherche, il existe de très nombreux entrepôts de données. Parmi eux, les **entrepôts dits « généralistes »** ou « pluridisciplinaires » permettent le partage de tous types de données de recherche. Leur utilisation est conseillée uniquement dans le cas où il n'existe pas d'entrepôt disciplinaire adapté.

Nous vous proposons aujourd'hui un **comparatif de quatre entrepôts généralistes** afin de mieux comprendre leurs spécificités : Recherche Data Gouv (que nous vous présentons ici), Zenodo (que nous vous présentons ici), Dryad et Figshare.com.

Commençons par leurs points communs : ces entrepôts sont assez **simples d'utilisation** : en quelques étapes, le scientifique peut charger ses fichiers de données, les décrire par des métadonnées, choisir une licence et publier son jeu de données. Un **DOI** est alors attribué au jeu de données qui devient un produit de recherche citable et trouvable via des moteurs de recherche.

Faisons maintenant le point sur les différences entre ces entrepôts :

Qui développe, héberge et finance l'entrepôt ?

- Recherche Data Gouv (RDG), Zenodo et Dryad sont soutenus par des institutions à **but non lucratif**.
- Figshare est la propriété de *Digital Science*, une filiale du groupe allemand *Holtzbrinck Publishing Group* (actionnaire majoritaire de Springer Nature). Retrouvez dans cette note de décryptage produite par le GTSO Données de Couperin les **risques encourus** lors du dépôt de données dans un **entrepôt d'éditeur commercial**.

Qui peut utiliser l'entrepôt ?

- RDG peut être utilisé pour partager des jeux de données dont au moins un des co-auteurs est **affilié à un établissement français**.
- Zenodo, Dryad et Figshare.com peuvent être utilisés par toute la communauté scientifique (NB : le logiciel figshare peut être utilisé par différentes organisations, éditeurs ou organisations de recherche par exemple, leur entrepôt n'est alors pas ouvert à tous).

L'entrepôt est-il dédié aux données de la recherche ou peut-il être utilisé pour partager d'autres types de contenu ?

- RDG et Dryad sont des entrepôts dédiés au partage des données de la recherche.
- Zenodo et Figshare.com sont ouverts à une large diversité de produits de recherche (publications, jeux de données, rapports, présentations, posters...).

L'entrepôt assure-t-il une curation des jeux de données pour s'assurer qu'ils sont correctement décrits et réutilisables, conformément aux principes FAIR ?

- Dans RDG, tous les jeux de données passent par une étape de curation, par les curateurs de l'espace générique ou des espaces institutionnels, suivant des règles communes.
- Zenodo n'a pas mis en place de curation par défaut. Une curation peut avoir lieu si le jeu de données est publié dans une communauté, selon des règles définies au sein de chaque communauté.
- Dryad dispose d'une équipe de curateurs qui se charge de la curation de chaque jeu de données déposé.
- Figshare.com n'assure pas de curation des jeux de données (certaines institutions utilisant Figshare comme entrepôt institutionnel peuvent toutefois assurer une curation).

Le volume des jeux de données est-il limité ?

- Dans RDG, la taille des jeux de données n'est pas limitée. Cependant, chaque institution peut fixer une limite de taille afin d'éviter d'atteindre trop rapidement son quota (un espace de 5 To est alloué à chaque espace institutionnel à sa création).
- Dans Zenodo, le volume de chaque jeu de données est limité à 50 GB, avec possibilité d'augmenter exceptionnellement le quota à 200 GB.
- Dans Dryad, la limite est de 300 GB par jeu de données.
- Dans Figshare.com, les comptes gratuits disposent d'un espace de stockage limité à 20 GB. L'utilisation de Figshare+ est nécessaire pour un volume supérieur (service payant).

Le dépôt de données est-il payant ?

- Dans RDG et Zenodo, le dépôt de données est gratuit pour les scientifiques.
- Dans Dryad, si vous ne faites pas partie d'une institution membre, le dépôt est payant pour couvrir les coûts liés à la curation et à la préservation des données sur le long terme (*Data Publishing Charges* de 150\$, auxquels s'ajoutent des frais au-delà de 50 GB). Certaines revues peuvent prendre en charge ces frais.
- Figshare.com est gratuit jusqu'à 20 GB par compte individuel. Avec Figshare+, le coût est de 450\$ jusqu'à 100 GB, puis 875\$ par tranche de 250 GB.

Quelles sont les possibilités offertes par l'entrepôt concernant les conditions d'accès et la licence à associer au jeu de données ?

- RDG offre la possibilité de définir des conditions d'accès différentes pour chaque fichier (accès ouvert, ouvert après embargo, restreint ou fermé). RDG propose une licence Etalab par défaut mais il est possible de choisir une autre licence.
- Dans Zenodo, les conditions d'accès sont définies au niveau du jeu de données (accès ouvert, ouvert après embargo, restreint ou fermé). Zenodo laisse la possibilité de choisir n'importe quelle licence disponible publiquement.
- Dans Dryad, tous les jeux de données sont en accès ouvert (après un éventuel embargo) sous licence CC0.
- Dans Figshare.com, tous les jeux de données sont en accès ouvert (après un éventuel embargo) sous licence CC0 ou CC BY.